

# When is an alternative possibility robust?

Simon Kittle 

Department for Christian Philosophy,  
University of Innsbruck, Innsbruck, Austria

**Correspondence**

Simon Kittle, Department for Christian  
Philosophy, University of Innsbruck, Karl-  
Rahner Platz 1, Innsbruck 6020, Austria.  
Email: simon@kittle.co.uk

**Abstract**

According to some, free will requires alternative possibilities. But not any old alternative possibility will do. Sometimes, being able to bring about an alternative does not bestow any control on an agent. In order to bestow control, and so be directly relevant qua alternative to grounding the agent's moral responsibility, alternatives need to be robust. Here, I investigate the nature of robust alternatives. I argue that Derk Pereboom's latest robustness criterion is too strong, and I suggest a different criterion based on the idea that what agents need to be able to do is keep open the possibility of securing their blamelessness, rather than needing to directly ensure their own blamelessness at the time of decision.

## 1 | INTRODUCTION

The literature on Frankfurt-style cases (FSCs) is well known. In contrast to Harry Frankfurt's intentions, it has provided several useful insights into how the avoidability criterion on moral responsibility will need to work if it is to be successful.<sup>1</sup> My aim in this paper is to discuss when an alternative possibility is *robust* (better: when the abilities which give rise to those possibilities are robust)—that is, when they explain, in and of themselves, an agent's moral responsibility. My starting point will be a consideration of Derk Pereboom's latest indeterministic Frankfurt-style case, *Tax Cut*. I will suggest that Pereboom's robustness criterion is too strong and propose an alternative understanding of robustness which explains both Pereboom's case and several further examples I will present.

Frankfurt-style cases purport to show that an agent can be responsible for something she does despite not having been able to avoid it. They were proposed as counterexamples to the Principle of Alternative<sup>2</sup> Possibilities (PAP), which states that “a person is morally responsible for what he has done only if he could have done otherwise” (Frankfurt, 1969, p. 829). I will call the person who holds that alternative possibilities (APs) are required for free will a *choice-theorist*, as such theorists typically think that alternatives are important because of their connection to a person's being able to choose.<sup>3</sup> I will call the person who thinks FSCs are successful a *Frankfurtian* or a *Frankfurt-defender*. One early Frankfurt-style case, presented by John Martin Fischer, runs as follows:

---

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors European Journal of Philosophy Published by John Wiley & Sons Ltd

*Black is a nefarious neurosurgeon ... [who] inserts a mechanism into Jones's brain which enables Black to monitor and control Jones's activities. Jones, meanwhile, knows nothing of this. Black exercises this control through a computer .... If Jones shows an inclination to decide to vote for Carter, then the computer, through the mechanism in Jones's brain, intervenes to assure that he actually decides to vote for Reagan and does so vote. But if Jones decides on his own to vote for Reagan, the computer does nothing but continue to monitor ... Jones. Suppose that Jones decides to vote for Reagan on his own (Fischer, 1982, p. 26).*

The case is meant to work like this: we are invited to agree that Jones is responsible for deciding to vote for Reagan because it is part of the case that Jones does this “on this own” and that Black’s device does not interfere in any way. The proponent of the case then points out that the *mere presence* of Black’s device meant that Jones could not have voted (nor even decided to vote) for Carter. But that means that Jones is responsible despite not having any options, and so PAP is false.

In the discussion which followed it was quickly pointed out that many of the early cases relied on what has come to be known as a *prior sign*: an indicator which the intervener—Black, in the case above—uses to work out whether to intervene. Fischer talks above of Jones “showing an inclination” of voting for Carter without specifying the nature of that inclination. Others talked about Jones *blushing* or *exhibiting some neural pattern* which reveals what he is going to do. Whatever the precise form, a worry arises for the proponent of the cases. If the prior sign is a perfectly reliable indicator of Jones’s decision, then it seems that there must be a deterministic connection between the prior sign and the decision. While the compatibilist proponent of PAP will have no grounds on which to object to this, the incompatibilist proponent of PAP now has a reason to dismiss the case. That’s because the incompatibilist holds that the making of free decisions is incompatible with those decisions being determined. On the other hand, if the prior sign isn’t a reliable indicator of Jones’s subsequent decision, then it seems like Jones might have had an alternative after all. If these latter FSCs demonstrate anything, then it is has not yet been shown that it is anything which needs to worry the incompatibilist. This reply is known as the Dilemma Defence and was advanced by Robert Kane (1985, p. 51 fn. 25), Carl Ginet (1996), and David Widerker (2003).

## 2 | PEREBOOM AND HIS ROBUSTNESS CRITERION

One line of reply to the Dilemma Defence has been to develop explicitly indeterministic FSCs without any deterministic connection between the prior sign and the agent’s subsequent decision. The most powerful of these are the so-called buffer cases invented independently by David Hunt (2005) and Derk Pereboom (2001, pp. 18ff). The latest such case put forward by Pereboom runs as follows:

**Tax Cut:** *Jones can vote for or against a modest tax cut for those in his high-income group by pushing either the “yes” or the “no” button in the voting booth. Once he has entered the voting booth, he has exactly two minutes to vote, and a downward-to-zero ticking timer is prominently displayed. If he does not vote, he will have to pay a fine, substantial enough so that in his situation he is committed with certainty to voting (either for or against), and this is underlain by the fact that the prospect of the fine, together with background conditions, causally determines him to vote. Jones has concluded that voting for the tax cut is barely on balance morally wrong, since he believes it would not stimulate the economy appreciably, while adding wealth to the already wealthy without helping the less well off, despite how it has been advertised. He is receptive and reactive to these general sorts of moral reasons: he would vote against a substantially larger tax cut for his income group on account of reasons of this sort, and has actually done so in the past. He spends some time in the voting booth rehearsing the relevant moral and self-interested reasons. But what would be required for him to decide to vote against the tax cut is for him to vividly imagine that his boss would find out, whereupon due to her political leanings she*

would punish him by not promoting him to a better position. In this situation it is causally necessary for his not deciding to vote for the tax cut, and to vote against it instead, that he vividly imagine her finding out and not being promoted, which can occur to him involuntarily or else voluntarily by his libertarian free will. Jones understands that imagining the punishment scenario will put him in a motivational position to vote against. But so imagining is not causally sufficient for his deciding to vote against the tax cut, for even then he could still, by his libertarian free will, either decide to vote for or against (without the intervener's device in place). However, a neuroscientist has, unbeknownst to him, implanted a device in his brain, which, were it to sense his vividly imagining the punishment scenario, would stimulate his brain so as to causally determine the decision to vote for the tax cut. Jones's imagination is not exercised in this way, and he decides to vote in favor while the device remains idle (Pereboom, 2012, pp. 308–309).

Ignoring the neuroscientist intervener for the moment, the structure of this case is as follows. Jones is considering whether to vote for a tax cut. We are to suppose that at any point in the scenario he can decide to vote for the cut. But he cannot at any point decide to vote *against* the cut. To make such a decision, Jones must first imagine his boss finding out and subsequently punishing him. Only if he were to imagine this would be able to decide to vote against the tax cut. Crucially, this wouldn't settle anything: after Jones imagines being punished, he would be able to decide either way. Jones can voluntarily imagine being punished, and he understands that if he does so he will then be able to decide to vote against the cut. Now let's bring the intervener back in. Given this set-up, Jones's imagining is a necessary but not sufficient condition for Jones's deciding to vote against the tax cut. Therefore, the intervener can use this as the prior sign without assuming a deterministic connection that would allow the incompatibilist to dismiss the case.

Now, Pereboom accepts of course that the choice-theorist can explain Jones's responsibility for his *not imagining being punished*: Jones has a robust alternative for that. What Pereboom denies is that Jones has a robust alternative with respect to *his decision to vote for the tax cut*. Pereboom doesn't think that Jones's alternative is *robust* because it is not *exempting*. Carlos Moya used the word "exempting" as a way of explaining and expanding on Pereboom's early robustness criterion (Moya, 2006, p. 58). Since then, Pereboom has adopted and endorsed the terminology. For Pereboom, an alternative is exempting only if, were the agent to avail himself of the alternative, *and were there no intervener*, the agent would succeed in bearing no blame for performing an action of the type in question (Pereboom, 2014, pp. 16–17). To be explicit: Pereboom assesses whether an alternative is exempting by asking what would happen in that alternative "without the device in place" (Pereboom, 2014, p. 16). In Jones's case, an alternative would be exempting if, were Jones to avail himself of it and were there no intervener, Jones would (or would be likely to) bear no blame for voting for the tax cut. In other words, exempting alternatives are such that if realised, the agent would *ensure* that she bears no blame for performing the relevant type of action. Jones's alternative clearly doesn't satisfy this requirement because it is stipulated as part of the case that even after imagining being punished Jones still has to make a free decision about what to do. Therefore, according to Pereboom's understanding of robustness, *imagining being punished* does not in and of itself preclude Jones from blame for deciding to vote for the cut. Indeed, the Frankfurt defender can even stipulate that after imagining being punished Jones would still be *likely* to choose to vote for the tax cut (Pereboom suggests this move in a recent discussion of his earlier *Tax Evasion* (2) case [Pereboom, 2014, p. 17]). Pereboom formalises these ideas in the following robustness criterion (Pereboom, 2014, p. 13):

**Robustness (B):** For an agent to have a robust alternative to her immoral action A, that is, an alternative relevant per se to explaining why she is blameworthy for performing A, it must be that

- i. she instead could have voluntarily acted or refrained from acting as a result of which she would be blameless, and
- ii. for at least one such exempting acting or refraining, she was cognitively sensitive to the fact that she could so voluntarily act or refrain, and to the fact that if she voluntarily so acted or refrained she would then be, or would likely be, blameless.

Is Robustness (B) a satisfactory criterion of robustness? To answer that we need to know what it applies to. Taken at face value, “A” is a name for a particular action that some agent performs. This reading, I submit, is unsatisfactory. It states baldly that the agent is blameworthy for some particular action A, but a particular action is going to have lots of properties of which the agent is entirely unaware and which could not be controlled even if the agent was aware of them. Consider Jones’s voting. Let’s assume this involves making a mark in a box on a voting slip. Jones’s vote, like any particular instance of voting, will involve making a mark of some determinate length in one of the boxes, using ink of a particular shade, and so on. Let’s assume the mark Jones makes is 5.667 mm long and in a shade half way between Columbia blue and Cornflower blue. On the current reading, Robustness (B) has it that Jones is responsible for the particular act he performed—all of it. That includes the properties *being a vote made with a mark 5.677 mm long and being made in a shade of blue half way between Columbia and Cornflower blue*. But is Jones responsible for such aspects of his action? That seems doubtful. To begin with, if Jones isn’t aware of these properties of his action, they are not properties for which he can be responsible. But even if we gerrymandered a context where Jones was aware of such properties and where they had moral importance—e.g., Jones lives under a strange fundamentalist government which has decreed that all votes must be made with marks having an even length in microns, which would result in it being morally good to vote by making an even-length mark (because obeying the law, avoiding punishment for oneself, one’s family, etc., are moral concerns)—that would not show that Jones can be responsible for them. That’s because they lie outside the scope of his control: humans cannot control marks made on pieces of paper with that kind of precision. Obviously, this point generalises to all of our ordinary actions.

It is far more plausible to hold that agents only control *some aspects* of the particular actions they perform, and that therefore they are only *morally responsible* for some aspects of their particular performances. The following reframing of Robustness (B) aims to capture this idea while also retaining the spirit of Pereboom’s Robustness (B):

**Robustness (C):** For an agent to have a robust alternative to her **having A-ed**, which it was immoral to do, that is, an alternative relevant per se to explaining why she is blameworthy for **having A-ed**, it must be that

- i. she instead could have voluntarily acted or refrained from acting as a result of which she would be blameless with **respect to having A-ed**, and
- ii. for at least one such exempting acting or refraining, she was cognitively sensitive to the fact that she could so voluntarily act or refrain, and to the fact that if she voluntarily so acted or refrained she would then be, or would likely be, blameless **with respect to having A-ed**.

The idea with Robustness (C) is that “A” should be filled in with a description of what the agent has done; “A” is not a proper name of a particular action, and neither should it be replaced with a phrase that contains an indexical that refers to a particular action (such as “performed *that* action”). It should be replaced by phrases like “walked to work” or “punched Tom.” This allows for a more fine-grained treatment because phrases like “having punched Tom” and “having punched Tom hard in the face” name *facts* or *states of affairs*. These contrast with phrases like “the punch Jones gave Tom” and “Jones’s punching of Tom” which name particulars (See Jonathan Bennett (1988, ch 1) for elaboration). And while, say, Jones’s particular action might be a punching of Tom by Jones, and a very powerful punch, and a punch in the face, *the fact that* Jones punched Tom is different to the fact that Jones punched Tom in the face and different again to the fact that Jones punched Tom in the face at full force. Robustness (B) is unsatisfactory because (taken at face value) it applies to particular actions and when satisfied yields the result that an agent is responsible for the entirety of that particular action. This, I have suggested, is implausible. And Robustness (C) addresses that problem because it concerns states of affairs. Not only does it allow us to avoid the conclusion that Jones is responsible for, say, *punching Tom in the face with a force of precisely 99.67543 pounds* (the idea here being that Jones will lack the ability to control the level of force he exerts to that level of specificity, meaning that he can’t be responsible for such a determinate property of one of his actions), it also allows us to distinguish between, say, *punching Tom in the face with all his strength* and *punching Tom*. This is a welcome conclusion. If, for example, Jones’s character (through no fault of his own) made it inevitable that he punch Tom, but didn’t dictate how he did it (where

he punched him, and the rough level of force used), we might be inclined to say that Jones is not responsible for punching Tom simpliciter but he is responsible for punching him in the face with all his strength.

### 3 | ROBUSTNESS AS KEEPING ONE'S OPTIONS OPEN

To recap so far. Robustness (B) is Pereboom's latest robustness criterion, but it applies to particular actions (action tokens). I have suggested that it is implausible that agents are ever in control of (and so responsible for) particular actions (as a whole); rather, agents are in control of performing actions of a certain type—that is, they are in control certain aspects or properties of the actions they perform. This can be expressed by saying that the agent is responsible for certain states of affairs. And Robustness (C) is a re-rendering of the idea behind Pereboom's Robustness (B) which is framed in terms of states of affairs.

Both Robustness (B) and Robustness (C) codify the idea that to have a robust AP an agent needs to be able to ensure that she will be exempt from blame (Pereboom, 2014, pp. 11, 16, 21). This is the key idea behind Pereboom's understanding of robustness: robustness with respect to performing an action of type A requires being able to secure one's blamelessness with respect to performing an action of that type. It is this equating of robustness with exempting oneself from blame that I want to challenge. I will pursue two strategies. First, I will present some examples where Pereboom's robustness criterion isn't satisfied but where I think it is plausible to consider the agent responsible in virtue of the alternatives available. Now, given that I believe it's plausible to think Jones's alternative in *Tax Cut* is robust, but Pereboom and others who agree with him don't, I must present examples where it is *more plausible* that the agent has a robust alternative than it is in *Tax Cut* if I am to stand any chance of persuading my opponents. That is one of the things I will aim to do. Second, I will propose a robustness criterion which I think explains why Jones in *Tax Cut* and the agents in my additional scenarios have robust alternatives.

Here is the first case, which will help us to see that Robustness (C) (i.e., Pereboom's suggestion about robustness, but re-formulated with respect to states of affairs) is too strong:

*(Pete chooses dinner)* Pete is attending a fancy charity gala dinner and must choose in advance which meal he wants. He's had one quick conversation with a friend who is also going, and so happens to know that one of the options is a starter of olives served with a carrot hummus, followed by a baked butternut squash stuffed with jasmine rice. He doesn't know, however, what any of the other options are. To find out what the other options are he would have to dig out his invitation details, visit the website, login and inspect the full menu. He could choose the butternut squash option by firing off a quick email. Or he could investigate the other options and then make a fully informed decision. Crucially, investigating the other options wouldn't close off choosing the butternut squash option. Pete, prior to investigating those options, doesn't know what he would end up choosing were he to look at the full menu, although of course he knows that the only way to choose one of the other options is to first find out what they are.

In this case, there is no Frankfurt-style intervener, no other intervention, no blockages of any kind. Suppose Pete chooses the butternut squash. According to Pereboom's criterion, Pete has a robust alternative, namely, the alternative where he first decides to check the full menu then goes ahead and successfully decides to choose something other than the butternut squash. It is only this second, successful decision which ensures that Pete isn't responsible for deciding to choose the butternut squash, so only alternatives which include some such second decision count as robust.

My contention, however, is that the alternative where Pete decides to investigate the full range of meal options is robust regardless of what happens after Pete's decision to look at all the options. It is robust because by taking this option Pete understands that he is not yet deciding on his meal choice: he is keeping his options open. Pete will conceive of himself as postponing the decision a little bit and will understand that by doing this he is not yet committing himself to the butternut squash option (which was the only alternative available to him at that point). Therefore, in this alternative, he is *avoiding becoming someone who has decided to eat butternut squash* and who would be liable

for any moral responsibility attaching to that decision (if there were any such responsibility). Pete understands all of this. Moreover, Pete is not *just* doing something which he understands is avoiding committing himself then and there to the butternut squash option. He's also doing something which he believes will keep his options open; indeed, he's doing the only thing which he believes will keep his options open. My contention is that acting so as to keep one's options open in this way makes an alternative robust.

Let me expand on the above scenario in a way which I hope makes the above point clearer. Suppose that Pete is reminded on Tuesday at 9 a.m. that he must choose what to eat, and suppose if he does not choose, then a meal will be selected for him at random, such that he's reasonably motivated to choose himself. If Pete were to decide at 9 a.m. to find his details, login to the site and look at the options, he'd discover there were more options than he thought and so he would put them to one side and resolve to look them over properly at 12 p.m. during lunch. However, Pete decides at 9 a.m. to fire off a quick email to confirm his selection of the butternut squash dish.

Now suppose we stop things here and ask whether Pete at 9 a.m. had a robust alternative to choosing the butternut squash. It seems clear to me that the answer is yes. And it also seems clear that we can and should give that positive answer even if we know nothing about what would have happened after the decision he makes at 9 a.m. So suppose instead that Pete decided to login to the website, saw how many other options there were, and so put the issue aside until 12 p.m. Now it's evident that various things might unfold from this point on. Perhaps Pete nips out to get a coffee at 11 a.m. and is run over by a bus. Maybe a colleague distracts him at 11:45 a.m. and he makes no decision at 12 p.m. Perhaps at 12 p.m., Pete browses the menu and makes a decision, just as he intended. Or perhaps there is a Frankfurtian intervener standing by who has been monitoring Pete's access to the website and has resolved to install a brain-manipulation device into Pete's brain and to then use that device to make Pete choose the butternut squash if needed—all of which he can do in a matter of seconds without Pete noticing a thing, of course. None of those potential unfoldings of things seems to undermine the judgement, not just that Pete is responsible, but that *he's responsible because of the choice he had at 9 a.m.*

At 9 a.m., Pete faced a choice between choosing the butternut squash or choosing to look up his login details, visit the website and look at the menu. That choice is *not* symmetrical in the following sense: the second option does not resolve the same issue or uncertainty that the first option does. But does that matter? To be sure, it might be a *natural* enough thought that to be responsible for deciding to A an agent needs to be able to decide to B, where B is an alternative to A which settles the same issue that A would settle, in the way that choosing red is an alternative to choosing blue. Perhaps that is a *natural* thought. But is it *correct*? My suggestion is that the above case shows it to be too strong. **Pete chooses dinner** can be amended with a Frankfurtian intervener so that Pete is unable to *ensure* that he chooses other than the butternut squash. That doesn't show he doesn't have a robust alternative, nor that robust alternatives aren't needed for control—it shows that Robustness (C) is false.

In its place, I want to propose a robustness criterion based on the following idea: what's important for robustness is not that the agent secures her blamelessness with respect to the performance of an action of some type, but that she do something which she believes keeps her options open. In this vein, I suggest the following:

**Robustness (D):** For an agent to have a robust alternative to her **having A-ed**, which it was immoral to do, that is, an alternative relevant per se to explaining why she is blameworthy for A-ing, it must be that

- i. she instead could have voluntarily acted or refrained from acting as a result of which **she would have maintained the same level of blameworthiness with respect to having A-ed**, and
- ii. for at least one such acting or refraining, she was cognitively sensitive to the fact that she could so voluntarily act or refrain, and to the fact that if she voluntarily so acted or refrained she **would have maintained that same level of blameworthiness with respect to having A-ed**, and
- iii. for at least one of the actings or refrainings which satisfies both i and ii, the agent believed either: (a) that realising that alternative would have precluded her from blame, or (b) that realising that alternative would have kept her options open with respect to having A-ed.

Clauses (i) and (ii) here parallel clauses (i) and (ii) in Robustness (C), with the amendment that instead of requiring the agent to be able to do something which would *ensure* her blamelessness, the agent needs only to be able to do something that would maintain her current level of blame with respect to performances of the action type in question. This change, which is discussed more below, is due to not requiring that the agent secure her blamelessness but only that she be able to keep her options open. The construction “blameworthiness with respect to having A-ed” is due to the move to a robustness principle which covers states of affairs. Given that the reason this move was made is because agents are only ever in control of (and so responsible for) *some but not* all aspects of what they do, the states of affairs in question will necessarily be general in character (the whole point was to abstract away from those properties which are not controllable). Unless we then time index these states of affairs, which I have not required in Robustness (D), they might obtain more than once. So Stefan might steal £5 from his brother in the morning, and then again in the evening. Stefan is blameworthy for having stolen £5 after the first theft. After the second theft Stefan is blameworthy for having stolen £5 twice, or, equivalently, has increased his level of blameworthiness with respect to having stolen £5. Here is a more detailed example illustrating the need for this construction:

*(Lustful Larry) Larry is considering whether to entertain a lustful thought. Such thoughts occur to Larry every now and again. Sometimes Larry embraces them, taking what minimal pleasure they offer, and sometimes he dismisses them. Larry's upbringing and own subsequent reflection has convinced him that actively entertaining a lustful thought is wrong; and let's suppose for sake of argument that he is right—lustful fantasising is wrong. One morning while typing away at his keyboard a lustful thought pops into Larry's mind. He understands that he has a choice: he can actively develop the fantasy or he can decide to use one of a few techniques he's learned to help him control his thoughts (maybe a short breathing technique, or reminding himself of the projects he truly cares about from which any fantasy will divert attention, etc). Now let's say that Larry successfully employs one of these techniques several times that morning, but in the afternoon he actively embraces a lustful thought when it occurs to him.*

This example highlights the fact that some types of action are ones we can perform many times each day or even each hour. We might be repeatedly responsible for such actions. Robustness (B) did not need any special machinery to handle such cases because, by employing a proper name for the action in question, it applied to token actions. But as already argued agents are not responsible for action particulars but rather for performing actions of certain types with certain properties. So once we make this move we need to handle the case where agents perform actions of the same type multiple times. Hence, Robustness (D)'s talk of maintaining the same level of blameworthiness with respect to the type of action performed.

The third clause of Robustness (D) is intended to capture the kind of connection needed between the alternative and the decision which is in fact made (or the action which is in fact performed). The main idea here is that the agent needs to believe that she is keeping her options open. This is what sub-clause (b) aims to capture. Clause (a) allows for the kind of connection that Pereboom requires of robustness—my claim being not that the kind of connection Pereboom posits isn't sufficient, but that it is not necessary. Why the switch from *cognitive sensitivity* to *belief* here? Clause (ii) invokes Pereboom's idea of cognitive sensitivity because the agent needs to *correctly* understand that she can indeed do what clause (i) affirms that she is able to do. Being *cognitively sensitive* to something will involve an attitude that entails a good measure of accuracy (even if it doesn't quite amount to knowledge). But on the current view, clause (i) will first and foremost apply to the alternative decision made; e.g., Pete's decision to consider the full menu, Larry's decision to engage in a breathing technique. That is what needs to be successfully performed. The agent's attitude towards *that* decision will need to involve a good measure of accuracy. But clause (iii) concerns some of the future consequences of the proximate alternative decision. Once we consider such consequences, however, we have to allow that what the agent takes herself to be doing might be incorrect inasmuch as what she is aiming at might be frustrated because of the presence of Frankfurtian interveners or other blockages. In the Frankfurt-style amended version of the **Pete chooses dinner** scenario, Pete takes himself to be keeping open the option of choosing a different meal, but he isn't, because the presence of the intervener meant that was never a genuine option in the first place. So



because clause (iii) concerns the more temporally distant consequences of the proximate decision we can't require that the agent's beliefs about her options are accurate. As far as robustness is concerned, however, the contention is that this doesn't matter if clauses (i) and (ii) are satisfied *and* the agent believed that her acting was keeping her options open.

It might be wondered whether Robustness (D) is still too weak.<sup>4</sup> Return to the elaboration of **Pete chooses dinner**, and suppose that at 9:10 a.m. Pete can choose butternut squash by sending the email, or he can look up his login details and check out the full menu, or he can scratch an itch on his head. If he scratches the itch, Pete will face the decision he faces at 9:10 a.m. at say, 9:12 a.m. instead (after he has scratched his head). That doesn't seem to be a robust alternative yet, it might be urged, it satisfies Robustness (D): scratching the itch will not incur any new responsibility with respect to deciding what to have for dinner and it also keeps his options open (and Pete may well have understood this). I agree that this is a problem, but I also think there is a version of this problem which afflicts both Robustness (B) and (C). In other words, the issue raised by this kind of case is everybody's problem. Consider the following case which is inspired by one of Carolina Sartorio's cases (See Sartorio [2016, p. 36]):

*(Squeaky Button Reasons)* I am a big fan of squeaky sounds. I know that pushing a certain button would result in a squeaky sound. I also know that pushing the squeaky button will result in a remote village being wiped out. I'm cognitively sensitive to the fact that I can do something—refraining from pressing the button—which would avoid the village being destroyed. But that plays no role in my decision making—indeed, it cannot play any such role. For whatever psychological reasons, the only consideration for which I could refrain from pressing the button is if I wanted silence at just that moment. As it turns out, I decide I'd like to hear the sound so I press the button.

In this scenario, the alternative I have satisfies Robustness (B) and (C): if I didn't press the button the village would not be wiped out and I am cognitively sensitive to that fact. But it's not clear the alternative here is robust, because it is disconnected from the reasons the agent has. That is, it's stipulated here that I *cannot* refrain from pushing the button *in order to* spare the village. The only reason for which I can refrain from pushing the button is to have silence. In other words, the only alternative I have is one where I refrain from pushing the button in order to obtain silence. Is such an alternative robust? Tentatively, I want to suggest not. Similarly, the alternative where Pete scratches his itch for no other reason than to scratch the itch is not a robust alternative. But an alternative where Pete scratches the itch in order to keep his options open—perhaps he sees this as buying himself a bit more time—is a robust alternative. Whether an alternative is robust, then, depends on the reasons for which the decision in the “alternative sequence” was made. If this thought is on the right lines, the easiest way to accommodate the insight is to add a fourth clause as follows:

- iv. for at least one of the actings or refrainings which satisfies (i)–(iii), the agent is able to bring about the alternative either (a) in order to preclude herself being blameworthy, or (b) in order to keep her options open.

On this view, if Pete can scratch his head in order to buy himself some time (to keep his options open) then he would have a robust alternative (I'm not denying that a similar modification could be applied to Robustness [B] or [C]). Of course, that's not particularly evident in the **Pete chooses dinner** example because Pete can keep his options open simply by deciding to—he doesn't really need to buy any time. But sometimes agents are faced with decisions where the performance of an apparently inconsequential action makes a substantial difference. Consider the following:

*(Walter)* Walter is thinking about whether to go and meet his wife from the airport. His wife has been away on a short trip; there is no expectation for him to go and meet her, and she already has a train ticket home. But meeting her might be a nice surprise. Walter is close to the bus stop which he would need to leave from, but he just can't decide. He's had a long day, he's tired and most of all wants to go home which he could decide to do right away. Given how tired and hungry he is, Walter can't decide to go and surprise his wife at the airport. But what he can decide to do is have a little sit down, eat a Golden Oreo, and see how he feels in a few minutes. If he did this, he would then be in a position to



*decide to go and surprise his wife. But having a sit down and a small snack is not a means to that end. Walter would be deciding to see how he felt, to give himself the option. He's fully aware that he might still decide to go straight home; indeed, further reasons for going home might occur to him while he's eating, such as him then having time to get the dinner started. Walter's deciding to sit down, therefore, doesn't settle anything.*

In most contexts, eating a biscuit is not connected to deciding to go and meet someone from the airport. But what I hope the **Walter** scenario makes clear is that such an alternative *could* be so connected. If Walter decides to sit down and eat a biscuit in order to keep his options open—that is, in order to be better able to make a decision that is not based mainly on his tiredness levels—then it is a robust alternative to his deciding to go straight home. Walter has options: he can decide to go home, or decide to sit down and take stock. The latter would make it possible for him to make a particular decision which at the start of the scenario he cannot make. For that reason, if Walter were to decide right away to go home, he has exercised a robust form of control and so could be morally responsible for so deciding, if it is appropriate to assess that decision morally. This is so even if, say, were Walter to have sat down to eat a biscuit he would have been beaten-up and robbed and thus rendered unable to go anywhere. Walter, Larry, Pete and Jones all have robust alternatives for this same reason.

#### 4 | THE HUNT AND SHABO TEST FOR ROBUSTNESS

In the previous section I argued that Pereboom's robustness criterion was too strong and I suggested an alternative criterion according to which Jones in *Tax Cut* does have a robust alternative. In putting forward this criterion I must contend with an argument for a certain class of robustness criteria that can be found in a test for robustness that has been put forward (independently) by David Hunt (2005) and Seth Shabo (2007) and then further developed in Hunt and Shabo (2013). If this test is acceptable, it will serve as an argument against my robustness criterion because it excludes as robust alternatives which my criterion counts as robust.

Hunt states the key thought behind the test as follows: "if ... an agent is morally responsible in virtue of being blameworthy, a PAP-relevant alternative should be such that, by accessing it, the agent would avoid (or mitigate) blame" (Hunt, 2005, p. 138). The affinity to Pereboom's understanding of exempting alternatives is obvious. And like Pereboom, Hunt and Shabo would agree that the alternative Jones has—deciding to imagine being punished—is plausibly a robust alternative for Jones's not imagining being punished. They would identify the crucial question as whether Jones's being able to decide to imagine being punished is a robust alternative for Jones's decision to vote for the tax cut. They address this question in two ways. First, drawing from Hunt's earlier paper, they present the following case (Hunt, 2005, p. 138): Jack is a smoker who wants to quit; he correctly believes that he can quit only by joining and completing a 12-step programme. Mack is Jack's smoking buddy who doesn't want Jack to quit and who would, if Jack joined the programme, slip Jack a pill to make his desire to smoke irresistible. In this scenario, Jack cannot decide to quit smoking, but he can join the 12-step programme, and that seems like it would be a robust alternative for Jack's continued smoking. Hunt and Shabo agree and suggest that this is because were Jack to join the programme he would be doing so *in order to quit*: in other words, he'd be taking the first step towards deciding to quit (and quitting) and so already trying to quit. So Jack's alternative is robust because it's a means to an end—namely, deciding to quit—which is incompatible with ("symmetrical to") his decision to continue smoking. One (relatively) uncontroversial way of avoiding or mitigating blame for A is by trying to do *not-A*, so this satisfies the key thought mentioned above. But, Hunt and Shabo stress, the same does not apply to Jones. Were Jones to decide to imagine being punished, it would not be a means to an end because it's stipulated that Jones must still decide what to do after having imagined being punished. So Jones's alternative is not robust.

While I agree that Hunt and Shabo's means-end condition on robustness is sufficient for robustness, I do not think it is necessary. For exactly the reasons outlined with respect to Pereboom's robustness criterion, I think it is too strong. It is enough for robustness that an agent conceives of herself simply as keeping her options open. To

develop my reply somewhat, I conceive of the decision that Jones in *Tax Cut* would make if he decided to imagine being punished as a means to making a better informed decision. It's not part of the means of deciding to vote against the tax cut. But in making the decision to imagine being punished, Jones is making a decision to keep an option open, he's making a decision to alter his motivational state slightly, with the knowledge that might then make it easier for him to vote against the tax cut, and he's making a decision to avoid committing himself there and then to voting for the tax cut. And that is enough to make it a robust alternative.

Now Hunt and Shabo agree that the means-end constraint as they state it is in need of more defence, and so they go on to develop their key thought in a second direction which doesn't appeal to the role of reasons. This is the idea that Jones's alternative would not "by itself eliminate or mitigate" Jones's blameworthiness for deciding to vote for the tax cut (Hunt & Shabo, 2013, p. 619). They write the following (I have adapted this to be applicable to Pereboom's *Tax Cut* FSC rather than Hunt's own *Revenge* case):

*Suppose that [the neuroscientist] is absent from the scenario, and that Jones ends up deciding on his own [to vote for the tax cut], but that before so deciding, he pauses to [imagine being punished]. If [imagining being punished] by itself eliminated or mitigated Jones's blameworthiness for deciding to [vote for the tax cut], Jones would at least be less blameworthy than he when he decides ... straightaway (Hunt & Shabo, 2013, p. 619).*

But, Hunt and Shabo state, if Jones were to imagine being punished *before* going ahead and voting for the tax cut anyway, this would not make any difference to his blameworthiness. So if Jones were to exploit the leeway that is available in the FSC in a normal (i.e. no intervener) case, where he goes on to vote for the tax cut anyway, it would make no difference to his blameworthiness. But then why think that this leeway counts as robust in the case where we do have an intervener? Or to put it another way, given that the defender of PAP maintains that Jones's alternative is robust, it should be legitimate to ask what would happen to Jones's blameworthiness if, in a case where there is no intervener, he were to exploit the available leeway before going ahead and deciding just how he does in the FSC. And Hunt and Shabo conclude that "since exploiting this alternative possibility in [the neuroscientist's] absence would not eliminate (or mitigate) Jones's blameworthiness for deciding as he does, this unexploited alternative doesn't plausibly ground Jones's blameworthiness in the actual sequence in [*Tax Cut*]" (Hunt & Shabo, 2013, p. 619).

It's important to be clear about how Hunt and Shabo's argument is supposed to work. They are not claiming that in order to be robust, we must be able to answer the following question positively: if there had been no intervener, would exploiting the leeway available to the agent in an FSC prior to then deciding in the same manner have mitigated or eliminated the agent's blameworthiness? That is, they are not saying that a necessary condition on an alternative being robust is that, if it were realised in the absence of the intervener, prior to the agent then deciding as she (in fact) does, the agent's blame would be lessened. But they are saying that if a positive answer cannot be given to that question, then the burden of explanation is on the defender of PAP to explain why.

However, it is telling that Hunt and Shabo don't offer this test as a necessary condition on robustness. They *do* not do so, because they *cannot* do so: to make this a necessary condition on robustness would prove far too much because it would mean that *deciding not to A* would not—at least in most cases—count as a robust alternative for *deciding to A*. As Hunt and Shabo themselves explain: suppose we consider a standard (non-buffer) Frankfurt-style case where the agent, who actually decides to do A, has the alternative of deciding to refrain from A-ing (suppose that the intervener would then intervene, causing the agent to decide to A). Hunt and Shabo agree that it will be "plausible" to think that deciding to refrain from A-ing is a robust alternative to deciding to A (this is something of an understatement: if anything counts as a robust alternative for deciding to A, surely it is deciding to refrain from A-ing). But now suppose the agent is not subject to an intervener, decides to refrain from A-ing, then changes her mind and decides to A. Is her blame lessened or even removed altogether? No, at least not in most ordinary cases. Making a decision to not-A rarely *ensures* (by itself) that an agent incurs no blame for a decision to A because it is (usually) open to an agent to change her mind and decide to A after all. So Hunt and Shabo cannot possibly hold that their robustness test is a necessary condition on robustness.

Yet they suggest that it is incumbent on the defender of PAP to explain just why an alternative is robust if it fails this test. And they maintain this even though the paradigm case of a robust alternative—a decision to refrain from A-ing—fails this test. Given that the paradigm robust alternative fails the test, I think it fair for the defender of PAP to ask for further details on why the test should be thought in any way plausible. This point is strengthened, I think, once we recognise that the PAP defender can and should see the alternatives that remain in cases like *Tax Cut* as decisions (as I have suggested above). It is, to be sure, not a decision to vote against the tax cut, and neither is it a decision to take the first step towards an end—voting against the tax cut—that Jones has settled on (indeed, if Jones had that end, he would have already made a decision to vote against the tax cut). Nevertheless, Jones's decision to imagine being punished is a decision to keep his options open, to potentially adjust his motivational state in the direction of that which he knows is right, as well as a decision to not commit there and then to voting for the tax cut. For those reasons Jones's decision is a robust alternative. This decision fails Hunt and Shabo's test for robustness, but so does the paradigm case of a robust alternative, so their test as currently framed should do nothing to trouble the defender of PAP.

## 5 | CONCLUSION

In this essay, I have investigated the notion of robust alternatives, that is, the question of when alternative possibilities per se explain an agent's responsibility. I presented several cases which provide good reason for thinking that Pereboom's robustness criterion is too strong, and I have developed a replacement principle which does justice to the cases presented. I have also offered a reply to Hunt and Shabo who have presented a test for robustness which is at odds with my criterion; I have suggested that further reason needs to be given why we should take their test seriously given that the paradigm case of a robust alternative fails the test, and demonstrated how the alternatives available to the agents in the so-called buffer-style Frankfurt cases, being the result of decisions, are closer the paradigm case than is sometimes realised.

## ENDNOTES

- <sup>1</sup> This literature is best thought of as an investigation into the nature of the ability to do otherwise rather than a body of work on the kind of cases Frankfurt highlighted (especially since much of the discussion concerns the metaphysics of control which Frankfurt hoped to avoid).
- <sup>2</sup> Frankfurt deliberately called this the Principle of *Alternate*—and not *Alternative*—Possibilities. This was because he thinks that the alternative options are merely second best: actions the agent would perform only had he not been able to do “what he clearly preferred to do” (Frankfurt, 1999, p. 372). Proponents of the Principle will reject that idea out of hand, so the apparently innocent name change is in fact non-incidentally.
- <sup>3</sup> Some call those who endorse the need for APs *leeway theorists*. I find this name tendentious as it gives the impression that such theorists are engaged in a futile attempt to squeeze control out of alternatives.
- <sup>4</sup> I would like to thank an anonymous referee for raising this point.

## ORCID

Simon Kittle  <http://orcid.org/0000-0002-6056-4970>

## REFERENCES

- Bennett, J. (1988). *Events and their names*. Clarendon Press.
- Fischer, J. M. (1982). Responsibility and control. *The Journal of Philosophy*, 79(1), 24–40. <https://doi.org/10.2307/2026344>
- Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy*, 66(3), 829–839.
- Frankfurt, H. G. (1999). Responses. *Journal of Ethics: An international. Philosophical Review*, 3(4), 367–372.
- Ginet, C. (1996). In defense of the principle of alternative possibilities: Why I don't find Frankfurt's argument convincing. *Noûs*, 30, 403–417.

- Hunt, D. P. (2005). Moral responsibility and buffered alternatives. *Midwest Studies in Philosophy*, 29(1), 126–145. <https://doi.org/10.1111/j.1475-4975.2005.00109.x>
- Hunt, D. P., & Shabo, S. (2013). Frankfurt cases and the (in) significance of timing: A defense of the buffering strategy. *Philosophical Studies*, 164(3), 599–622. <https://doi.org/10.1007/s11098-012-9874-y>
- Kane, R. H. (1985). *Free will & values*. State University of New York Press.
- Moya, C. (2006). *Moral responsibility: The ways of scepticism* (1st ed.). London: Routledge.
- Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.
- Pereboom, D. (2012). Frankfurt examples, derivative responsibility, and the timing objection. *Philosophical Issues*, 22(1), 298–315. <https://doi.org/10.1111/j.1533-6077.2012.00231.x>
- Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford: Oxford University Press.
- Sartorio, C. (2016). *Causation and free will*. Oxford, UK: Oxford University Press.
- Shabo, S. (2007). Flickers of freedom and modes of action: A reply to Timpe. *Philosophia*, 35(1), 63–74. <https://doi.org/10.1007/s11406-007-9050-3>
- Widerker, D. (2003). Libertarianism and Frankfurt's attack on the principle of alternative possibilities. In G. Watson (Ed.), *Free will* (2nd ed.). Oxford University Press.

**How to cite this article:** Kittle S. When is an alternative possibility robust? *Eur J Philos*. 2018;1–12. <https://doi.org/10.1111/ejop.12369>