

# Robustness and Up-to-us-ness

**Simon Kittle**

University of Innsbruck

BIBLID [0873-626X (2017) 44; pp. 35–57]

## **Abstract**

Frankfurt-style cases purport to show that an agent can be morally responsible for an action despite not having any alternatives. Some critics have responded by highlighting various alternatives that remain in the cases presented, while Frankfurtians have objected that such alternatives are typically not capable of grounding responsibility. In this essay I address the recent suggestion by Seth Shabo that only alternatives associated with the ‘up to us’ locution ground moral responsibility. I distinguish a number of kinds of ability, suggest which kinds of abilities ground the truth of the ‘up to us’ locution, and outline how these distinctions apply to the indeterministic buffer cases.

## **Keywords**

Free will, control, alternative possibilities, Frankfurt-style cases, up to us

## 1 Introduction

Frankfurt-style cases purport to show that an agent can be morally responsible for an action despite not having any alternatives. Some critics have responded by highlighting various alternatives that remain in the cases presented, and Frankfurtians have in their turn responded that such alternatives are not *robust*—not capable of grounding responsibility. In this essay I distinguish two different kinds of ‘up to us’ statement which, I argue, are made true by different kinds of ability. I articulate a number of different concepts of ability and ask which ones ground the truth of the ‘up to us’ locution when applied to actions, and in the process I address the recent suggestion by Seth Shabo that only alternatives associated with the ‘up to us’ locution ground moral responsibility. Using this framework I then assess what I take to be the strongest Frankfurt-style cases, the indeterministic buffer cases, showing how they fail.

*Disputatio*, Vol. IX, No. 44, May 2017

Received: 16/06/2016 Revised: 04/11/2016 Accepted: 08/12/2016

## 2 Shabo on alternatives and robustness

Frankfurt-style cases (FSCs) are scenarios where an agent, typically Jones, performs some action “on his own”, while being monitored by an external agent who is capable of manipulating the subject’s brain so as to cause him to decide in a particular way. The external agent does not, in fact, affect the agent, but is said to be able to detect whether intervention is needed, and thereby leave Jones unable to do anything different. FSCs are thus taken to be counter-examples to the Principle of Alternative Possibilities (PAP), which states that an agent can be morally responsible for an action only if the agent could have done otherwise. One response to the FSCs is to argue that, despite initial appearances, Jones does in fact have alternatives. Because some alternatives are clearly irrelevant to responsibility the challenge is to find alternatives that are capable of grounding responsibility. I will follow John Martin Fischer (1994: 142) in calling such alternatives *robust*. The following scenario, which for reasons that will become apparent I’ve called **No Cup**, illustrates this point:

**(No Cup)** Jones intentionally detonates an explosive device by keying in the activation code on his cell phone. He was being monitored by Black, who wanted him to detonate the device and would have intervened if need be (had Jones been going to decide to refrain from detonating the device, he would have twitched, and Black was going to use this fact to intervene). Unknown to Black and Jones, however, there was a 1% objective probability that the low rumbles of an overhead plane would cause the roof of Jones’s office to cave in. Had this happened Jones would have fled and made no decision concerning the explosive device.

Assume for the time being that this is a successful Frankfurt-style case: Black removes all the alternatives that might have been thought to be relevant to Jones’s responsibility for detonating the device. The current point is just that the small probability of Jones’s office collapsing introduces a genuine alternative possibility, but one that is obviously irrelevant to any responsibility Jones might bear for detonating the device. Thus, not all alternatives are relevant. This much is agreed upon by all parties.

What we need then is a robustness criterion which spells out

when alternative possibilities are relevant to responsibility. One suggestion, which would explain the case above, is that robust alternatives will be those associated with an agent's *abilities*. Shabo (2014: 384) articulates the distinction between these kinds of alternatives as follows:

**(Mere alternative)** Consistent with the past and the laws of nature, it's possible that the agent performs a different action (or refrains from acting), but the agent lacks the ability to realize this possibility; the past and the laws preclude this possibility from being realized except as a result of circumstances the agent doesn't control, or with respect to which her control is markedly impoverished.

**(Enabling alternative)** Consistent with the past and the laws, it's possible that the agent performs a different action (or refrains from acting), and, further, she has the power or ability or to realize this possibility.

Shabo says that while it has long been recognised that **mere alternatives** do not satisfy any interesting avoidability requirement, it is often thought that **enabling alternatives** do. Shabo argues that this latter point is not straightforward. He aims to illustrate why with the following example, which I will call **Original Cup**:

**(Original Cup)** Jones intentionally detonates an explosive device by keying in the activation code on his cell phone, while (incidentally) resisting an urge to take a sip of coffee until he has completed the code. He was being monitored by Black, who wanted him to detonate the device, and who would have intervened if need be. Since Black has determined that whether or not Jones takes a sip of coffee during this interval reveals nothing about his intentions, his taking a sip would not have triggered Black's intervention. Unknown to Black and Jones, however, the coffee contains a fast-acting neurotoxin that would have paralyzed Jones before he had a chance to complete the code (Shabo 2014: 386).

Here, just as with **No Cup**, there is an alternative possibility where Jones does not detonate the explosive: if he sips the coffee, he'll be immediately paralysed. But in **Original Cup**, in contrast to **No Cup**, Jones is—in some sense—able to realise the alternative

where he bears no responsibility (by taking a sip of coffee). Shabo's point is that this is not enough. He rightly points out that behind the demand for alternatives lies the demand for control. It is typically supposed that if an agent is able to do otherwise then that agent has the distinctive kind of control required, but **Original Cup** shows that not every sense in which an agent might realise an alternative bestows this control. Shabo suggests that what's important is that the action be *up to* the agent. The connection between 'up to us' and responsibility goes all the way back to Aristotle and has been discussed at various points in the literature on Frankfurt-style cases (Frankfurt 1969: 836, Naylor 1984, Alvarez 2009: 64).

How can it be up to Jones whether he sips the coffee but not up to him whether he detonates the device, given that sipping his coffee would have prevented him from detonating the device? According to Shabo, the crucial thing to recognise is that the 'up to us' locution introduces an *intensional context* (Shabo 2014: 379). Typically, an intensional context is one where the substitution of co-referring terms does not preserve truth. To use Shabo's example: Rob is the masked man, and Peter believes that the masked man robbed the poor box. But we cannot infer from this that Peter believes that Rob robbed the poor box because 'believe' introduces an intensional context (Shabo 2014: 379). The intensional context introduced by the 'up to us' locution explains why it is possible that it is up to Jones whether he takes a sip of coffee, and that his taking a sip of coffee would lead to him not being blameworthy, and yet it *not* be the case that it was up to Jones whether or not he was blameworthy. A similar thing is true when Black's intervention is involved, as the following case shows:

**(Fresh Cup)** Suppose that Jones's coffee isn't poisoned but that Black knows that Jones's deciding to take a sip of coffee is a sure sign that he is having second thoughts about detonating the device. If Jones takes a sip of coffee, Black will intervene and force Jones to complete the activation code on time (Shabo 2014: 387).

As before, it's up to Jones whether he takes a sip of coffee. But although Jones's taking a sip of coffee will result in him being blameless, it is not up to him whether he is blameless. On the basis of these observations Shabo makes the following distinction between enabling alternatives (Shabo 2014: 385):

**(Non-robust enabling alternative)** Consistent with the past and the laws, it's possible that the agent performs a different action (or refrains from acting), and, further, she has the power or ability to realize this possibility. Even so, the agent lacks a robust alternative possibility, one that could plausibly ground her moral responsibility for what she does, because it isn't up to her whether this possibility is realized.

**(Robust enabling alternative)** As above, except the ability is such that, in virtue of possessing it, it is up to her whether this alternative possibility is realized. Such an agent has a robust alternative possibility, one that could plausibly ground her moral responsibility for what she does.

Shabo goes on to connect the 'up to us' locution to the epistemic component included in recent formulations of robustness criteria. Here is a recent robustness criterion from Derk Pereboom which includes an epistemic element:

For [an] agent to have a robust alternative to her immoral action A ... it must be that

- (a) she instead could have voluntarily acted or refrained from acting as a result of which she would be blameless, and
- (b) that for at least one such acting or refraining, she is cognitively sensitive to its being available to her, with the result that she believes to some significant degree that had she voluntarily so acted or refrained she would be, or would likely be, blameless (Pereboom 2012: 301).

Shabo contends that the need for "the epistemic dimension added by clause (b) is a consequence of the fact that an alternative possibility is robust only if it's up to the agent whether that possibility is realized; in short, this epistemic dimension is implicit in this kind of power attribution" (Shabo 2014: 395). Shabo appears to be saying here that the epistemic criterion<sup>1</sup> needs to be satisfied in order

<sup>1</sup> The relevant condition is frequently referred to as an *epistemic* condition even though it is widely accepted that it does not require *knowledge*. I will follow this terminology.

for the *power attribution* most relevant to free will to be attributable. That is, the kind of condition identified by Pereboom in the above robustness criterion belongs to the *power* or *ability* that is relevant to free will; it does not belong, for example, to a separate “epistemic condition on moral responsibility.” As we might put it: knowledge is power—or, control.

This is a departure from the view which separates the freedom or control condition on moral responsibility from the epistemic condition on moral responsibility—a view which, if not the orthodox position, is certainly one significant strand in contemporary writing. Such a view is widely endorsed by those on both sides of the debate over the compatibility of moral responsibility and determinism (See Fischer and Ravizza 1998, Ginet 2000, Timpe 2011). The following kind of example from Carl Ginet motivates the separation of epistemic and freedom conditions:

**(Simon)** Simon enters the hotel room he has just checked into and flips what appears to be, and what he takes to be, an ordinary light switch, but, to his surprise and consternation, the flipping of the switch sets off a loud fire alarm (Ginet 2000: 269).

On this view, Simon doesn’t bear any moral responsibility for setting off the fire alarm because he didn’t know he was setting off the fire alarm. Simon is in control of flipping the switch, and because flipping the switch results in the fire alarm going off, Simon is also in control of setting off the fire alarm. If, however, the relevant power attribution includes an epistemic component in virtue of which it introduces an intensional context then this conclusion does not follow. Alfred Mele (2010) has recently presented a number of arguments for thinking that what is typically treated as a separate epistemic condition on moral responsibility may in fact be an epistemic aspect of what is required in order to have the control associated with free will. And prior to both Shabo and Mele, Alvin Goldman (1970, 1972) and Peter Morriss (1987) made similar distinctions which they took to be helpful when discussing control. It is this approach—treating these epistemic requirements as relevant to the agent’s control—that I will pursue in the following section via a discussion of the ‘up to us’ locution. Although this locution is often used in discussions of free will it is not often discussed in depth.

In the following section I seek to answer two questions: first, what kind of abilities ground the truth of the ‘up to us’ locution? Second, are the abilities associated with the ‘up to us’ locution necessary for moral responsibility?

### 3 Up-to-us-ness and abilities

#### 3.1 The ‘up to us’ locution

Shabo suggests that the power attributions relevant to free will are those which ground the truth of the ‘up to us’ locution. Here we must be careful, however, because just as there are many different kinds of ability, so the ‘up to us’ locution can be used in different ways. In particular, it can take different kinds of complement. Sometimes it is used with actions:

- (1) It’s up to me whether or not *I turn on* the light.
- (2) It’s up to me whether or not *I invite you* to the party.

Other uses show that the complement may be filled in with a non-actional state of affairs:

- (3) It’s up to me whether or not the stone is on the grass.
- (4) It’s up to me whether or not the stream is blocked.

Both are acceptable but it is crucial to distinguish between them because the contention of those who defend the PAP concerns primarily the agent’s control *over his or her actions*. It is not a point about the control an agent has over arbitrary, non-actional states of affairs, as might be thought if statements such as (1) and (2) were confused with those like (3) and (4). The main reason for this is an asymmetry that exists between actions and (non-actional) states of affairs. Take a proposition  $P$  which asserts the occurrence of some dated state of affairs. Once the relevant time comes it must be the case that either  $P$  is true or  $\sim P$  is true. And this means that when the ‘up to us’ locution takes a non-actional state of affairs as its complement it requires the agent to have the power to render  $P$  true and also the power to

render  $\sim P$  true. For (3) to be true, for example, I need to be able to bring it about that the stone is on the grass and able to bring it about that the stone is not on the grass. With (4), I need to be able to block the stream and be able to unblock the stream. When the ‘up to us’ locution has a non-actional complement, then, it requires that the agent have the ability to bring about and the ability to prevent an event of the very same type.

This is not the case when the ‘up to us’ locution has an action as its complement. The natural way of reading (1), for example, is that I’m able to turn on the light and also able to *refrain* from turning on the light. In other words, the negative side of the locution—what the ‘not’ in the ‘whether or not’ attaches to—requires that the agent be able to refrain from something. But to refrain from something is not equivalent to the negation of the state of affairs of the agent’s performing the action. Let  $P$  be the proposition *that I turn on the light*. The negation of that proposition is *it is not the case that I turn on the light*. But it’s not being the case that I turn on the light does not entail that I refrained from turning on the light. If I drop dead before even thinking about the light then the statement ‘it is not the case that I turned on the light’ is true but I didn’t refrain from turning on the light. Ezio Di Nucci (2011: 119, 124) has discussed how this distinction is important for zeroing in on what PAP is about; part of my purpose here is to place this distinction in the larger context of the workings of the ‘up to us’ locution.

To summarise: in statements (1) and (2) the ‘not’ in ‘whether or not’ does not generate the negation of the locution’s complement, but a proposition describing a refraining. In (3) and (4), however, the ‘not’ produces the negation of the proposition which is the complement to ‘up to us’. To put it somewhat differently: a proposition affirming the occurrence of a non-actional, dated state of affairs and the negation of that proposition are contradictories; a proposition affirming the performance of an action and its “negation” are contraries.

This asymmetry is hugely important. The dual-power or ability that the defender of PAP—the leeway incompatibilist—thinks is important is not, as might be suggested by (3) and (4), the power to bring about non-actional state of affairs  $P$  and the power to bring about non-actional state of affairs  $\sim P$ . Now, it’s common enough for



agents to have such powers. I might have the power to bring about the stone's being on the grass and the power to bring about the stone's not being on the grass. But this kind of power, common as it might be, is stronger than what the leeway incompatibilist requires. The leeway incompatibilist's demand for avoidability only need concern actions. It is the conflation of these two sets of abilities, I suggest, that leads people to find Frankfurt-style cases in any way plausible, and in section 4 I'll show how being aware of it helps us to see where the buffer cases go wrong. In the remainder of this section I will only be concerned with the 'up to us' locution as it appears in (1) and (2).

There is a further complication that needs addressing before we can continue. Shabo claims that the abilities which ground the truth of the 'up to us' locution include an epistemic component and that as a result the locution *introduces an intensional context*. I agree with the former point, but the latter point is not straightforward. As we've seen, intensional contexts are standardly defined as those contexts where the substitution of co-referring terms does not preserve truth. But with respect to the 'up to us' locution, what is it that the co-referring terms refer to? The obvious answer is: actions. But here we encounter a difficulty, for whether different descriptions of what someone does *ever* co-refer depends on how one individuates actions. For example, on Goldman's (1970) view actions are exemplifications of properties at times. Moreover, Goldman adopts a fine-grained view of property individuation and his comments make it clear that he would consider *taking a sip of coffee* to be a different property to *rendering oneself morally blameless*. Therefore, these descriptions of what Jones does will never co-refer.

I will bypass this complication by adopting a coarse-grained theory of action individuation such as that found in Anscombe 1963. Such a view of action will make the idea easier to apply: 'Jones's taking a sip of coffee' may indeed refer to what 'Jones's rendering himself morally blameless' refers to, so we can accept Shabo's characterisation of things in terms of intensional contexts.

### 3.2 Abilities: non-intentional vs. doxastic, weak vs. reliable

In order to answer the question of which kind of abilities ground the truth of the 'up to us' locution I will outline a few different notions of

ability. Each kind of ability outlined below should be understood as an intrinsic ability that is *maximally specific*, that is, fully specified with respect to the circumstances referenced. I also assume here that to have free will the agent will need an opportunity to exercise the relevant ability. I have argued for both these claims at length in Kittle 2015.

The first kind of ability to consider is, plausibly, that behind non-robust enabling alternatives; I shall label it *non-intentional ability*:

**(Non-intentional ability)** An agent *S* is non-intentionally able to *A* in circumstances *X* if and only if

- (i) *S* is able to intentionally *B* in circumstances *X*, and
- (ii) *S*'s intentionally *B*-ing in *X* would be his *A*-ing.

A few words of explanation are in order. First, this is evidently not a reductive analysis: clause (i) of the explanans requires the agent to *be able* to intentionally *B*. Moreover, because clause (i) requires that the agent be able to *intentionally B*, the account is parasitic on an account of what it is to be able to do something intentionally. Second, **non-intentional ability** defines an ability property in terms of an action *and* a set of circumstances, *X*, even though it is a kind of intrinsic ability. In other words, it is the ability-to-*A*-in-*X* that is possessed rather than merely the ability-to-*A*. Again, I have argued for this in Kittle 2015.

**Non-intentional ability** is a very broad notion of ability. If the agent can intentionally *B*, then the agent will have the non-intentional ability to do anything that would result from his *B*-ing. Here's an example: suppose that Phillip flips a light switch and as a result the tungsten filament heats up to its standard operating temperature of around 2500 degrees C. Phillip doesn't know this—he has no idea the filament is made of tungsten. Still, Phillip is non-intentionally able to raise the temperature of the filament to 2500 degrees C. Similarly for all the causal consequences which *would follow reliably* from his flipping the switch—the reliability condition is built into clause (ii). Jones, in the **Fresh Cup** example, has the non-intentional ability to prevent himself from being blameworthy. This explains the sense in which we can affirm that Jones *is able* to prevent himself from being blameworthy. But it also explains why that isn't much

help, and doesn't render the alternative robust: non-intentional abilities bestow little, if any, control.

Now consider the following kind of ability:

**(Doxastic ability)** An agent *S* is doxastically able to *A* in *X* if and only if

- (i) *S* is able to intentionally *B* in circumstances *X*, and
- (ii) *S*'s intentionally *B*-ing in *X* would be his *A*-ing, and
- (iii) *S* is "cognitively sensitive" to the fact that that by *B*-ing he would be *A*-ing.

By using the phrase 'cognitively sensitive' I intend to draw on Pereboom's recent work on the robustness criteria (Pereboom 2009, 2012). I therefore use the phrase as a placeholder for the precise conditions which the agent needs to satisfy in order to be 'cognitively sensitive'. Moreover, it is expected that those conditions will be just those conditions which are usually discussed under the banner of "the epistemic conditions on moral responsibility" (See Ginet 2000: 270ff and Timpe 2011: 18). It is not my purpose here to give an account of such a set of conditions but rather to provide additional support for Mele's suggestion that such conditions might be best thought of as part of the control condition and then, given that background, to discuss the 'up to us' locution and the latest Frankfurt-style cases. Why think that the 'cognitive sensitivity' condition belongs to the control condition? Consider the following example:

**(Trapped Tom)** Tom is in a building which is currently ablaze. He has only one route of escape not blocked by the flames, but the route is blocked by a locked door protected by an electronic keypad. The combination is '453322' but Tom does not know this.

Tom has the non-intentional ability to escape. That's because he is able to intentionally type '453322' into the keypad, and if he did so, the door would open and he could get to safety. Is Tom in control of whether he escapes? According to some of those who endorse a strict separation of the control and epistemic conditions on responsibility he is. But this is implausible. Suppose that Tom's brother Tim is

trapped in a different corridor also protected by a locked door with a keypad. Tim is unfortunate, but not as unfortunate as Tom, because Tim knows the combination to the door which blocks his way. Intuitively, Tim has more control than Tom: it is up to Tim whether he escapes but it is not up to Tom whether he escapes. The account of doxastic ability can explain this: Tim has but Tom lacks the doxastic ability to escape.

The two notions of ability so far articulated require that there be a reliable connection between the agent's *B*-ing and the agent's *A*-ing (indicated by the 'would' in clause (ii)). Sometimes, however, it is only the case that an agent's *B*-ing *might* be his *A*-ing. If the agent knows this, would that be enough for control? In the following section I will suggest that sometimes the answer is yes. I will therefore define two further notions of ability which parallel the two outlined above but which replace the occurrence of 'would' in each of the clauses (ii) and (iii) with a 'might.' I will call these *weak abilities*:<sup>2</sup> **weak non-intentional ability** (though we will have little use for this) and **weak doxastic ability**. I will not write both definitions out but for illustrative purposes the definition for weak doxastic ability is as follows:

(**Weak doxastic ability**) An agent *S* is doxastically able to *A* in *X* if and only if

- (i) *S* is able to intentionally *B* in *X*, and
- (ii) *S*'s intentionally *B*-ing in *X* **might** be his *A*-ing, and
- (iii) *S* is "cognitively sensitive" to the fact that that by *B*-ing he **might** be *A*-ing.

With that framework in place, I will now seek to answer the questions posed at the close of section 2: first, what kind of abilities ground the truth of the 'up to us' locution? Second, are the abilities associated with the 'up to us' locution necessary for moral responsibility?

<sup>2</sup> Of course, the singling out of 'reliable' and 'weak' abilities in this way is a simplification; the reliability of abilities is best thought of as a spectrum.

### 3.3 Abilities and up to us

Pereboom and Shabo each require that an agent be able to bring about the alternative *reliably*; witness Pereboom's use of "would or would likely be" in his robustness criterion. This suggests that what Pereboom and Shabo want to endorse is the idea that the 'up to us' locution should be understood as involving **reliable doxastic abilities**:

**(Up to us)** It is up to the agent whether or not she *A*-s in circumstances *X* if and only if:

- (i) The agent has the reliable doxastic ability to *A* in *X*, and
- (ii) The agent has the reliable doxastic ability to refrain from *A*-ing in *X*.

In fact, however, neither is committed to this. This is because in the literature on Frankfurt-style cases the robustness criterion is typically formulated only with reference to the "alternative sequence". That is, it is assumed that the agent has performed an action and then it is asked what it takes for the alternative to be robust. On this retrospective approach nothing is said about the kind of ability that the agent actually exercised in producing her behaviour; the focus is entirely on the kind of ability the agent needed to have to realise the alternative possibility. So while Pereboom and Shabo are committed to something like clause (ii) they are not committed to anything like clause (i).

This is fortunate because I want to suggest that when it comes to the control associated with free will (i) is too strong. If this is right, then either **Up to us** is an incorrect characterisation of the 'up to us' locution or an agent can freely *A* without *A* being up to the agent. Here is an example, adapted from Robert Kane (1996: 55), which motivates the idea that (i) is too strong:

A plant worker places some radioactive material in his boss's office with the intention of killing him. Over a period of time, before being discovered, the radioactive material emits enough radiation into the worker's boss in order to produce a fatal cancer. It was genuinely indeterminate whether the material would emit enough radiation to give the worker's boss cancer.

Among the plant worker's abilities are the following:

The reliable doxastic ability to place the material in his boss's office

The reliable doxastic ability to refrain from placing the material in his boss's office

Of course, the plant worker is responsible not just for putting some radioactive material in his boss's office but also for killing his boss: that is what he intended to do, and that is what he did. What kind of ability did the plant worker have to do this? I think the answer is the following:

The weak doxastic ability to kill his boss

This is because the worker can do something which might—but only might—result in his boss's death. If this is right, then the worker has the ability to kill his boss and the ability to refrain from killing his boss but these are different kind of abilities: his *doxastic ability to kill his boss* is weak while his *doxastic ability to refrain from killing his boss* is reliable.

Is it up to the worker whether he kills his boss? No. It would be strange to say that it was up to the worker whether he killed his boss and that is because the 'up to us' locution does indeed require a certain level of reliability. More precisely, it suggests that clause (i) of **Up to us** is on the right lines. This might be challenged in the following way. Though it's somewhat intuitive that it's not up to the plant worker whether he kills his boss, nevertheless it clearly is up to the plant worker whether he refrains from killing his boss. He can ensure that he refrains from doing so. And it might be thought that this is because the following statements are true:

The worker has the reliable doxastic ability to refrain from killing his boss

The worker has the weak doxastic ability to kill his boss

This, it might be thought, speaks in favour of requiring a reliability condition on only clause (i) of the account of 'up to us'. But this is too quick and would be to make a mistake similar to that made by those who conflate dual-power over non-actional, dated states of af-

fairs with dual-power over actions. If the complement to ‘up to us’ is *he refrains from killing his boss*, then the second ability which **Up to us** requires the worker to have will be the *reliable doxastic ability to refrain from refraining to kill his boss*. And this, I think, is an ability the worker does have because he has the reliable ability to *try* to kill his boss. This is an interesting point about the ‘up to us’ locution: the phrase ‘it’s up to me whether or not I *A*’ requires that I be able to reliably refrain from *A*-ing, but to refrain from refraining to *A* does not require being able to reliably *A*, so the truth of the phrases ‘it’s up to me whether I *A*’ and ‘it’s up to me whether I refrain from *A*-ing’ are not grounded by the same abilities—they both require the ability to reliably refrain from *A*-ing, but the second ability in each case differs.

Now, if it’s not up to the plant worker whether he kills his boss, yet he is morally responsible for it, then the control which grounds that responsibility does not require the control required by the ‘up to us’ locution. I offer the following as a tentative proposal for the kind of control which is capable of grounding responsibility:

**(Freely able)** An agent is able to *A* freely in circumstances *X* if

- (i) The agent has the weak doxastic ability to *A* in *X*, and
- (ii) The agent has the reliable doxastic ability to refrain from *A*-ing in *X*.

On this account, when an agent is freely able to *A* it will not (necessarily) be up to the agent whether he *A*-s, but it will be up to the agent whether he refrains from *A*-ing. Clause (i) is subject to an ‘if’ but not an ‘only if’ because possessing a stronger ability to *A* would also secure the required control. (Recall too that my discussion has been limited to intrinsic abilities; the agent also needs the opportunity to *A* in *X*.) If correct, this account would (a) explain the close association between alternative possibilities and ‘up to us’, (b) retain the reliability criterion on ‘up to us’, and (c) explain the agent’s responsibility in cases like that of the plant worker. In the final section I will show how to apply this account to Frankfurt-style cases.

## 4 The mixed response to indeterministic Frankfurt-style cases

### 4.1 The leeway incompatibilist's task

If we accept some kind of doxastic condition on control then finding a robust alternative is not just a matter of finding an enabling alternative, we need to find the right kind of enabling alternative. Shabo thinks that this makes it harder for the leeway incompatibilist to respond to Frankfurt-style cases. In particular, Shabo argues that once we recognise this doxastic condition it will be all the more evident that cases like **Fresh Cup** are counterexamples to the Principle of Alternative Possibilities.

I will show why Shabo's contention is mistaken, but it is important to focus on the right Frankfurt-style cases. The leeway incompatibilist in fact needs to say very little about **Fresh Cup** because it is an illegitimate case: it employs a prior sign which Shabo takes to be a *reliable indicator* of Jones's decision and behaviour. But the leeway incompatibilist will reject the idea that there could be such a *reliable indicator* of a free decision. This is because **Fresh Cup** falls prey to what is known as the Dilemma Defence and which runs as follows: either the prior sign that indicates what Jones will do is deterministically connected to his decision or not. If it is, then the leeway incompatibilist will be under no obligation to agree that Jones is responsible; if it isn't, then Black cannot expunge all of Jones's alternatives.<sup>3</sup> The **No Cup** and **Original Cup** cases also succumb to the Dilemma.

The only cases the leeway incompatibilist needs to address are those cases designed to avoid this dilemma. I think the so-called *buffer cases* are the strongest cases in this category and will address them below. In pursuing this strategy I am suggesting that the Dilemma Defence and the so-called Flicker Defence—the pointing to “flickers” which are capable of grounding the agent's responsibility (Fischer 1994)—should be deployed together. This strategy is not *ad hoc* because the Dilemma Defence simply forces the Frankfurtian to provide a case which does not prejudge the very question at issue.

<sup>3</sup> David Widerker's (1995) is a useful formulation of the Dilemma Defence.



#### 4.2 The indeterministic buffer cases: a mixed response

The buffer cases have the following structure. Rather than employ a prior sign which operates as a reliable indicator of what the victim in the Frankfurt-style case will do, they employ a prior sign which is a necessary but not sufficient condition for the agent acting in some way. Thus, the agent cannot perform the alternative action unless the prior sign occurs, but after the prior sign occurs the agent still needs to make a decision to perform that alternative action. Pereboom, one of the first to present a buffer case, has offered the following example:

**(Tax Evasion (2))** Joe is considering claiming a tax deduction for the registration fee that he paid when he bought a house. He knows that claiming this deduction is illegal, but that he probably won't be caught, and that if he were, he could convincingly plead ignorance. Suppose he has a strong but not always overriding desire to advance his self-interest regardless of its cost to others and even if it involves illegal activity. In addition, the only way that in this situation he could fail to choose to evade taxes is for moral reasons, of which he is aware. He could not, for example, choose to evade taxes for no reason or simply on a whim. Moreover, it is causally necessary for his failing to choose to evade taxes in this situation that he attain a certain level of attentiveness to moral reasons. Joe can secure this level of attentiveness voluntarily. However, his attaining this level of attentiveness is not causally sufficient for his failing to choose to evade taxes. If he were to attain this level of attentiveness, he could, exercising his libertarian free will, either choose to evade taxes or refrain from so choosing (without the intervener's device in place). However, to ensure that he will choose to evade taxes, a neuroscientist has, unbeknownst to Joe, implanted a device in his brain, which, were it to sense the requisite level of attentiveness, would electronically stimulate the right neural centers so as to inevitably result in his making this choice. As it happens, Joe does not attain this level of attentiveness to his moral reasons, and he chooses to evade taxes on his own, while the device remains idle (Pereboom 2012: 302–3).

In this example Joe has to decide whether to claim an illegal tax

deduction. The necessary but not sufficient condition for his refraining is that he considers the moral reasons against evading taxes. But Joe might consider the moral reasons and—if there were no Frankfurtian intervener—still go on to evade taxes. It is this feature that supposedly circumvents the force of the Dilemma Defence because there is no way it can be argued that the prior sign deterministically causes Joe’s decision to evade taxes. However, given the availability of this prior sign it seems that the neuroscientist is able to arrange things such that Joe has no robust alternatives. The alternative which does remain—Joe’s reaching the required level of attentiveness to moral reasons, and subsequently having his brain electrically stimulated by the neuroscientist’s device to choose to evade taxes—is, Pereboom argues, not robust. This is because although the alternative is one where Joe doesn’t bear any moral responsibility, Joe himself does not realise this.

At this point, leeway incompatibilists might wonder whether they could accept the two intuitions that we are supposed to have with respect to Joe and explain Joe’s moral responsibility in terms of *derivative responsibility*. When an agent is derivatively responsible, the agent is responsible for some unavoidable action because there was some prior time at which the agent could have acted so as to not subsequently face an unavoidable action. The paradigm examples are drunk driver cases: Derek drives drunk and hits someone at  $t_2$ ; he could not have done any different at  $t_2$  because his drunken state drastically reduced the control he had; still, he could, at  $t_1$ , have decided not to drink, or given his keys to a friend, and so on.

David Widerker has applied this idea to Joe’s case, suggesting that Joe is “derivatively blameworthy for the decision he made, because he has not done his reasonable best ... to avoid making it” (Widerker 2006: 173). Granted, Joe couldn’t have decided to not evade taxes. But Joe could have considered moral reasons and this was something *he should have done*. Moreover, if Joe had been more attentive to moral reasons, he would have been “forced by the neuroscientist to [decide to evade taxes]” and so would not himself be responsible for his decision (Widerker 2006: 173). So Joe’s failure to consider moral reason makes him derivatively responsible for his subsequent decision to evade taxes.

The Frankfurtian can easily avoid Widerker’s objection, however,

by amending the example to remove the feature on which it relies, namely, that the available option is something the agent *should* do. Pereboom (2012: 307) has subsequently pursued this line of thought, proposing instead that in order to be able to make the free choice to decide not to evade taxes Joe has to, say, imagine being severely punished. Given that imagining being severely punished is not something agents have a moral obligation to do when deliberating, Widerker's derivative responsibility objection cannot get started.

Although useful this amendment is, as Pereboom (2012: 307) says, not needed. Widerker's appeal to derivative responsibility gains illicit plausibility from the fact that we would ordinarily assume that Joe would think that his considering moral reasons might lead him to decide not to evade taxes. That is, we ordinarily think that one way of deciding not to do something is by attending to the moral reasons there against doing it. Attending to the moral reasons against evading taxes is, in other words, plausibly thought to be a means to the end of deciding not to evade taxes. But Pereboom is clear that Joe does not know that considering moral reasons is a means to the end in question. Pereboom's amendment is useful because by making the prior mental action something that is a little unusual, we are not tempted to think that Joe will see it is a means to the end.

Pereboom is correct in his evaluation of Widerker's derivative responsibility objection; however, this does not show that appeals to derivative responsibility are out of place. It simply shows that Widerker was off target in his choice of what to trace back to in order to ground Joe's responsibility. Widerker's attempt to trace back to Joe's failure to consider moral reasons is a bit like attempting to explain a drunk driver's responsibility for running someone down by tracing back to his decision to get into his car. By the time the driver "decides" to get into his car, it's already too late: he's already lost control. To explain the driver's responsibility, we need to trace back to a time when the driver had the requisite control, i.e., to a time when he wasn't drunk.

One thing to recognise at this point is that we need to appeal to derivative responsibility *only if Joe doesn't know* that the necessary but not sufficient prior sign is a means to the end of not evading taxes. There is where the mixed nature of the response comes in. If the

agent in a buffer case knows<sup>4</sup> that the necessary but not sufficient prior sign is a means to an end where he would be blameless, then we can affirm that the agent has the requisite control. This is one reason why it is crucial to recognise that the cognitive conditions on robustness belong in an account of ability and not on the epistemic condition on moral responsibility: the agent's knowledge affects his control.

In other words, if Joe had the doxastic ability to claim the illegal tax deduction, and he had the reliable doxastic ability to refrain from evading taxes (which he would have, if he knew that attending to moral reasons was the means to that end), then Joe would have the requisite control and the example would be no problem. Before moving on to the scenario where Joe doesn't have that knowledge I want to offer a few comments regarding the case if we assume that Joe does have this knowledge. I have just stated that Joe would have the reliable doxastic ability to refrain from evading taxes. It might be objected that this is patently false: it's inevitable that Joe evades taxes in the scenario, because if he attends to moral reasons, the neuroscientist will begin his intervention. It is at this point that the distinction introduced in section 3.2 becomes vital. It is *actions* that need to be avoidable, not non-actional, dated states of affairs. An agent will bear responsibility for a particular decision only if that particular decision is avoidable. And the decision to evade taxes for which Joe is responsible was avoidable. Moreover, on the current assumptions the alternative is robust: Joe knew just what he had to do to avoid making the decision that he in fact made.

Now, the Frankfurtian might object here, complaining that there is an inevitable state of affairs concerning Joe's decision—that *Joe decides to evade taxes*—which shows that the decision wasn't avoidable. Note that this is only the case if we allow that a person can *be caused to decide*, and, moreover, *caused to decide in a particular way*. Many leeway incompatibilists would reject (at least) the latter idea. But in any case, if we clearly distinguish between the performance of a particular decision and a state of affairs which affirms that a decision of that type has been made, the leeway incompatibilist could accept that a person can be caused to decide something and still hold on to PAP.

<sup>4</sup> Or is suitably "cognitively sensitive" to this fact.

For if we allow that people can be caused to decide, then what the Frankfurtian needs to show is not just that Joe can be responsible when the state of affairs *that Joe decides to evade taxes* is inevitable, but that he can be responsible even when the state of affairs *that Joe decides on his own to evade taxes* is inevitable. And that he has not shown.

Let us now return to the case where Joe doesn't have these beliefs, and so doesn't have the required doxastic abilities (see 3.2). My suggestion is that if Joe is responsible, then what we have is a case of derivative responsibility. Consider the situation: we are told that Joe is deliberating about whether to claim the illegal tax deduction. And we're told that Joe's moral character is such that he "invariably" follows his self-interested desire. Indeed, he doesn't have the ability that the leeway incompatibilist thinks is required to decide not to evade taxes—Joe could only (unknowingly) gain that ability by doing something unrelated to it. But this means that the "decision" to evade taxes that it is stipulated he makes, if we can indeed call it a decision, is something which flows, pretty much automatically, from Joe's character. Indeed, it is hard to see what kind of control Joe exercises over this "decision." In what sense is he active? Maybe he's able to affect just *when* he makes the decision: he might make the decision at  $t_1$ , or perhaps at  $t_2$ , or maybe he could postpone it until tomorrow. But sooner or later this "decision" is going to flow from his desires, and there is nothing he can do (in the sense relevant to free will and control) to stop it.

Now, the leeway incompatibilist can readily agree that a person might be responsible for behaviour which flows from his character in this way. But as highlighted by the scare quotes, he will probably hesitate to say that intentions acquired in this way are *decisions*, and he might even hesitate to call the resulting behaviour an *action*. Still, he can affirm that the person is responsible by appealing to derivative responsibility. And this provides a plausible way of handling Joe's case. We might imagine, for example, that Joe was brought up by pious parents who inveighed against the greed inherent in much of modern society and who tried to bring Joe up to be a good citizen. Joe listened patiently enough but decided he was going to put the gratification of his desire for money ahead of all others. He vigorously pursued this goal by deliberately choosing, on many different occasions, to shut out moral considerations. Indeed, Joe did this with

joy, knowing that he would soon develop the habit of pursuing his own self-interests. Now, as Joe faces this decision about taxes, he will evade taxes pretty much “on autopilot,” as we might say. Nevertheless, he’s derivatively morally responsible for making the illegal tax deduction in virtue of those decisions which were made in order to form his character in a certain way. So the leeway incompatibilist can accept all the details of the case, if it is suitably filled out. Of course, if the case is filled out differently, such that it’s clear Joe isn’t responsible for the character traits from which the decision flows, then the leeway incompatibilist can simply deny that Joe is responsible. Either way, we do not have a successful Frankfurt-style case.

## Conclusion

In this essay I have investigated the interesting idea that some of the conditions typically treated as part of the “epistemic condition on moral responsibility” are better thought of as affecting what the agent controls. I have distinguished between notions of ability which include a doxastic component and those that do not, and have investigated how this distinction helps us to understand different levels of control that an agent might possess. I made what I took to be an important distinction between an agent’s control over her actions and her control over non-actional, dated states of affairs, highlighting two different uses of the ‘up to us’ locution which reflect this distinction. And I then applied these insights to the indeterministic buffer Frankfurt-style cases, arguing that a mixed response to these cases, dependent on whether the agent has any doxastic abilities, shows how these cases pose no problem for the defender of the Principle of Alternative Possibilities.<sup>5</sup>

Simon Kittle  
Department for Christian Philosophy  
Karl-Rahner-Platz 1, A-6020 Innsbruck, Austria  
simon.kittle@uibk.ac.at

<sup>5</sup> This work was in part supported by the Templeton World Charity Foundation through Grant No. TWCF0078/AB46.

## References

- Alvarez, Maria. 2009. Actions, thought-experiments and the 'principle of alternate possibilities'. *Australasian Journal of Philosophy* 87(1): 61–81.
- Anscombe, G.E.M. 1963. *Intention*. 2<sup>nd</sup> edition. Oxford: Blackwell.
- Di Nucci, Ezio. 2011. Frankfurt versus Frankfurt: a new anti-causalist dawn. *Philosophical Explorations* 14(1): 117–131.
- Fischer, John Martin. 1994. *The Metaphysics of Free Will: An Essay on Control*. Oxford: Blackwell Publishers.
- Fischer, John Martin and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Frankfurt, Harry. 1969. Alternate possibilities and moral responsibility. *Journal of Philosophy* 66(3): 829–39.
- Ginet, Carl. 2000. The epistemic requirements for moral responsibility. *Noûs* 34(s14): 267–277.
- Goldman, Alvin. 1970. *A theory of human action*. Englewood Cliffs New Jersey: Prentice Hall.
- Goldman, Alvin. 1972. Toward a theory of social power. *Philosophical Studies* 23(4): 221–268.
- Kane, Robert. 1996. *The Significance of Free Will*. New Edition. New York: Oxford University Press.
- Kittle, Simon. 2015. Abilities to do otherwise. *Philosophical Studies* 172(11): 3017–35.
- Mele, Alfred. 2010. Moral responsibility for actions: epistemic and freedom conditions. *Philosophical Explorations* 13(2): 101–111.
- Morriss, Peter. 1987. *Power: A Philosophical Analysis*. Manchester: Manchester University Press.
- Naylor, Margery Bedford. 1984. Frankfurt on the principle of alternate possibilities. *Philosophical Studies* 46(2): 249–258.
- Pereboom, Derk. 2009. Further thoughts about a Frankfurt-style argument. *Philosophical Explorations* 12(2): 109–118.
- Pereboom, Derk. 2012. Frankfurt examples, derivative responsibility, and the timing objection. *Philosophical Issues* 22(1): 298–315.
- Shabo, Seth. 2014. It wasn't up to Jones: unavoidable actions and intensional contexts in Frankfurt examples. *Philosophical Studies* 169(3): 379–399.
- Timpe, Kevin. 2011. Tracing and the epistemic condition on moral responsibility. *The Modern Schoolman* 88/1/2: 5–28.
- Widerker, David. 1995. Libertarianism and Frankfurt's attack on the principle of alternative possibilities. *Philosophical Review* 104(2): 247–261.
- Widerker, David. 2006. Libertarianism and the philosophical significance of Frankfurt scenarios. *Journal of Philosophy* 103(4): 163–187.

