

2 Ignorance, Alternative Possibilities, and the Epistemic Conditions for Responsibility*

Carolina Sartorio

Introduction

The relation between responsibility and ignorance is multifaceted.¹ On the one hand, it has been suggested that being responsible requires being ignorant of some facts about our world. It might require, for example, being ignorant that we lack alternative possibilities of action because we live in a deterministic world (or a world that is deterministic for all practical purposes), or at least being ignorant of what it is that we will do as a result of the deterministic causes of our behavior, which rob us of real alternatives. The very possibility of deliberation, it is sometimes argued, requires believing that we have alternative possibilities, for we cannot genuinely deliberate unless we think that there is more than one choice open to us at the time.² On the other hand, it is commonly assumed that certain forms of ignorance exculpate.³ In this sense, responsibility—more specifically, blameworthiness, a form of responsibility—seems to require the opposite of ignorance, in that being ignorant of certain facts is exactly what *prevents* us from being blameworthy for things for which we would otherwise be blameworthy.

In this chapter I will explore a potential connection between responsibility and ignorance that combines these two lines of thought, in that it concerns a form of ignorance that is exculpating, *but* one that is also motivated by considerations concerning alternative possibilities. The chapter focuses on the epistemic conditions for blameworthiness, i.e., the conditions that we must meet in order to be blameworthy and that have to do with our beliefs or, more generally, our epistemic state. I will discuss cases where, although we actually *do* have alternative possibilities, we believe that we don't. In those cases, ignorance of the relevant facts about the world, which results in the belief that there are no alternatives, seems to relieve the agent of responsibility.

The Belief in Alternatives View

Consider the following scenario:

Two Doors: I must come out of a room or I will asphyxiate. There are two doors, A and B. As a matter of fact, opening door A would result in an explosion in an adjacent room and cause some damage, and opening

door B would not. However, I have acquired (through no fault of my own) good reasons to believe that opening *either* door would cause the explosion. I randomly pick door A and freely choose to open it. As a result, the explosion occurs.

It seems clear that I am not blameworthy for the explosion in this case. Moreover, it seems clear that my lack of blameworthiness is grounded in my epistemic state, in particular, in my (blameless) ignorance of the relevant facts about the world.

On the other hand, notice that I opened door A freely (no one forced me to do it, I did it on the basis of reasons, etc.). Moreover, I did so foreseeing that it would result in an explosion. So note, in particular, that my ignorance in this case is limited to certain facts of the world but not others: I am ignorant of the consequences that opening door B would have, but I am not ignorant of the consequences that opening door A would have.

Interestingly, as a result of this, this case poses a challenge for a quite standard way of understanding the epistemic conditions for blameworthiness for harms. For it is common to suggest that the harms for which we are blameworthy are the reasonably *foreseeable* harms that (non-deviantly) result from our free acts. For example, it is common to suggest that a reckless driver can be blameworthy for running over a pedestrian and causing him harm because she could reasonably foresee that driving in that fashion was likely to result in harm of that type (at least much more likely than if she were to drive more carefully). The epistemic conditions for blameworthiness for harms are often cashed out in terms of a foreseeability condition of this kind, one that identifies the harms that we are responsible for with the harms that we could foresee would happen or were (to a considerable degree) likely to happen as a result of our behavior, or, sometimes, with the harms that a reasonable person in our circumstances could foresee would happen or were (to a considerable degree) likely to happen as a result of our behavior.

As a specific example, consider one of the most sophisticated accounts of the epistemic conditions for blameworthiness for harms: Carl Ginet's account.⁴ Ginet's account is quite complex; however, simplifying a bit in ways that are not relevant for our purposes here, it basically amounts to this:

S is blameworthy for bringing about a harm by acting in a certain way only if (i) [*freedom condition*] it was open for S to act in a different way that would not have resulted in the harm, and (ii) [*epistemic condition*] either S knew that by acting in that way she would or might bring about that kind of harm, or there is something S did earlier that she knew would or might eventually lead to her bringing about that kind of harm.

As stated, this account offers only necessary conditions for blameworthiness for harms, but Ginet's paper ends with the conjecture that the conditions are also sufficient.⁵

But note that, if (ii) is the epistemic condition that I have to meet to be blameworthy for a harm, I clearly meet that condition in Two Doors. I could foresee (we may assume, with close to complete certainty, and on the basis of solid evidence) that opening door A would result in an explosion and in some harm. So the foreseeability condition is met. The other conditions for responsibility are met; in particular, I freely opened the door, and (unbeknownst to me) it was possible for me to avoid causing the harm without asphyxiating.⁶ So it follows from this type of account that I am blameworthy for the explosion. However, clearly I am not. And, again, the reason I am not blameworthy seems to concern the epistemic component of responsibility. For, intuitively, the reason I am not blameworthy is that I was ignorant about some of the relevant facts: I didn't know that opening door B would not have resulted in the explosion; in fact, I reasonably believed the opposite. So this suggests that this way of understanding the epistemic conditions for blameworthiness for harms is flawed. In particular, it seems to show that Ginet's conjecture is wrong, in that the conditions he offers are not sufficient for blameworthiness, and they are not sufficient because the epistemic condition is incomplete as it stands.

To clarify, here I am focusing on blameworthiness as a specific form of moral responsibility. Sometimes philosophers distinguish blameworthiness (and its counterpart, praiseworthiness) from a broader or more neutral concept of moral responsibility, one according to which one can be morally responsible for something without being blameworthy or praiseworthy for it. Perhaps there is a sense in which I am morally responsible for the explosion in Two Doors, and it is a more neutral sense of that kind. Perhaps all that's required to be responsible for an outcome in that sense is that the outcome be the foreseeable result of something we did freely, or something we knew would happen as a result of our free agency. But (like Ginet) here I am interested in blameworthiness, not in this broader sense of responsibility, and I take it that it is clear that I am *not* blameworthy for the explosion in a case like Two Doors.

So Ginet's account of the epistemic condition for blameworthiness for harms is incomplete in some significant way. What is missing? A natural suggestion is that what's missing is something like a 'Belief in Alternatives' condition:

(BA) S believed (knew?) that she had alternatives.⁷

That is to say: S believed (or knew) that there was something she could have (reasonably) done that would not have resulted in the harm. In particular, the suggestion is that being blameworthy for a harm requires, in cases where one has alternatives, that one *not* be ignorant of the existence of those alternatives (for that kind of ignorance is exculpatory). Note that here, following common usage, 'alternatives' should be read as *reasonable* alternatives. The way I am imagining the case, there is something I could have done in Two

Doors that wouldn't have resulted in the explosion, namely, not opening any door and asphyxiating. Obviously this isn't a reasonable alternative. Thus, on this natural way of understanding the BA condition, it is not satisfied in Two Doors.

We have seen that, if BA were a necessary condition for blameworthiness for harms, then the failure of this condition would account for my lack of blameworthiness in Two Doors. For then my ignorance about the relevant facts of the world, which resulted in the absence of a belief in alternatives, would be exculpatory. But is BA a genuine epistemic requirement for responsibility? More precisely, is BA a genuine requirement for blameworthiness for harms?

In the remainder of this chapter I will examine this issue. As we will see, there are other scenarios that suggest that it is not a genuine requirement. If that is the case, then there must be a different way of accounting for the agent's lack of blameworthiness in cases like Two Doors. But it's not obvious what this other way could be; as a result, there is an interesting puzzle.

The main aim of this chapter is to draw attention to that puzzle. Toward the end of the chapter I will also give a sketch of a possible solution, one that I find promising, but I hasten to add that it's far from a fully worked out proposal, as it stands. Again, my main aim here is to explain how the problem arises, and why it is an interesting problem.

Nelkin's Variations on Frankfurt-Style Cases

Dana Nelkin once tried to undermine the thesis that deliberation requires the belief in alternatives by appeal to a certain variant on the famous 'Frankfurt-style cases.'⁸ The most popular Frankfurt-style cases are scenarios where an agent makes a certain choice completely on his own, but a neuroscientist has been secretly monitoring the agent's thoughts and has the ability to intervene in order to guarantee that the agent will make the choice that he wants. Given that the agent ended up making the choice that the neuroscientist wanted him to make, the neuroscientist never intervenes, and thus the agent appears to be responsible for his choice (a choice that he couldn't have avoided). In Nelkin's variations, and in contrast with the kinds of scenarios originally imagined by Frankfurt, the agent *is aware* of the presence of the neuroscientist and so he is aware of the fact that he cannot make a different choice, but he still makes the choice completely on his own, on the basis of his own reasons, and without the neuroscientist's intervening or entering the deliberation process in any way. Nelkin argued that in this kind of case the agent is able to deliberate about whether to make the relevant choice, even though he doesn't believe he has alternatives. If so, this is a counterexample to the idea that deliberation requires belief in alternatives.

Here I will not take a stand on this issue, for I am not interested in the conditions for *deliberation* but, instead, in the conditions for responsibility (more specifically, blameworthiness).⁹ So I will use the Nelkin variants for

a different purpose, namely, casting doubt on BA as a genuine requirement for blameworthiness. As I will attempt to show next, regardless of what one thinks about the relevance of the Nelkin variants for deliberation, it is plausible to argue that cases of that kind constitute counterexamples to the BA condition as a necessary condition for blameworthiness.

It will help to work with a specific example. Consider this scenario (a variant on one of Frankfurt's original cases):

Jones and the Neuroscientist: A neuroscientist wants Jones to shoot Smith, but he prefers to avoid showing his hand unnecessarily. If it were to become clear that Jones is going to decide to do something else, he would take effective steps to ensure that Jones decides to shoot Smith, by directly manipulating the relevant processes in Jones's brain. (The neuroscientist bases his prediction of what Jones will decide to do on some reliable sign, such as a certain twitch in his face that he invariably shows when he is about to make decisions of that kind.) As it turns out, the neuroscientist never has to intervene because Jones decides to shoot Smith for reasons of his own—say, to get revenge for something that Smith did to him earlier. At some point during this process, Jones becomes aware of the neuroscientist's presence and intentions. However, this does not affect his reasoning in any way: he still makes the decision to shoot Smith on the basis of his own reasons, in exactly the same way he would have made it if he had never become aware of the neuroscientist's presence.

It seems quite plausible to claim that Jones is blameworthy for his choice to shoot Smith, and for the possibly harmful consequences of that choice (such as Smith's death), even though he knew that he couldn't have made a different choice—in other words, even if he failed to satisfy the BA condition. After all, he made the choice on his own, on the basis of his own reasons (the desire for revenge), and not at all as a result of the neuroscientist's presence, or as a result of his awareness of the neuroscientist's presence.

To motivate this some more, recall Frankfurt's own reasoning about the original Frankfurt-style cases. Frankfurt argued that the neuroscientist's presence is not relevant to the agent's responsibility in those cases because, given that the agent decided to make the choice on his own, the neuroscientist never intervened and thus was never part of the causal chain of events resulting in the agent's choice. Frankfurt argued that factors that are in this way irrelevant to the actual causal history of the choice are irrelevant to the agent's responsibility for his choice.¹⁰ Now, arguably, if this reasoning is sound, similar reasoning can be used to show that the agent is responsible for his choice *even* when he is aware of the neuroscientist's presence, in cases where the neuroscientist's presence did not affect the agent's deliberation in any way. If the agent still made the choice purely on the basis of his own reasons (the same desire for revenge), in the same way he would have made

it if the neuroscientist hadn't been there, then he seems just as responsible for his choice and whatever harmful consequences result from it, even if he could not have done otherwise, and even if he was fully aware of that fact.

In other words, assuming one agrees with Frankfurt that the agent is responsible for his choice in a standard Frankfurt-style case, it is plausible to extend that reasoning to scenarios like Jones and the Neuroscientist. Here is one way to reconstruct that argument:

- 1 The agent in a standard Frankfurt-style case is responsible for his choice (By assumption about standard Frankfurt-style cases).
- 2 If 1 is true, then the agent in a Frankfurt-style case would still be responsible for his choice in cases where he is aware of the neuroscientist's presence, *unless that awareness affected his deliberation in some significant way*.
- 3 In Jones and the Neuroscientist, Jones's awareness of the neuroscientist doesn't affect his deliberation in any significant way (By design about the case).
- 4 Therefore, in Jones and the Neuroscientist, Jones is responsible for his choice (From 1, 2, and 3).

The only premise that needs justification seems to be 2. The justification for 2 is that the reasons Frankfurt offers in support of the idea that the agent is responsible for his choice in a standard Frankfurt-style case seem to carry over to this kind of case. If the awareness of the neuroscientist did not make a difference to the agent's deliberation—if the agent arrived at his choice in exactly the same way he would have if he had not been aware of the neuroscientist's presence—then it seems that the awareness is irrelevant to the agent's responsibility, and so he is just as responsible as in the original case. If so, Jones and the Neuroscientist is a counterexample to the BA condition.

Of course, although Frankfurt's argument has convinced many, others remain unconvinced. This is not the place to assess the merits of that argument.¹¹ All I want to draw attention to here is the conditional claim: *if* the argument works for the original cases, *then* it is plausible to believe that it also works for a variant like Jones and the Neuroscientist, and thus the BA condition fails. I personally find Frankfurt's reasoning plausible and Frankfurt-style cases illuminating, so I am tempted to think that the right view about the epistemic requirements for responsibility shouldn't be committed to the truth of the BA condition. But, more importantly, I think this at least shows that the BA condition is *controversial*. So, even if one isn't completely persuaded by the reasoning about Frankfurt-style cases, it is worth investigating what other account of the epistemic conditions could be offered (one that didn't rely on a condition of that kind). This is enough to motivate the search for a different account.

So we are back to the drawing board. In Two Doors, somehow, the agent's ignorance of the relevant facts exculpates her. Her failure to meet the BA

condition seemed like a natural explanation of why she is not responsible. Assuming this is not the real explanation, what is it? What could be the difference between Two Doors and Jones and the Neuroscientist? In what follows I examine possible answers to these questions.

The Inherited Responsibility View and Zimmerman's View

Recall that we have been dealing with blameworthiness for harms. Now, harms are *consequences* of a certain kind, and responsibility for consequences is standardly considered to be a form of derivative (that is, non-basic) responsibility. In cases of derivative responsibility, the agent's responsibility is inherited from other things for which she is responsible—ultimately, from things for which she is basically responsible. Thus, if the form of responsibility at issue is a kind of derivative responsibility, then the agent's blameworthiness for the harm in a case like Two Doors would have to be inherited from other things. Presumably, the agent would have to be blameworthy for something like an act or choice, and her blameworthiness for the harm would then be inherited from her blameworthiness for those other things, assuming that the relevant epistemic conditions for responsibility are satisfied.

So this suggests that perhaps, in cases of this kind, the conditions for responsibility should not be analyzed in terms of a freedom condition and an epistemic condition, but, instead, as something like an *inherited responsibility* condition and an epistemic condition. The idea would be to suggest that the freedom condition plays a role only in cases of *basic* responsibility or basic blameworthiness (the ultimate grounds for responsibility), and in all other cases the inherited responsibility condition takes its place.

According to this line of thought, then, the right account of blameworthiness for harms would roughly look like this:

(Inherited Responsibility View)

S is blameworthy for a harm just in case (i) [*inherited responsibility condition*] S is blameworthy for some act or choice that (non-deviantly) resulted in the harm, and (ii) [*epistemic condition*] S meets the relevant epistemic conditions with respect to that harm (e.g., some kind of foreseeability condition).¹²

Then one way to use this account to explain the difference between Two Doors and Jones and the Neuroscientist would be to argue that the inherited responsibility condition is not satisfied in Two Doors, but it is satisfied in Jones and the Neuroscientist. For example, one could argue that I am not blameworthy for my choice to open door A in Two Doors, but Jones is blameworthy for his choice to shoot Smith in Jones and the Neuroscientist.

All of this seems very plausible. But, does it work as an answer to our puzzle? The worry is that this just pushes the problem back one step. For, now the question arises: *why* am I not blameworthy for choosing to open door A in Two Doors? In particular, why am I not blameworthy for that choice, if I made the choice freely, and knowing that it would result in an explosion? Again, it seems very natural to suggest that this has something to do with my belief that I did not have reasonable alternatives: although I knew I had a choice about which door to open, I did not believe I had a (reasonable) choice about whether to cause harm. But, if this were the right answer, then it would follow that Jones (in Jones and the Neuroscientist) cannot be blameworthy for his choice to shoot Smith, since he also didn't believe he had alternatives. If so, we have not made any real progress: we're still in need of an explanation of the difference between Two Doors and Jones and the Neuroscientist.

At this point another possibility might come to mind, one that seems at least a bit more promising: it is to appeal, not to a belief about alternatives, but to a belief about *wrongdoing*. Let us examine this possibility.

In the literature on responsibility and ignorance, some philosophers have forcefully argued for the view that culpability for ignorant behavior can always be traced back to, and is always ultimately grounded in, culpability for *non-ignorant* behavior. Michael Zimmerman, in particular, argued for a view of this kind according to which the lack of ignorance that ultimately grounds the agent's culpability in cases of culpable ignorance is or involves a belief on the agent's part that she acted wrongly.¹³ In a nutshell, Zimmerman's reasoning is that, unless we can find an episode of awareness of that kind in which to root the agent's culpability, it is unfair to hold him blameworthy for his behavior. Although this is not the place to assess the general merits of such a view, or of Zimmerman's arguments, for our purposes here it is worth noting that this kind of reasoning motivates a different requirement for blameworthiness, one that can be cashed out as something like a 'belief in wrongdoing' condition:

(BW) S believed that she was acting wrongly.¹⁴

(This is quite rough, but it will do for our purposes here. Again, the thought is not that all behaviors for which agents are blameworthy must satisfy this condition, but only that all culpable ignorance must be ultimately rooted in *some* behavior that satisfies this condition.)

Can this help with our puzzle? Recall that what we are looking for is an account that could explain why Jones is blameworthy for his choice to kill Smith in Jones and the Neuroscientist, but I am not blameworthy for my choice to open door A in Two Doors. So one would have to argue that Jones is blameworthy for his choice because he meets the BW condition, but I am not blameworthy for my choice in Two Doors because I fail to meet that condition.

The claim about Two Doors seems clearly true. Given that I was ignorant of the relevant facts (the facts that make it the case that I should have opened door B instead of door A), I did not believe that I was acting wrongly by opening door A. But things are much less clear when it comes to Jones and the Neuroscientist. On the one hand, it might seem natural to read into my description of the case that Jones took himself to be acting wrongly in making the choice to shoot Smith because he made the choice just on the basis of his own reasons (his desire for revenge), in a way that was not affected by his awareness of the neuroscientist's presence. However, it is not clear that this is the *best* way to understand the case, or the *only* possible way to understand the case (note that the puzzle would persist if there were at least one version of the case that could not be interpreted in that way).

Let me explain. The standard view on wrongness is that its being wrong for S to do A amounts to, or at least entails that, S ought to have refrained from A-ing. Now, many people believe that 'ought implies can,' roughly, the principle that S ought to have done something only if S could have done it. By assumption, however, the agents in Frankfurt-style cases couldn't have avoided making the choice they made. Thus the ought-implies-can principle entails that it is not the case that those agents ought to have made a different choice. If so, the standard view on wrongness entails that it *was not* wrong for them to make those choices. In particular, Jones didn't act wrongly when he made the choice to shoot Smith, in Jones and the Neuroscientist.¹⁵

Of course, Jones could still *believe* that he was acting wrongly, even if the ought-implies-can principle were true and he was not really acting wrongly. What an agent is likely to believe in a case of that kind would depend on many things, including his psychology, his power of reflection, his philosophical proclivities, etc. So let us fill in the details of the case some more. Imagine that Jones is a moral philosopher—as it happens, a stern believer in the ought-implies-can principle—and has reflected on the situation carefully enough *not* to believe that he is acting wrongly in making the choice to shoot Smith, on the basis of the fact that he lacks alternatives. Still, the same argument from the last section seems to apply equally well to this case. Assuming (following Frankfurt's original reasoning) that Jones *would* have been responsible for his choice if he *had not* been aware of the neuroscientist's presence, then it seems that he must also be responsible when he *is* aware of the neuroscientist's presence, given that such awareness did not affect his decision-making process in any way. And, again, it does seem possible to imagine a case of that kind, one where the awareness of the neuroscientist's presence does not affect Jones's deliberation and decision-making process in any way. Arguably, all one needs to imagine is that he still makes the decision to shoot Smith on his own, on the basis of his own reasons (his desire for revenge), and in exactly the same way he would have made it if he hadn't been aware of the neuroscientist's presence.

As a result, there are reasons to think that Jones is blameworthy for his choice to kill Smith, even if he didn't believe that he was acting wrongly in

making that choice. Therefore, the BW condition has trouble accommodating the Jones and the Neuroscientist case, at least when the case is interpreted in this particular way.

On reflection, the failure of the BW condition should not come as a surprise. For the BA condition and the BW condition are related: at least for someone who is committed to the ought-implies-can principle, the belief about wrongdoing is likely to be connected to the belief about alternatives, in a way that gives rise to the same kinds of problems in cases like Jones and the Neuroscientist. So it is natural to expect that, if the appeal to the BA condition fails, so does the appeal to the BW condition.

The Awareness View

Arguably, what the preceding discussion suggests is that we should look for a *looser* epistemic condition for blameworthiness, one that doesn't require a belief in alternatives or a belief in wrongdoing. In this section, I will sketch a way in which I think this could be done. This is really just intended as a sketch, not as a fully developed proposal, but hopefully it is enough at least to point us in the right direction.

In the literature on free will, it is common to see the epistemic condition for responsibility characterized in a way that is quite vague, as something like 'the awareness of what one is doing' or 'the awareness of the moral significance of what one is doing.'¹⁶ Although this is too vague as it stands, perhaps it could be made more precise in a way that could help with our puzzle. So, as a first step, this is the alternative account I'm proposing:

(Awareness View)

The agent was (or perhaps should have been) aware of the moral significance of what she was doing.

The second step is the suggestion that there are different ways in which we could be aware of the moral significance of what we do, and that not all of them involve a belief that one is acting wrongly, or an underlying belief in alternatives. In particular, the agent in Jones and the Neuroscientist is aware of the moral significance of what he is doing, in a way that makes him blameworthy for his choice, even if he doesn't believe that he is acting wrongly, and even if he does not believe that he has alternatives.

Recall the argument from the section on Nelkin's variations on Frankfurt-style cases, which appeared to show that Jones is blameworthy for his choice in that case. That argument extended the reasoning about standard Frankfurt-style cases, where the agent is ignorant of the fact that he lacks alternatives, to the Nelkin variants (like Jones and the Neuroscientist), where the agent is aware that he lacks alternatives. The argument relied on

the claim that the *same* reasons that (following Frankfurt) show that the agent is blameworthy in the standard cases also show that he is blameworthy in the Nelkin variants. Those reasons had to do with the *actual* deliberation or decision-making process that led Jones to make his choice. The thought was that, if that actual process is all that matters, and if the awareness of the neuroscientist doesn't affect that process or make a difference to how the agent made his choice, then the agent in Jones and the Neuroscientist is just as blameworthy as in the original case. He is just as blameworthy because he still makes the choice completely on his own, on the basis of his own reasons (morally reproachable reasons, such as a desire for revenge), in exactly the same way he would have made it if he hadn't been aware of the neuroscientist's presence.

If this reasoning is sound, then it can help shed some light on how one could argue that the awareness condition is satisfied in a case like Jones and the Neuroscientist. For it motivates the idea that what makes Jones blameworthy is something like *his recognizing that he is acting from morally reproachable reasons*. He is aware that he is making the choice on the basis of his own reasons, because he wants to make it, and he is aware that the reasons that are motivating him to act are not acceptable on general moral grounds. In contrast, when I make the choice to open door A in Two Doors, I am not acting from morally reproachable reasons, but I am motivated only by the desire not to asphyxiate (which, given my ignorance about the relevant facts concerning door B, leads me to make the choice to open one door at random).

Again, the main thought is that the relevant condition—*being aware of the moral significance of our behavior*—could be satisfied in different ways in different circumstances. In circumstances where we act wrongly, it could be satisfied by the awareness that we were acting wrongly, or by the awareness that one ought to have behaved differently. In circumstances where we don't act wrongly, and perhaps are aware that we do not act wrongly, it could be satisfied simply by virtue of recognizing that we are acting from morally reproachable reasons.

Note that the view is actually neutral on the issue of whether ought implies can. This is intentional.¹⁷ If ought *does not* imply can, then of course agents in Frankfurt-style cases could still be acting wrongly; if, on the other hand, ought *does* imply can, then agents in Frankfurt-style cases do not act wrongly. But the account entails that the agents in those cases (including the Nelkin variants discussed above) can satisfy the awareness condition even if they don't believe that they're acting wrongly, to the extent that they recognize that they are acting from morally reproachable reasons.

Of course, as anticipated, this is just a sketch of the view; the details would still need to be worked out.¹⁸ But I think it is enough to see that a different account of the epistemic conditions is possible, one that doesn't rely on the truth of the BA condition or the BW condition, and one that could provide a plausible answer to our puzzle.

Conclusions

This chapter examined the epistemic conditions that agents must satisfy in order to be blameworthy. We have seen that pairs of scenarios like Two Doors and Jones and the Neuroscientist give rise to an interesting puzzle. The puzzle is to explain how the ignorance of the relevant facts about the world in a case like Two Doors can absolve me of moral responsibility in a way that is consistent with the claim that Jones is responsible for his choice in a case like Jones and the Neuroscientist. As we have seen, this puzzle resists an easy solution. We saw that, in particular, principles like the BA condition and the BW condition, which seem to naturally explain the fact that the agent fails to be responsible in Two Doors, have trouble accommodating the fact that the agent seems to be responsible in Jones and the Neuroscientist.

This motivated the search for an alternative account of the epistemic conditions for blameworthiness. I outlined an account that focuses on the agent's awareness of the moral significance of her behavior. I explained how that type of account could, at least in principle (if fleshed out in more detail), provide a successful solution to our puzzle. It could do that, I argued, because it could explain how the ignorance of certain facts about the world can be morally exculpatory in cases like Two Doors without relying on the truth of the BA condition or the BW condition, but simply on the truth of a more general principle of moral awareness.

Notes

- * Thanks to E.J. Coffman, Juan Comesaña, Michael McKenna, Rik Peels, and Jan Willem Wieland for helpful comments.
- 1 For the purposes of this chapter I will remain as neutral as possible on the nature of ignorance itself. I think everything I say here is compatible with ignorance being lack of knowledge, as it is traditionally assumed (see, e.g., Zimmerman 1988, p.75), as well as with weaker views on ignorance, such as the view that ignorance is mere lack of true belief (Peels 2010). In particular, the cases I will be focusing on are cases of lack of true belief, which both views would count as cases of ignorance.
- 2 See, e.g., van Inwagen (1983); Coffman and Warfield (2005).
- 3 See, e.g., Smith (1983).
- 4 Ginet (2000).
- 5 Ginet (2000, p.277).
- 6 We may also assume that the way in which opening door A results in the explosion exactly matches the way in which I expected it to happen, so the non-deviance condition would also be met in this kind of case.
- 7 A bit less strongly, one could argue that it requires the absence of the belief that one lacks alternatives. I will not consider this possibility here because it's subject to the same kinds of problems that I'll point out for the BA condition. Another (related) possibility is to argue that what is missing is the agent's belief that she ought to have acted in a different way, or the agent's belief that she was acting wrongly. I will consider this possibility later—see my discussion of the BW condition below in the section on the Inherited Responsibility View and Zimmerman's view.

- 8 See Nelkin (2004). Frankfurt-style cases are originally from Frankfurt (1969). Frankfurt used them to undermine the traditional view of responsibility—the idea that responsibility requires the agent’s ability to do otherwise, or the agent’s access to alternative possibilities of action.
- 9 For responses to Nelkin, see Coffman and Warfield (2005) and Coffman (forthcoming).
- 10 Thus Frankfurt wrote: “When a fact is in this way irrelevant to the problem of accounting for a person’s action it seems quite gratuitous to assign it any weight in the assessment of his moral responsibility. Why should the fact be considered in reaching a moral judgment concerning the person when it does not help in any way to understand either what made him act as he did or what, in other circumstances, he might have done?” And: “The circumstances that made it impossible for him [the agent in a Frankfurt case] to do otherwise could have been subtracted from the situation without affecting what happened or why it happened in any way. Whatever it was that actually led the person to do what he did, or that made him do it, would have led him to do it or made him do it even if it had been possible for him to do something else instead.” (Frankfurt (1969), in Watson (2003, p.174)).
- 11 For a survey of the main views on the prospects of Frankfurt-style cases, see Sartorio (2016a). In Sartorio (2011; 2016b, ch.1) I argue, against Frankfurt’s view, that an agent’s beliefs can sometimes be relevant to his responsibility without bearing on the actual explanation of the choice he made. However, these are beliefs that can make an agent *blameworthy* for something (for example, my awareness that I am acting wrongly can make me blameworthy for what I do even if I don’t do it *because* it’s wrong). Jones and the Neuroscientist is different: in this case, if his awareness of the neuroscientist made a difference to his responsibility, it would be by rendering him *non-blameworthy*, instead of blameworthy. This clashes with the idea that, if a factor has no bearing on the actual explanation of your behavior, you cannot use it as an excuse, and it does not absolve you of moral responsibility (see Frankfurt (1969), in Watson (2003, p.175)).
- 12 See, e.g., Feinberg (1970) and Smith (1983). Presumably, some epistemic considerations (perhaps different from those that concern condition (ii)) will also enter into the satisfaction of condition (i), since an agent’s being basically blameworthy for an act or choice arguably involves the satisfaction of some conditions that are epistemic in nature.
- 13 See Zimmerman (1986; 1988; 1997).
- 14 Note that, just as with the BA condition, another possibility is to require something stronger, such as *knowledge* that one is acting wrongly. Similar problems to those that arise for BW arise for the stronger versions, so I will not be concerned with the stronger versions here.
- 15 Those who agree with Frankfurt that the agent in a Frankfurt-style case is blameworthy for his choice are of a divided opinion on the issue of the ought-implies-can principle. Some believe that Frankfurt-style cases are themselves counterexamples to that principle, because the agents in those cases ought to have made a different choice despite the fact that they could not have made a different choice (see, e.g., Fischer (2003) and Frankfurt (2003)). In turn, others believe that Frankfurt-style cases are not counterexamples to the ought-implies-can principle, but to the principle that blameworthiness entails wrongdoing (see, e.g., McKenna (2008), Capes (2010), and Pereboom (2014)). (Note that, if I am right about the Nelkin variants of Frankfurt-style cases, then this might call for a revision, not just of the principle that blameworthiness requires wrongdoing, but also of the principle that blameworthiness requires *belief in* wrongdoing.)
- 16 For example, Fischer and Ravizza write: “The first condition, which may be termed the “epistemic condition,” corresponds to the excuse of ignorance. It

captures the intuition that an agent is responsible only if he both knows the particular facts surrounding his action, and acts with the proper sort of beliefs and intentions.” (Fischer and Ravizza 1998, p.13) Fischer and Ravizza (as well as other authors) distinguish the epistemic condition from the metaphysical freedom or control condition, and they trace this distinction to the two kinds of excuse originally identified by Aristotle: ignorance and force. For discussion of the distinction between the freedom and epistemic condition (in connection with Fischer and Ravizza’s view but also more generally), see Mele (2010).

- 17 As I pointed out above (see note 15), advocates of Frankfurt-style cases are of a divided opinion on this matter, in particular, they disagree about whether Frankfurt-style cases themselves are counterexamples to the ought-implies-can principle.
- 18 Some examples of the questions that arise for the view are: What does the relevant awareness consist in? (Is it mere true belief, or justified true belief/knowledge, etc.?) How does that awareness interact with the awareness that one is acting wrongly, in cases where one is acting wrongly and is aware of that? (Does the awareness that one is acting from morally reproachable reasons make one more blameworthy in those cases?)

Bibliography

- Capes, Justin. 2010. “The W-Defense.” *Philosophical Studies* 150: 61–77.
- Coffman, E.J. “Deliberation.” In *Routledge Companion to Free Will*, edited by Meghan Griffith, Neil Levy, and Kevin Timpe. New York: Routledge.
- Coffman, E.J., and Ted Warfield. 2005. “Deliberation and Metaphysical Freedom.” *Midwest Studies in Philosophy* 29: 25–44.
- Feinberg, Joel. 1970. “Sua Culpa.” In *Doing and Deserving*, edited by Joel Feinberg, 187–221. Princeton, NJ: Princeton University Press.
- Fischer, John. 2003. “‘Ought-Implies-Can,’ Causal Determinism, and Moral Responsibility.” *Analysis* 63: 244–250.
- Fischer, John, and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.
- Frankfurt, Harry. 1969. “Alternate Possibilities and Moral Responsibility.” *Journal of Philosophy* 66: 829–839.
- Frankfurt, Harry. 2003. “Some Thoughts Concerning PAP.” In *Moral Responsibilities and Alternative Possibilities*, edited by David Widerker, and Michael McKenna, 339–345. Aldershot: Ashgate.
- Ginet, Carl. 2000. “The Epistemic Requirements for Moral Responsibility.” *Philosophical Perspectives* 14: 267–277.
- McKenna, Michael. 2008. “Frankfurt’s Argument Against Alternative Possibilities: Looking Beyond the Examples.” *Noûs* 42: 770–793.
- Mele, Alfred. 2010. “Moral Responsibility for Actions: Epistemic and Freedom Conditions.” *Philosophical Explorations* 13: 101–111.
- Nelkin, Dana. 2004. “The Sense of Freedom.” In *Freedom and Determinism*, edited by Joseph Campbell, Michael O’Rourke, and David Shier, 105–134. Cambridge, MA: MIT Press.
- Peels, Rik. 2010. “What Is Ignorance?” *Philosophia* 38: 57–67.
- Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
- Sartorio, Carolina. 2011. “Actuality and Responsibility.” *Mind* 120: 1071–1097.
- Sartorio, Carolina. 2016a. “Frankfurt-Style Examples.” In *Routledge Companion to Free Will*, edited by Meghan Griffith, Neil Levy, and Kevin Timpe. New York: Routledge.

- Sartorio, Carolina. 2016b. *Causation and Free Will*. Oxford: Oxford University Press.
- Smith, Holly. 1983. "Culpable Ignorance." *The Philosophical Review* 92: 543–571.
- Van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.
- Watson, Gary, ed. 2003. *Free Will*. New York: Oxford University Press.
- Zimmerman, Michael. 1986. "Negligence and Moral Responsibility." *Nous* 20: 199–218.
- Zimmerman, Michael. 1988. *An Essay on Moral Responsibility*. Totowa, NJ: Rowman and Littlefield.
- Zimmerman, Michael. 1997. "Moral Responsibility and Ignorance." *Ethics* 107: 410–426.