The University Of Sheffield.

Fairness, "Ought Implies Can", and the Origin of Alternate Possibilities.

By:

Sam Waters

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Arts and Humanities
Department of Philosophy

September 2018
Word Count: 72,583

The University Of Sheffield.

Fairness, "Ought Implies Can", and the Origin of Alternate Possibilities.

By:

Sam Waters

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Arts and Humanities
Department of Philosophy

September 2018
Word count: 72,583

Abstract

The goal of this thesis is to set out an argument which demonstrates that the Principle of Alternate Possibilities (PAP), which holds that an agent may only be blameworthy for performing a given action if they were able to do otherwise than it, can be logically derived from a certain interpretation of the common moral principle that "ought implies can" (OIC). In pursuing this "Derivation", I follow David Copp in claiming that both PAP and OIC are in large part motivated by the same considerations about fairness and the role that our genuine moral obligations may play in our processes of moral deliberation.

I begin by outlining a case in favour of my preferred sense of OIC which is grounded in these shared motivations, before protecting it from a varied range of counterexamples. I then turn my attention to defending what I call the "connecting premises" of the Derivation- a set of reasonably uncontroversial moral beliefs which, in conjunction with my understanding of OIC, show that PAP is true. In arguing for these premises, I draw on a broader normative picture in which our moral obligations, blameworthiness for failing in those obligations, and the morally relevant sense of ability which underpins both of them are inherently linked.

I conclude my argument in support of the Derivation with an extensive discussion of the concept of fairness as it applies to discussions of blameworthiness and moral obligation, and develop the aforementioned normative picture into an argument that *all* genuine moral obligations are necessarily fair in this sense. Hence, the Derivation will serve not only as a proof of PAP, but also as evidence for an interrelated set of moral principles which govern the conditions under which an agent may be obligated to act and/or be correctly said to have acted wrongly.

Acknowledgements

"Talent is a pursued interest. In other words, anything that you're willing to practice, you can do." - Bob Ross, *The Joy of Painting.*

This thesis would not have been written without the help and support of many of the staff of the University of Sheffield's Philosophy department, who have been my teachers, colleagues, and friends throughout my efforts to bring this work to completion over the past six years. I am grateful to all of them for their kindness and understanding during a long and often difficult period of study, and to Professor Jenny Saul in particular for her patience and support in helping me return from the multiple periods of absence which have interrupted my research.

Having said this, I would like to extend my sincerest gratitude to my past and present supervisors Dr Yonatan Shemmer, Dr Steve Makin and Professor Miranda Fricker for their outstanding philosophical advice and genuine engagement with every argument, no matter how shockingly bad, that I ever brought to them. I can honestly say that I would not be where I am today if not for their inexhaustible faith in my abilities, and I do not have the words to express my gratitude for their efforts.

As in every other area of my life, my mother's positive influence on my professional work cannot possibly be underestimated. In addition, I owe a great debt to both Anaïs White and Morrigan Auxland for supporting me through the horrendously difficult final months. You never gave up on me, and ensured I did not give up on this volume. I hope it is deserving of your unwavering faith in me, then and now.

I also wish to thank Dr Graeme A. Forbes of the University of Kent for the many months spent discussing topics of mutual philosophical and political interest with me. His most valuable gift by far, however, has been his consistently sound guidance through the terrifying world of postgraduate life. I am indebted to him for his help in fine-tuning the ideas which now comprise this project, and I am proud to call him my friend.

Finally, I offer my genuine thanks to Jennifer Sharp, whose contribution to the success of this thesis requires no explanation.

- Sam.

Table of Contents

Chapter I - The History of Alternate Possibilities

The Principle of Alternate Possibilities (PAP), the view that an agent can only be morally responsible or blameworthy for an action if they "could have done otherwise" than that action, holds a generally unfavourable reception in the world of moral philosophy. Whilst criticism of PAP is far from a new phenomenon, having gained increasing ground ever since Harry Frankfurt's landmark article in the 1960's,[1] it is arguably more disputed at the present moment than it has ever been before. Despite the continued efforts of the modern defenders of PAP, it is safe to say the principle is no longer regarded as settled ground within the debate on moral responsibility and blameworthiness.

I believe this general trend away from PAP over the past few decades to have been in error. In this thesis, my goal is to help to correct that particular mistake. I am a supporter of the "classic" formulation of PAP given above, although my interest is specifically in regard to the question of blameworthiness rather than moral responsibility. I suggest that the general consensus that PAP is false has arisen from an improper understanding of the moral principles on which PAP is (or rather, should be) based. Furthermore, in many cases this misunderstanding has been committed by those who are avowed advocates of these foundational principles, whom I believe ought to be supporters of PAP and yet, disturbingly, are not. This thesis is therefore founded on the idea that there are many more people who *ought* to hold PAP, given what they already believe, than currently do.

Specifically, what I wish to set out in this work is a philosophically sound argument which deduces PAP, in respect to blameworthiness, from the broader and more widely accepted idea that "Ought Implies Can" (OIC). The latter principle, which is commonly held to have originated with Immanuel Kant,[2] states that in order for an agent to have a moral obligation to do or avoid doing something, they must be able to do or refrain from doing that action(s), respectively. By doing so, I am seeking to anchor in PAP in a different foundation than its adherents have generally tried to ground it - a foundation which, I believe, many more philosophers are prepared to stand on than are apparently willing to

[1] Frankfurt, H.G. 1969. 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy,* Volume 66, 23: 829-839.
[2] Kant, Immanuel. 1998. *Religion Within the Boundaries of Mere Reason and Other Writings,* ed. by Allen Wood and George di Giovanni (Cambridge: Cambridge University Press). There is some dispute about in what sense(s) Kant believed OIC to be true- Stern (2004) contains an excellent introduction to the discussion.

back PAP. Whilst there is considerable debate among moral philosophers regarding OIC's precise meaning and scope, there is a broad agreement that the principle is true at least in certain senses and conditions. This type of consensus simply does not exist with PAP at the present time.

As such, if I am successful in showing that PAP follows from a certain interpretation of "ought implies can" - and further, that this interpretation is both widespread and plausible on its own merits - then we will have found new reason to accept PAP, since it logically follows from a moral principle many of us rightly take for granted. This leaves the opponents of PAP with a difficult choice - to continue with their opposition, they will need to abandon at least one explanatorily powerful interpretation of OIC, and to accept the not insignificant problems that come with doing so. My wider objective is thus not merely to support a formal derivation of PAP from OIC, but also to explain how both principles fit into (and are a logical product of) a much wider pattern of moral thinking concerning the nature of obligations and blameworthiness.

The purpose of this initial chapter is primarily an expository one, and there are two topics I wish to cover before I move onto the work of the thesis proper. The first is to present a detailed summary of the history of the debate over PAP in Sections I - III, so that we may understand the current state of the principle and the wider context into which my attempt to link PAP with OIC fits. In Section IV and V I will turn to my second aim, where I set out in detail the central argument of this thesis - what I call the "Derivation" of the Principle of Alternate Possibilities from "ought implies can." This latter half of the chapter will also set out the precise structure of the thesis as a whole, to make clear how I intend to argue for each premise of the Derivation in the rest of this work.

But before I present my argument in favour of the Principle of Alternate Possibilities, I wish to examine how we arrived at the status quo that I intend to deviate from. How did PAP first become so unpopular? What are the major challenges which have been raised against it, over the decades, that we will need to counter or account for? And will the other defences of the principle over the years be of any help to us, at least as a starting point?

Section I.I: The Historical Debate

Though the Principle of Alternate Possibilities only gained its modern name when Frankfurt turned his attention to the subject, the debate over the necessity of alternatives to moral responsibility - the question, in short, of "could have done otherwise" - far precedes him. It has also been historically bound up with the related but much more general division between compatibilism and incompatibilism (in essence, whether free will is consistent with the theory of determinism), with compatibilist philosophers having been more willing to question or reinterpret the requirement of "could have done otherwise."

Though I think he is ultimately wrong in both his analysis of PAP and how it relates to the principle of "ought implies can", such is the seminal influence of Frankfurt's work on the current state of the debate over PAP that I have divided this review according to those works published *before* and *after* Frankfurt's attack upon the principle. This arrangement is intended to best highlight the ways in which the debate about PAP has changed over the years, particularly in recent decades as various camps have sprung up and strategies for attacking and defending the principle have adapted over time. In this section, I want to account for some of these historical contributors to sketch a picture of the consensus which existed pre-Frankfurt, so that it is clear just how it changed in the years thereafter.

Opposition to what we now know as PAP began, according to Robert Kane,[3] in the 17th century with the work of Thomas Hobbes. His debates with Bishop Bramhall (featured in, amongst others, "Of Liberty and Necessity",[4] "A Defence of True Liberty",[5] and "Questions Concerning Liberty, Necessity and Chance"[6]) contained not only one of the earliest forms of compatibilism, but also one of the first genuine attacks upon PAP. Hobbes was a determinist who, unusually by modern standards, regarded the classical idea of a free agent able to choose between various genuine alternatives as contradictory nonsense. Since it is obvious that at least some agents are morally responsible and/or

[3] Kane, Robert. 1985. *Free Will and Values* (Albany: State of New York Press).
[4] Hobbes, Thomas. 2008a. 'Hobbes' treatise of liberty and necessity', in *Hobbes and Bramhall on Liberty and Necessity*, ed. by Vere Chappell (Cambridge; Cambridge University Press), pp. 14-43.
[5] Bramhall, John. 2008. 'Selections from Bramhall, A Defense of True Liberty', in *Hobbes and Bramhall on Liberty and Necessity*, ed. by Vere Chappell (Cambridge; Cambridge University Press), pp. 43-69.
[6] Hobbes, Thomas. 2008b. 'Selections from Hobbes, The Questions concerning Liberty, Necessity, and Chance', in *Hobbes and Bramhall on Liberty and Necessity*, ed. by Vere Chappell (Cambridge; Cambridge University Press), pp. 69-91.

blameworthy for what they do, he claimed, the thesis that alternate possibilities are necessary for these concepts was therefore plainly false.

This line of questioning was later taken up by David Hume, another compatibilist who rejected the very idea of the ability to do otherwise. In his famous *"Treatise on Human Nature"*[7] and later *"Enquiry concerning Human Understanding"*,[8] Hume helped lay the groundwork for the classical compatibilist accounts of free will and responsibility. He did so by redefining free actions, as Hobbes had previously done, as those which are the product of an agent's wants and desires as opposed to choices between differing possible futures. His assault upon PAP is thus essentially the same as Hobbes' had been before him - that since the ability to do otherwise is impossible in any relevant sense, and agents are morally responsible for what they do, the former cannot be necessary for the latter.

Fast-forward to the 20th century, and opposition to PAP had started to become more nuanced. The increasingly popular compatibilist position had by now crafted two prominent forms of argument regarding the "could have done otherwise" question. The first, now rarer school joined Hobbes and Hume in attacking PAP on the grounds that the notion of being "able to have done otherwise" was either incoherent and/or simply not the relevant thing to be concerned with when judging an agent's moral responsibility[9]. G.E. Moore attempted the latter strategy in his *"Ethics"*,[10] broadly following Hume's account of free and responsible actions, and was followed up in this effort by Schlick's *"Problems of Ethics."*[11] Schlick's attack notably identified the curious problem of "exactly the same circumstances" in defences of PAP - for any point in time, if *all* the relevant circumstances (including my motivations, available information and so on) are assumed to be the same as they were when I acted, on what basis am I supposed to have been able to act otherwise than I did? What reason have we to think that if I were given a do-over, as it were, that the outcome would have been any different?

But a second type of strategy to account for - or depending on your point of view, undermine - PAP also emerged, and its line of attack was significantly more insidious. Rather than attempt to deny the importance of alternate possibilities, these new

[7] Hume, David. 2004. *A Treatise of Human Nature*, new edn (Mineola: Dover Publications, Inc.).
[8] Hume, David. 2008. *An Enquiry concerning Human Understanding*, ed. by Peter Millican (Oxford; Oxford University Press).
[9] The most successful example of this type of strategy is, of course, Frankfurt himself.
[10] Moore, G. E. 1912. *Ethics* (London; Williams and Norgate).
[11] Schlick, Moritz. 2008. *Problems of Ethics*, tr. by David Rynin (New York; Nielsen Press), Chapter 7.

compatibilists sought to redefine the meaning of "could have done otherwise" in such a way that the principle was consistent with determinism, and so moral agents in a deterministic world could still be meaningfully said to be able to act otherwise than they were determined to do. The purpose of such a strategy was simple: hitherto, PAP in its various interpretations had been solely the domain of incompatibilists - generally believers in so-called "libertarian" free will - and being able to call on the intuitive force of freedom being the choice between alternative futures was one of incompatibilism's greatest attractions. But if defensible accounts of alternate possibilities as consistent with determinism existed, as it turned out they did, then this intuitive advantage would be lost.

This type of strategy proved both popular and effective, and so the primary battleground over PAP quickly shifted from *whether* the principle was true to the question of *in what senses* it was true, and in particular whether any of those senses were consistent with a deterministic universe. One such competing analysis of "could have done otherwise" was provided by A.J. Ayer,[12] who endorsed the popular Humean analysis of free action mentioned above and used it as a basis for defining "could have done otherwise": "To say that I could have acted otherwise is to say, first, that I should have acted otherwise if I had so chosen; secondly, that my action was voluntary in the sense in which the actions, say, of the kleptomaniac are not; and thirdly, that nobody compelled me to choose as I did: and these three conditions may very well be fulfilled [if determinism is true]."[13]

The prevalence of this "redefining strategy" lasted several decades, and eventually led to a predictable and powerful backlash from incompatibilist supporters of PAP. In 1964 and 1966, Roderick Chisholm[14] and Carl Ginet[15] struck back against what they considered the hollow compatibilist senses of "could have done otherwise" with a compelling pair of arguments - compelling, at least, if one bore any sympathy whatsoever with the incompatibilist position. Chisholm's target were the conditional analyses which were, at that time, the broadly favoured descriptions by compatibilists. As mentioned before, on these accounts free actions are those which stem from our goals and desires, and we act freely so long as we are not frustrated in pursuing them. To say therefore that an agent

[12]Ayer, A. J., 'Freedom and Necessity', in *Philosophical Essays* (London; Palgrave Macmillan, 1954) pp. 271-284.
[13] Ayer (1954), p.278.
[14] Chisholm, Roderick. 1964. 'Human Freedom and the Self', in *Lindley Lectures* 4 (University of Kansas, Department of Philosophy).
[15] Ginet, Carl. 1966b. 'Might We Have No Choice?', in *Freedom and Determinisim*, ed. by Keith Lehrer (Random House) pp. 87-104.

"could have done otherwise" is merely to say that if they had wanted to do otherwise, they *would* have done otherwise. Chisholm demolished this position by pointing out that such analyses gave clearly incorrect verdicts in a certain subset of cases - cases where although it is true that *if* an agent desired A they would have done A, that agent is unable for whatever reason to desire A. Thus, the conditional compatibilist analysis became trapped as it had to say such agents could in fact do so. Yet it was (rightly) objected by Chisholm that this was ridiculous - how is it possible for one to do A if one cannot even do what is necessary for A?

Ginet's target was much the same, but he employed different tactics. Following up on the blow Chisholm had inflicted on compatibilist attempts of "could have done otherwise", he crafted what is now commonly called the Consequence Argument. Put simply, this argument claimed that since no-one at any given point in time has any power to alter the past and/or the laws of nature, and determinism by its nature holds that the past and the laws of nature together entail everything which will occur in the future, it follows that no-one at any point in time has ever had any power over the form the future will take. So it is madness, Ginet argued, to suggest we can retain the power to do otherwise than we actually do if determinism is true, since every action ever taken is ultimately the product of a deterministic chain of cause and effect stretching back to the beginning of the universe. By extension, PAP and determinism are fundamentally incompatible.

By the late 1960's, then, the advantage in the debate over PAP had largely shifted back to the incompatibilists. The preferred notion of free action preferred by compatibilists had suffered serious damage, and although the compatibilists had no choice but to reject the Consequence Argument it was far from clear there were philosophically sound grounds on which to do so. It seemed, then, that the relatively new strategy of taking the sting out of PAP by interpreting it in a light more sympathetic to determinism was a failure. Mercifully for the compatibilists, however, new life was about to be breathed into the original strategy of simply rejecting the principle altogether. It is now time to examine Frankfurt's revolutionary argument against PAP - one which proved more problematic for the incompatibilist position than any we have discussed so far.

Section I.II: Black and Jones: The Frankfurt-Style Cases

In 1969, Frankfurt published his "Alternate Possibilities and Moral Responsibility." I will cover the contents of this specific paper in more detail than the brief overview of the works discussed so far, so as to make clear its effects on the debate over PAP in the years which followed. Frankfurt believed that PAP was false no matter the chosen definition of alternate possibilities under discussion, and his goal was the same as Hobbes' or Hume's had been: to establish that regardless of one's understanding of the relevant terms in the debate, alternate possibilities are an irrelevant concern in determining what one is morally responsible (or blameworthy) for.

Frankfurt's central idea was that "there may be circumstances that constitute sufficient conditions for a certain action to be performed by someone and that therefore make it impossible for the person to do otherwise, but that do not actually impel the person to act or in any way produce his action."[16] In other words, there is a difference between a state of affairs which simply mandates that an action will be performed (i.e. that constitutes sufficient conditions for it), and a state of affairs which *influences* or *encourages* a person to act in that specific way. In many cases where we would naturally invoke PAP to exempt an agent from blameworthiness or moral responsibility, Frankfurt argued, the former is present but the latter is not. As such, the agents in these cases are left with no choice but to act as they do, but they may still be motivated to act thus as a result of their own decisions, beliefs, and so on: factors which have nothing to do with the fact that they cannot do otherwise, in whatever sense we take that to mean.

If cases such as these are plausible, Frankfurt claimed, then we have reason to doubt that it is *PAP* which exempts an agent from responsibility or blameworthiness in situations such as these, where we would intuitively feel an agent is not at fault for their action(s). He suspected that PAP had piggy-backed on the work of other principles which Frankfurt was more sympathetic toward, such as the idea that coercion would remove one's moral responsibility. Since in Frankfurt's mind it was certainly true that not all potential cases where an agent cannot act otherwise are the result of coercion (et al), it is only logical that he would conclude that PAP was suspect.

[16] Frankfurt (1969), p.830.

The kind of example Frankfurt had in mind for undermining PAP would go on to become infamous in the literature of free will and responsibility. These Frankfurt-cases (hereafter, FFCs) would be endlessly presented, attacked, and refined in a constant dialogue between supporters and critics of Frankfurt's argument, and we will study that dialogue in the next section. But here is the original case that spawned the legion of variants, which Frankfurt believed was sufficient to prove PAP was false:

"Suppose someone - Black, let us say - wants Jones to perform a certain action…he waits until Jones has made up his mind about what to do, and he does nothing unless it is clear to him (Black is a very good judge of such things) that Jones is going to decide to do something **other** than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to, and that he does do, what he wants him to do. Whatever Jones' initial preferences and inclinations, then, Black will have his way."[17]

Here we have a case where all parties agree Jones cannot do otherwise than Black wants him to do, for Frankfurt invites us to use our own understandings of "could have done otherwise" and "effective steps" to determine exactly what Black would have to do to ensure Jones' compliance. If Jones attempts to do otherwise than Black wishes, the verdict in terms of blameworthiness and moral responsibility is clear: Jones will not be responsible for his eventual action, because he has been coerced into doing it by Black's intervention rather than anything on his own part. Note, however, that even here the absence of alternate possibilities on Jones' part is irrelevant on Frankfurt's view.

But what if, Frankfurt asked, Black did *not* have to intervene? What if Jones just decided on their own to do what Black wanted, and no meddling on Black's part were required? Now the case becomes interesting. For Frankfurt, the answer is just as clear as before: Jones is obviously responsible (and/or blameworthy) for their actions in this case, alternate possibilities be damned. Specifically, they are just as responsible for their actions as they would have been if Black did not exist and thereby Jones *had* been free to act otherwise. This is because everything which caused and motivated Jones to act in this case is identical, given that Black does nothing, to what would have caused and motivated Jones to do that selfsame thing in the world where Black is absent. The absence of

[17] Frankfurt (1969), p. 835.

alternate possibilities, Frankfurt maintained, played no role in what lead Jones to decide as he did and to act on that decision. He would have done exactly the same thing, for the exact same reasons, whether they possessed alternate possibilities or not. Those possibilities are therefore irrelevant to the question of Jones' responsibility or to the question of whether he warrants praise or blame, and PAP is false. For how could Jones' lack of alternate possibilities be the crucial factor in determining their responsibility, Frankfurt claimed, if everyone agrees that that fact had nothing to do with why Jones acted as he did?

The impact of this argument was immediate and revolutionary. Each defender of PAP, up to and including myself, now had a dangerous challenge to answer regardless of their own preferred understanding of the principle. What's more, it suggested a startling point about the debate between compatibilists and incompatibilists more generally: if determinism is inconsistent with our being morally responsible, praiseworthy and blameworthy agents, as has long been argued, it will not be thus because it threatens our ability to do otherwise. Hence, arguments that the ability to act otherwise can be retained in a deterministic universe would be not just ineffective, as we had noted in the previous section, but irrelevant to the *real* question of what is required to be a morally functioning agent. Let us now, therefore, turn our attention to those reactions to Frankfurt, and see how the debate about PAP evolved following his initial work on the subject.

Section I.III: PAP, from Frankfurt to Present

For the rest of the 20th century, the debate over PAP and its importance to the questions of moral responsibility and blameworthiness had found a new central question: was Frankfurt right, or at least onto something important, in claiming alternate possibilities were irrelevant? Many compatibilists, tired of being on the defensive for so long, eagerly jumped on Frankfurt's line of argument and a new consensus formed shortly thereafter. An enormous amount of material in support of Frankfurt's attack on PAP was produced during this period, and so for brevity's sake I will note only a few of the highlights - Davidson's "Freedom to Act"[18] (1973) Dennett's "Elbow Room"[19] (1984), and

[18] Davidson, Donald. 2001. 'Freedom to Act', in *Essays on Actions and Events*, 2nd edn (Oxford; Clarendon Press), pp. 63-82.
[19] Dennett, Daniel C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting* (The MIT Press).

Zimmerman's "Acts, Omissions, and Semi-Compatibilism"[20] (1994) being particularly popular.

Of course, against this legion of imitators and supporters an equally numerous army of critics of FFCs (in both their original and refined forms) eventually assembled, and it was with the advent of these new critics as the debate as it stands today took form. Broadly speaking, two significant camps of objection to Frankfurt's argument developed among supporters of PAP, and I will cover each in turn here.

On the one hand, some objected that the basic circumstances set out in the Frankfurt-style objections are incoherent when applied to an indeterministic universe. This is the line originally adopted by Robert Kane in his *"Free Will and Values"*, and later taken up by Ginet,[21] Widerker,[22] and Keith Wyma.[23] This "Indeterminist World" group of objectors holds that if the world is indeterministic it will not be possible for Frankfurt's hypothetical intervener to lurk unseen before affecting an agent's judgement at precisely the right moment, as is part of the basic structure of such cases. This is because the choices of an indeterminstic agent, occurring as the result of whatever non-deterministic factors may be relevant in the case at hand, cannot *in principle* be predicted even knowing their casual history and the full laws of nature (as they can be in the deterministic world). Because there is no such "prior sign" that gives away Jones' decision Black cannot possibly be in a position to intervene, whatever the specifics of the FFC, *before* Jones has reached their decision about what to do. If the world is indeterminstic, it was maintained, Black cannot know what Jones will do until Jones decides to do it. The earliest Black can intervene is therefore *after* Jones has reached their decision about what to do, because that is the earliest point Jones' action can be predicted.

This line, if correct, cripples the FFCs because now there is a problem: assuming Jones decides to do what Black wants without Black's intervention, Jones will indeed be morally responsible for that decision and the resulting action. But this will, the objectors

[20] Zimmerman, David. 1994. 'Acts, Omissions and "Semi-Compatibilism"', *Philosophical Studies*, 73, 2-3: 209-223.
[21] Ginet, Carl. 1996a. 'In Defence of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing', *Philosophical Perspectives*, 10, 403-417.
[22] Widerker, David. 1995. 'Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities', *The Philosophical Review*, 104, 2: 247-261.
[23] Wyma, Keith D. 1997. 'Moral Responsibility and the Leeway for Action', *American Philosophical Quarterly*, 34, 1: 57-70.

maintain, be because Jones' decision was undetermined and they therefore had the power to do otherwise at the relevant point (when the decision was made, before Black could have intervened). Such a case is thus not a counterexample to PAP. And if Black does intervene to force Jones' hand after Jones has decided otherwise than Black wants, then by Frankfurt's own admission that is coercion and Jones will *not* be responsible for their eventual action. Whichever way, therefore, PAP stands. However, despite the cleverness of this line of argument it was commonly rejected as a successful defence on the grounds that it begs the question. In other words, it was (and is) simply disputed by compatibilists and other critics of PAP that an indeterministic agent's decisions are so totally unpredictable. For example, Frankfurt himself, Mele and Fischer pointed out that Black might still be able to predict with reliability what Jones will choose to do even if that choice is *not* the result of deterministic processes - say, if they were possessed of inhuman observational powers concerning Jones' indeterministic thought processes.

The other major category of objection to FFCs, and the one I personally favour, is the "flicker of freedom" defence. First popularised by van Inwagen in 1978,[24] its essence is this: for FFCs to serve as true counterexamples to PAP, they must show that there are possible situations in which 1) an agent cannot do otherwise than they in fact do and yet 2) they remain morally responsible or blameworthy for having done that specific thing. Despite appearances, it is objected, FFCs fail to do this. They certainly identify a case where Jones cannot do otherwise than a specific thing - to act as Black wishes - and a thing which Jones is at fault for, but these are, crucially, two different things. We agree that Jones is morally responsible for his actions, good or evil, in the case where he acts of his own accord to do what Black wants. But doing what Black wants *of Jones' own accord* is a different matter than simply *doing what Black wants,* and here the Frankfurt case comes apart. For it is obviously *not* the case that Jones cannot do otherwise than "doing what Black wants of his own accord". It is inherent in the makeup of FFCs that Jones can attempt to act otherwise than Black desires, which is what would prompt Black to intervene.

This ability to resist Black's desired course of action - ultimately useless, but clearly present - is Jones' "flicker of freedom". It grants him the ability to do otherwise in reference to the specific thing we agree he is morally responsible or blameworthy for -

[24] Van Inwagen, Peter. 1978. 'Ability and Responsibility', *Philosophical Review*, 87, 2: 201- 224.

acting as Black wants, but of his own will. So like the Indeterminist World argument, it is claimed, FFCs fail because all they accomplish is to identify another situation where someone is morally responsible for something for which they did have the power to act otherwise.

Of the two defences detailed here, the "flicker of freedom" proved the more generally popular primarily due to its lack of other philosophical commitments. It retained a surprisingly sizeable following throughout the post-Frankfurt decades, including Naylor (1984),[25] Pettit (2001),[26] and Speak (2002).[27] However, as with the Indeterminist World argument it did not take long for the compatibilists to strike back. The most dangerous attack on the "flicker of freedom" defence of PAP came from the man who popularised its name, the compatibilist J.M. Fischer in 1994.[28] Fischer was prepared to grant that the alternate possibilities the defence identifies in the FFCs are legitimate, and therefore that they technically fail to demonstrate an act for which one is both morally responsible and cannot do otherwise than. He still, however, rejected the flicker defence as a basis for PAP on its own merits, for a simple reason: the alternatives identified by the flicker defence are not sufficiently "robust", in his terminology, to serve as the basis for moral responsibility which would otherwise be absent. The flicker defence, Fischer argued, is incomplete without an explanation of the role that these alternate possibilities play in making one deserving of responsibility and blame.

Furthermore, it was unintelligible to him that *these* possibilities in particular - the possibilities that are supposed to make the difference in the FFCs - should be the ones that play this role. Ordinarily, we on the pro-PAP side think of the type of control over our actions that is required for responsibility in terms of the "garden of forking paths" - distinct alternatives open to us which we have the power to choose between. This is what Fischer refers to as "regulative control", as opposed to the "guidance control" favoured by compatibilists more generally in which the source of one's free actions is the agent's own will. Yet the alternate possibilities highlighted by the flicker defence are not like this,

[25] Naylor, M. B. 1984. 'Frankfurt on the principle of alternative possibilities', *Philosophical Studies*, 46, 249-58.
[26] Pettit, Philip. 2001. 'The Capacity to Have Done Otherwise', in *Relating to Responsibility: Essays in Honour of Tony Honoré on his 80th Birthday*, ed. by Peter Cane and John Gardner (Oxford: Hart Publishing) pp. 21-35.
[27] Speak, Daniel. 2002. 'Fanning the Flickers of Freedom', *American Philosophical Quarterly*, 39, no. 1, 91-105.
[28] Fischer, John Martin. 1994. *The Metaphysics of Free Will; An Essay on Control* (Oxford; Blackwell Publishers Inc.).

since they identify alternatives in which the agent does *not* freely undertake a different course of action at all. They merely showcase an alternate possibility where that agent performs the *same* course of action decidedly *less* freely - that is, as the result of coercion (Black's nefarious work rather than Jones' own will). How then, Fischer asked, do these flicker possibilities add to Jones' level of *control* over his eventual action in such a way as to make him morally responsible for it? Especially given that it is not up to Jones, being ignorant of Black, which of these alternate possibilities will even be actualised?

Other critics of this vein included Robb/Mele[29] and Eleanor Stump,[30] who objected to the distinction of "doing X" from "doing X of one's own free will/own accord/etcetera." For the latter, it was argued, is not a distinct action warranting its own separate considerations of responsibility, but simply one of the factors which should be taken into account when morally assessing the former action. In what would become a frustratingly common occurrence among critics of PAP from this point forward, it was also suggested (contra Fischer) that the flicker defence *can* be outwitted by simply rewriting the Frankfurt - examples in such a way as to omit the possibility of any such flickers of freedom - perhaps by making it so that Black is prepared to intervene earlier, when Jones is still forming their decision.

But the defenders of the flicker of freedom did not give up so easily, and for each new variant of the Frankfurt case which purported to be the one which finally ended PAP for good, its defenders would find holes, inconsistencies and fledgling alternate possibilities to put them down. For instance, Capes (2014)[31] offers a specific and lengthy retort to Stump's modified FFCs, and Janzen (2013)[32] attacks the general methodology used to undermine the flicker defence of removing the "prior sign" component of traditional FFCs (i.e. that which prompts Black to make his move on Jones). At the same time, Robinson (2012)[33] took aim at Fischer's attack on the flicker defence in its entirety in a well-intentioned, but incomplete attempt to put a stop to the fine-tuning of the FFCs altogether.

[29] Mele, Alfred R. and David Robb. 1998. 'Rescuing Frankfurt-style cases', *Philosophical Review*, 107, 1: 97-112.
[30] Stump, Eleonore. 1999. 'Alternative Possibilities and Moral Responsibility: The Flicker of Freedom', *The Journal of Ethics*, 3, 4: 299–324.
[31] Capes, Justin A. 2014. 'The Flicker of Freedom: A Reply to Stump', *The Journal of Ethics*, 18, 4: 427-435.
[32] Janzen, Greg. 2013. 'Frankfurt Cases, Alternate Possibilities and Prior Signs', *Erkenntnis*, 78, 5: 1037-1049.
[33] Robinson, Michael. 2012. 'Modified Frankfurt-type counterexamples and flickers of freedom', *Philosophical Studies*, 157, 2: 177-194.

At this point, the debate about whether Frankfurt was right that PAP is false, and of the relevance of alternate possibilities to responsibility and blameworthiness in general, collapsed into an endless sequence of counterexample and criticism. PAP's defenders continued to find holes, however small or intuitively irrelevant, in the latest types of Frankfurt-case which in turn were refined by compatibilists and other critics of the principle to escape them, and the circle continued ever unbroken. This dance would eventually grow so convoluted that it began, in the minds of some contributors, to lose sight of what made the original Frankfurt-style objections so powerful in the first place. Instead, as Simkuhlet (2015) despaired the so-called "Neo" FFCs became so obsessed with proving PAP as written false that they were "silent on the big metaphysical issues at the centre of the free will debate."[34]

By the early 21st century, this feeling of disillusionment with the Frankfurt debate led to the growth of a new body of writing, focusing on approaching (or attacking) the defence of PAP from more original directions. These efforts to circumvent the deadlock I have just described are the last category of historical work on PAP which remains for us to complete the picture of the debate as it exists today and are perhaps, in many regards, the most interesting.

In 1997, David Copp attempted to sidestep the now-traditional debate on FFCs by adopting a new tactic. In his paper "Defending the Principle of Alternate Possibilities: Blameworthiness and Moral Responsibility,"[35] and later "Ought Implies Can, Blameworthiness, and the Principle of Alternate Possibilities,"[36] and "Ought Implies Can and the Derivation of the Principle of Alternate Possibilities,"[37] Copp argued in favour of PAP on the basis that PAP ought to be understood as a subsidiary principle of an idea which most of PAP's critics already accepted (that "ought implies can"), and we therefore ought to understand the "can" of OIC and the "able" of PAP as referring to the same

[34] Simkulet, W. 2015. 'On the Signpost Principle of Alternate Possibilities: Why Contemporary Frankfurt-Style Cases are Irrelevant to the Free Will Debate', *Filosofiska Notiser*, 2, 3: 107–120.
[35] Copp, David. 1997. 'Defending the Principle of Alternate Possibilities: Blameworthiness and Moral Responsibility', *Noûs*, 31, 4: 441-456.
[36] Copp, David 2006. 'Ought Implies Can, Blameworthiness, and the Principle of Alternate Possibilities', in *Moral Responsibility and Alternative Possibilities* (Aldershot: Ashgate Publishing Ltd), pp. 265-301.
[37] Copp, David. 2008. ''Ought' Implies 'Can' and the Derivation of the Principle of Alternate Possibilities', *Analysis*, 68, 1: 67-75.

notion of morally relevant ability. This followed Widerker[38] having independently reached this same conclusion. Whilst these papers received comparatively little attention at the time of their publication, in part because they failed to present a complete argument for what this supposedly shared sense of "can" would be, Copp's work in this area is amongst the most influential on the overall purpose and content of this thesis. As we will see later, both its central idea of tying PAP to a version of OIC and the ideas which will eventually ground the links between the two principles are based in the original arguments Copp made toward the same conclusion I am attempting here.

In 2010, Larvor took the more unorthodox approach of defusing the FFCs by denying not simply that Jones is responsible for the action they cannot do otherwise than, as the ordinary PAP defender would, but also that the action for which Jones is supposedly responsible is *his* action to begin with. He claims the schemes put into place by Black mean the relevant action in this case (the one Jones cannot do otherwise than), is in fact an action of Black's, and all associated responsibility and blameworthiness fall on him. Larvor denies that Jones is responsible for the unavoidable act because, as he puts it "This is Black's deed. The fact that the causal chain passes through parts of Jones's body does not make it Jones's deed. Jones is merely Black's unwilling instrument…which is why we would hold Black responsible."[39]

This is also the position taken by Alvarez,[40] who regards FFCs as structurally incoherent - in such cases, Black's power is necessarily assumed to be sufficient to determine Jones' decision to act in a certain way, and hence ensure he *will* act in that way, and Alvarez denies such a thing is possible. For although Black's power may be strong enough to cause Jones' brain and body to do the things Black wishes, she argues, there is no reason to think this is a decision of Jones' and so no responsibility should attach to him for the resulting action. If this line is correct it is obvious how FFCs would fail to undermine PAP, since the unavoidable action in the case and its associated moral responsibility now do not just belong to different actions, as in the flicker of freedom defence, but to different agents altogether.

[38] Widerker, David. 1991. 'Frankfurt on 'Ought Implies Can' and Alternative Possibilities', *Analysis*, 51: 222-224.
[39] Larvor, Brendan. 2010. 'Frankfurt counter-example defused', *Analysis*, 70, 3: 506-508.
[40] Alvarez, Maria. 2009. 'Actions, thought-experiments and the 'Principle of alternate possibilities'', *Australasian Journal of Philosophy*, 87, 1: 61-81.

Alongside these new defences of the traditional form of PAP would also come attempts to, unusually, meet its critics halfway. An example of this method is Whittle's "Ceteris Paribus, I Could Have Done Otherwise,"[41] born from a desire to keep the essence of PAP whilst also conceding that Jones is morally responsible in the FFCs. Whittle's replacement principle, the "Ceteris Paribus" PAP, redefines its scope by changing the ability to do otherwise from a necessary condition on all moral responsibility to a more limited constraint - one which applies "ceteris paribus", or "under normal or typical circumstances." The basis of this new PAP is that although Frankfurt succeeds in proving that PAP is not an exceptionless principle, he fails in the more important task of showing that alternate possibilities are *irrelevant* in our accounts of moral responsibility. PAP is thus understood as a generalisation about how moral responsibility works, not an absolute - but a true and explanatorily useful generalisation nonetheless. The historical division between pro and anti-Frankfurt camps, Whittle maintains, is a false dilemma. One can and should accept the intuitive force of both.

This tactic of replacing PAP with a structurally similar principle immune to FFCs is repeated by Young,[42] who opts to radically redefine PAP's supposed purpose in such a way as to sidestep them entirely. Like Whittle, Young is persuaded of the intuition behind the FFCs that Jones is morally responsible for their unavoidable action but is loath to abandon PAP altogether. Young's PAP is a sufficient but deliberately not necessary condition for moral responsibility (et al). Rather, what is needed for an agent to be properly responsible or blameworthy for their actions, it is argued, is that either PAP is satisfied or what Young calls the "twin world" condition is satisfied. This latter condition is met if and only if the existence of alternate possibilities holds no weight in what an agent did and/or what motivated them to do what they did, as in the FFCs, and so both the action and its causes are the same in each "world" under consideration. Once this move is made the defender of PAP has nothing to fear from FFCs, for what they are defending has fundamentally changed: now that they are interested in just one of multiple ways in which someone might come to be morally responsible for what they do, it no longer matters that the "PAP description" of the necessary conditions for responsibility is not applicable to Jones.

[41] Whittle, Ann. 2016. 'Ceteris Paribus, I Could Have Done Otherwise', *Philosophy and Phenomenological Research*, 92, 1: 73-85.
[42] Young, Garry. 2016. 'The Principle of Alternate Possibilities as Sufficient but not Necessary for Moral Responsibility: A way to Avoid the Frankfurt Counter-Example', *Philosophica*, 44, 3: 961-969.

With this, we come to the end of our discussion of history. Having now thoroughly examined the debate over PAP, it should be apparent why I have sought this path in my attempts to defend the principle in this thesis. It has been nearly fifty years since Frankfurt's "Alternate Possibilities and Moral Responsibility" was published, and yet we have made very little ground in figuring out whether his cases do, in fact, serve as counterexamples to PAP. More depressingly, it is clear to me that the conversation about PAP has long since stagnated. To the credit of the more recent writers on this subject, I feel I am not alone in this assessment of the debate. Whilst I do not deny that Frankfurt's work has warranted an extensive discussion, it is also clear to me that we have wrung the juice from this particular fruit. As I think Vihvelin rightly puts it, "Inevitably, the discussion turns to an argument about which side has the burden of proof, always a sure sad sign of a philosophical impasse. I think we should have avoided this mess. Things went wrong from the start."[43]

Therefore, I suggest a change of tactics from those who support PAP is in order. As today's writers on the subject have realised, it is not enough for any principle (however plausible) to simply be defended from the charges that have been laid against it. A positive case must be made in favour of that principle, and reasons provided to motivate it which do not simply appeal to those whom, in this case, are already sympathetic to the libertarian position. Nor, however, does the more recent trend of defending PAP by retreating from the universal scope that has been historically associated with the principle appeal to me. Interesting though the approach of a more limited PAP taken by Whittle and Young is, I cannot help but feel they have cured the disease of Frankfurt-style objections by euthanising the patient. Conceding that PAP is not a necessary requirement for blameworthiness is, I feel, tantamount to conceding the argument over the relevance of alternate possibilities in metaethics as a whole. As such, the task of this thesis will be to find a more productive method of approaching the same objective as the historical defenders of PAP, and thus argue for the principle as a necessary and universal constraint on an agent's blameworthiness for any particular action.

Given the longstanding trench warfare between the two sides of the debate that I have just documented, one might blanch at the prospect of crafting a positive case for it that could potentially be acceptable to both. Regardless of their view on PAP itself, though, most of

[43] Vihvelin, Kadri. 2000a. 'Freedom, foreknowledge, and the principle of alternate possibilities', *Canadian Journal of Philosophy*, 30, 1: 1-23 (p.8).

the parties engaged in the debate over the merits of the principle do possess some common ground. OIC is a popular moral belief among compatibilists and incompabilists alike, who may each accept that an action is only genuinely morally obligatory if it "can" be carried out even whilst disagreeing about the precise meaning of that "can." Moreover, many of the compatibilists that have supported OIC, including Frankfurt himself, explicitly reject PAP. Here then is our way in: following Copp, I believe that PAP in respect to blameworthiness follows deductively from a certain interpretation of OIC. If we can craft a version of OIC from which we can deduce PAP *and* which is also acceptable to a compatibilist audience, therefore, we have (in principle) a way of breaking the deadlock over PAP by grounding it in something which a good many compatibilists already believe. That this approach also sidesteps the entire issue of FFCs which has dominated the discussion of PAP for the past half-century is a welcome bonus.

This will be no easy task, and more will need to be said about how it might be accomplished before I can proceed with the main argument of this thesis - what I have termed, borrowing Copp's language, the "Derivation" - and show how we might argue from OIC to PAP. We must now sadly leave behind our discussion of history, and concern ourselves with improving the present.

Section I.IV: The Derivation

At first glance, it might seem odd to suggest that PAP is a subsidiary principle of "ought implies can." The Kantian principle is fundamentally concerned with the nature of our moral obligations and PAP, as I want to argue for it, is a principle about blameworthiness. In addition, whilst the two principles are linked in that they reference an agent's ability - the "able" of "able to complete an obligation" (OIC) and of "able to do otherwise" (PAP) - they appear to deal with conditions for different subjects. As Frankfurt put it, "the relation between Kant's doctrine and PAP is not as close as it seems to be. With respect to any action Kant's doctrine has to do with the agent's ability to perform *that* action**.** PAP, on the other hand, concerns his ability to do *something else.* Moreover, the Kantian view leaves open the possibility that a person for whom only one course of action is available...is morally praiseworthy for doing so. On the other hand, PAP implies that such a person cannot earn any moral credit for what he does. This makes it clear that

renouncing PAP does not require denying "ought implies can" and that PAP is not entailed by the Kantian view."[44]

Needless to say, I disagree. To that end, this thesis seeks to defend two claims:

I) It is possible, given certain connecting premises, to derive the principle of alternate possibilities (PAP) from a certain interpretation of "ought implies can" (OIC).

II) "Ought implies can" in that certain interpretation is true, as are the aforementioned connecting premises. Therefore, PAP as regards blameworthiness is also true.

In what remains of this chapter, I will show that I) is correct - there exists a logically valid argument deriving PAP from OIC. In the remainder of the thesis as a whole I will argue that II) is also correct; that this logically valid argument is also philosophically sound. Having tied the two principles together in this way, I will have shown that anyone who accepts OIC in the sense which I understand it will also be committed to PAP, and additionally that the version of OIC I have in mind is a plausible and defensible principle on its own merits. As such, the ultimate conclusion of this work is not only that PAP in regard to blameworthiness follows from a certain interpretation of OIC, but a version of OIC which we *ought to accept* - and thus, that PAP should be accepted as well.

Before we proceed, I wish to clarify certain things about the parameters of the Derivation I will argue for. Although I have said that I believe PAP to follow from a particular interpretation of OIC (the details of which shall be provided in Chapter II), that interpretation is not exclusionary. I believe the sense of OIC that I will soon argue for to be one of several ways in which the principle of "ought implies can" is true, and most certainly not necessarily the *only* sense in which OIC holds. I am not (presently) interested in writing a treatise on which interpretations of the principle are true and which are false. It may or may not be the case that there are other valid interpretations of OIC besides the I one I will defend here, but I wish to stress that this question is beyond the scope of this thesis.

[44] Frankfurt, H.G. 1998. 'What We Are Morally Responsible For', in *The Importance of What We Care About* (New York; Cambridge University Press), 95-104 (pp. 95-96)

In addition, whilst most writers on the subject have chosen to make PAP a much broader principle about moral responsibility in general, I am only concerned with the principle as it pertains to an agent's blameworthiness for their actions. Nor do I intend for PAP as I defend it to apply to the concept of "agent-regret", most often found in the discussions of moral luck popularised by Bernard Williams[45] and Thomas Nagel.[46] This is, briefly, the idea that in many situations it is not just acceptable for an agent to feel regret, shame, or otherwise morally at fault over an action(s), but also warranted (in the normative sense). It may turn out that the arguments I will advance in defence in PAP in respect of blameworthiness can also be adapted to reference our more general moral responsibility et al, but this is not my goal and any arguments which tend in that direction are incidental. It is clear to me what is wrong with the moral picture in which PAP does not factor to assessments of blameworthiness, and my Derivation's sole purpose is to resolve that mistake. The focus of this thesis is narrowed accordingly.

Finally, my Derivation is designed to stand neutral regarding the much wider debate about whether moral responsibility, blameworthiness, praiseworthiness and so on are consistent with determinism. If I am successful, I believe it will be possible to interpret my claims about the morally relevant abilities, blameworthiness and the connection between OIC and PAP in ways which are consistent with both compatibilism and incompatibilism. I have taken great pains in what follows to choose definitions of the terms which the two camps so often clash over that could, at least in principle, be interpreted in favour of either of them. I am only concerned with one relatively small part of the so-called "free will" question - the necessity of the ability to do otherwise for culpability and blameworthiness - and I intend that determinists be able to accept my arguments without any inconsistency in their overall position.

Here is my Derivation of the Principle of Alternate Possibilities:

1) "Ought implies can": an agent may only be morally obligated to perform (or refrain from performing) an action if they are able to perform (or refrain from performing) that action.

[45] Williams, Bernard. 1981. *Moral Luck; Philosophical Papers 1973-1980* (Cambridge; Cambridge University Press).
[46] Nagel, Thomas. 1979. 'Moral Luck', in *Mortal Questions* (Cambridge; Cambridge University Press) pp. 24-39.

2) If an agent is morally blameworthy for an action, that action is objectively morally wrong.

3) If an action is objectively morally wrong, we are morally obligated not to do it.

4) Thus, an agent is morally blameworthy for an action only if they were able not to do it (from 1 - 3).

5) If an agent who performs an action was able not to do it, they were able to act otherwise than they did (even if this is just to do nothing).

6) Thus, an agent is morally blameworthy for an action only if they could have acted otherwise than they did (from 4, 5).

Conclusion: PAP in regard to blameworthiness is true.

As mentioned in Section III, this Derivation maintains the same essential structure as the one defended by David Copp in his three papers on the subject. Copp was attempting to construct an argument that deduced PAP from a version of OIC, yet by his own admission his work was incomplete. I intend to use his arguments in favour of the Derivation you see above as a starting point in making my own case for its soundness. One particularly useful observation taken from Copp's prior efforts at deriving PAP, I should mention now: for the purposes of this argument, whatever we take the morally relevant sense of "can" or "able" to be for OIC, it ought to be understood as the same sense of "able" used in PAP in respect to blameworthiness (which states, remember, that one may only be blameworthy for performing or not performing an action if they were "able" to do otherwise than it).

Only if this unstated premise holds will be the steps made in 1-6) be logically valid. For as noted previously, the Derivation is predicated on the assumption that the senses of "able" used in PAP and OIC are identical. If they are not, the move from 4) to 6) collapses as it would then be possible for an agent to possess one principle's relevant type of ability, but not the other. If such a thing is possible, my Derivation would fail in its intended goal of tying the adherents of (a certain version of) OIC inescapably to PAP. We will discuss this question in Chapter IV, where an argument will be made that we have good reason

to believe these two principles are interested in the same kind of normatively relevant ability, since both are sourced in and motivated by the same core moral principles. But once we have granted this premise, Copp argues, PAP will follow from OIC without issue:

"Suppose that a person (M) is blameworthy for doing something (A). It follows, on plausible assumptions, that what M did was wrong, and from this it follows that she was morally required not to do A. But then, given the Maxim ["ought implies can"], if M was morally required not to do A, she was able not to do A. Therefore, a person is morally blameworthy for doing A only if she was able not to do A. But if a person who did something was able not to do it, she was able to do otherwise than she did. It follows, then, that a person is morally blameworthy for doing something only if she could have done otherwise. That is, PAP with respect to blameworthiness follows from the Maxim."[47]

Section I.V: The Thesis Structure

At this stage, we now possess a logically valid argument for deriving PAP from OIC. Now I shall explain how I intend to argue for the soundness of my Derivation, for it has many enemies if we accept the (non-trivial) claim that the senses of "can" and "able" used in OIC and PAP should be viewed identically. The obvious first point of attack would be to deny 1) outright and reject OIC altogether, or argue that it is only true in a sense which does not grant 4) and 5). Given my claim that my preferred version of OIC is merely one among many true interpretations of the principle, there is certainly intellectual room for this latter sort of criticism and we will examine an objection of this type in Chapter III. If either of these were successful, PAP would simply be a false principle deduced from another false principle and the Derivation could not even begin.

2) and 3) can be denied, though not without some difficulty. One might, for instance, reject 3) by asserting that one can act wrongly and hence be blameworthy for performing an action even though they have no moral obligation to avoid doing so, thus severing the link between "ought implies can" and PAP. One could also opt for the similar plan of denying 2) and asserting that one can be blameworthy for acts which are not themselves

[47] David Copp (2006) p.270.

morally wrong. 4) and 6) are purely deductive and cannot be directly attacked, but the move between them can be undermined by taking Yaffe's[48] line that a moral obligation not to do something does not automatically translate into a moral obligation to do something else instead. If true, then even given OIC it will not automatically follow that we will be able to do otherwise than a particular action simply by being obligated not to do it. Thus 5) and by extension 6) are challengeable, for once an obligation to avoid performing a certain action no longer mandates (via OIC) that we have the power to *act otherwise* than that action - merely the ability to refrain from performing it - an agent could potentially be blameworthy for failing in their moral obligation despite lacking the ability to do anything else instead.

What should we do, when faced with this dizzying array of challenges? Rather than attempt to rebut every possible objection to the argument I have laid out above, my plan will be to tackle the case in favour of the Derivation systematically. Going forward, each of the premises which compose it will be individually assessed, and a case shall be laid out in their favour that engages with and accounts for a variety (though, as said, by no means all) of the criticisms which may be made against them in turn.

In Chapter II I will present my chosen interpretation of OIC alongside a detailed discussion of the nature and purpose of our genuine moral obligations. This discussion will then be used to argue for the specific variation of OIC from which I intend to derive PAP in regard to blameworthiness, and also includes a thorough account of the sense of morally relevant ability which I believe is required by OIC and PAP alike. In Chapter III, I then turn to examining the most pressing challenges to my account of OIC and the underlying motivations which lead me to favour it, including attacks on how I interpret the "implies" and "can" portions of the principle. Once my defence of my preferred formulation of OIC is completed, Chapter IV will move onto what I have called the "connecting premises" of Derivation - 2), 3), and 5) - and explore some ways in which each of them may be challenged before offering corresponding reasons in their favour.

In Chapter V, I devote considerable attention to a matter that, whilst not an explicit premise of the Derivation itself, factors so heavily into my overall argument in support of

[48] Yaffe, Gideon. 1999. ''Ought' implies 'can' and the principle of alternate possibilities', *Analysis*, 59, 3: 218–22. See also Graham, Peter. 2011. 'Ought and Ability', *Philosophical Review*, 120, 3: 337-382 for a similar line of argument.

it that I consider it to be the most important piece of the reasoning which grounds my interpretations both of OIC and PAP: the question of whether morality and our moral obligations are *fair*, and what fairness in this context actually means and requires. As we shall see in the next chapter, the claim that our genuine moral duties must be fair to us occupies an important role in my defence of "ought implies can" and since the Derivation as a whole is dependent on the truth of OIC as I wish to understand it, this "Fairness Principle" therefore provides a great deal of the grounding for my project as a whole. Indeed, it is often claimed as a point in favour of various forms of OIC (as I intend to do) that without such a principle, an agent who was unable to complete their moral obligations might be blamed, and more worryingly *deserve* blame, for their inevitable failure to do so. Because we judge this would be unfair to the agents under those duties, many believe such obligations to be impossible.

Yet this intuitive assumption may be called into question, and if it is wrong - if our genuine moral obligations could be unfair to their agents in the sense I am interested in - then the case in favour of OIC will be sorely weakened, and my overall argument along with it. Through my defence of this claim, I will explore the broader view of the nature and purpose of our moral obligations that lies at the heart of the Derivation, and show how it provides a strong foundation on which to not only base my particular interpretation of OIC, but also why we should accept the type of ability I will put forward as the morally relevant one whether we are assessing whether an agent is "able" to fulfil their moral obligation, or whether they are "able" to act otherwise than they in fact do.

In this chapter, I have sought to provide an outline of the thesis as a whole. Its purpose has been to inform and prepare, not to defend. It should now be clear where my project stands in relation to the history of the Principle of Alternate Possibilities, the conclusions I am seeking to accomplish in the course of it, and how exactly I intend to accomplish the goals I have set out on this chapter. Whilst there have been advocates of the Derivation of PAP in the past and more, no doubt, in future, I believe that my work here represents the most rigorous and well-grounded attempt at a strategy of this sort that I have encountered. My work here will take the best of the arguments that have been made previously and mould them into a more well-rounded and defensible position, one which goes beyond the defence of OIC and PAP to seek to understand the broader observations about the character of our moral obligations and blameworthiness which underpin them.

Now I shall take the first steps in showing how those principles may lead us to each of the premises of the Derivation, beginning with my defence of "ought implies can."

Chapter II - Obligation and Ability

Thus far, we have discussed the history and current state of the debate regarding PAP, and articulated how I intend to argue in favour of the principle. In this chapter, I aim to mount a defence of the first of the premises which comprise my Derivation of PAP - the principle of "ought implies can", often first attributed to Immanuel Kant[49] - in the particular sense which I believe PAP to be deducible from it. Here, I shall cover my own preferred interpretation of OIC and explore what I consider to be the most powerful motivations in favour of it. Once this is complete, the following chapter will then consider the case *against* this interpretation of OIC, including both arguments against OIC in any form as well as those targeted at my understanding of this principle more specifically, and will detail my responses to the most compelling and/or noteworthy objections to the arguments I will present in this chapter.

My goal is to offer an account of "ought implies can" (OIC) which draws on intuitions about both the function and purpose of our genuine moral obligations as its starting point. To begin, Sections I - III deal with each of the different aspects of the OIC principle, first establishing exactly what I mean by "ought", then "implies" and "can" in turn. In Sections IV - VI I discuss three reasons why we should accept OIC, drawing on David Copp's prior work on the subject, and explain how my chosen interpretation of the principle is grounded in and supported by each of these motivations. Taken together, I believe these reasons - reasons concerning fairness, the action-guiding nature of moral requirements, and the nature of morality in the counterfactual world where OIC is false - do much to show why OIC is so commonly adopted, the difficulties and consequences of attempting to do away with it, and why my own variant of the principle is so plausible. Once this is done, we shall gain a strong base for the future premises of the Derivation that are built upon OIC as I understand it.

[49]Immanuel Kant, *Religion within the Boundaries of Mere Reason*, ed. by and tr. by Allen Wood and George di Giovanni (Cambridge: Cambridge University Press, 1998), 6:47, 6:50, 6:62 et al.

Before I move onto the specifics of my definition of OIC, however, it is helpful to make a certain distinction between two basic questions which this chapter will be concerned with. The first of these is a straightforward matter of truth - does OIC hold true on any interpretation of the relevant terms? The consensus in regard to this question has been that it is, which is partially why I pursue the strategy of tying the much more controversial PAP to it. The second, considerably less settled question is one of *interpretation* - that is, in what sense and under what conditions are OIC true? Answers to this later dispute have varied significantly across the philosophical spectrum, from those like myself who view OIC as a hard and fast logical entailment to be applied to all genuine moral duties, to those who regard the principle merely as a matter of social implicature.[50] In essence, then, whilst the essence or "spirit" of OIC is widely - though not universally - agreed upon, its particulars are a veritable minefield of disagreement.

Part of the challenge of these two chapters which will focus on OIC, therefore, is to present a version of the principle which answers these two questions without being drawn into a battle with other advocates of OIC that will, almost inevitably, interpret the principle differently to myself (and perhaps significantly so). It is not my intention, nor would it be wise for my project to be dragged into a murky search for a "complete" definition of OIC - that is to say, one which encompasses all the true interpretations of the principle and excludes the false ones. All I am seeking to motivate and defend in this chapter is a single sense in which OIC is correct - albeit, in my view, a highly plausible and intuitive one - because all I am seeking to conclude is that PAP in regard to blameworthiness is derivable from a certain interpretation of the former principle. This does not necessarily mean my project will have no conflict with my fellow supporters of OIC - far from it! - merely that I intend not to let the discussion ahead become dominated by disputes over the endless number of subtly different variations of the principle.

Section II.I: Defining "Ought"

When discussing how best to define "ought" for the purpose of "ought implies can", there are two distinct subjects which one might be asking about. The first is to ask what types of "ought", specifically, we are saying are constrained by the relevant "can." Are we, for example, only referring to our all-things-considered or "genuine" moral obligations? Or

[50] Sinnott-Armstrong, William. 1984. ''Ought' Conversationally Implicates 'Can'', *The Philosophical Review*, 93, 2: 249-261.

is it also intended to apply to less demanding "oughts", like Kant's hypothetical imperatives - "if you want X, then you ought to do Y"? What about other types of genuine obligations, such as epistemic, legal, or societal "oughts"? If only some of the above, why only those and not the rest?

The second, perhaps more compelling reading of this question is to ask what makes a particular "ought" obligatory to begin with, and how these obligations exert their obligatory nature (what I will call their *authority)* over their subjects. In other words, we may be asking what it *means* when we say a moral agent "ought" to do something, as opposed to it merely being praiseworthy or "good if" we were to do it. This second question matters to the discussion of OIC because if the source of a moral obligation's authority can be tied in some way to the abilities of the obligated agent, as I believe it can, an obligation which commands the impossible would therefore lose its authority over the agent in question and cease to be genuine. If this is so, we will have strong grounds to favour at least some conception of OIC. In the process of explaining my understanding of the relevant "oughts" to OIC as I wish to defend it, I shall give answers to both of these questions. In fact, I believe the two enquiries are connected: part of the explanation I shall offer for why my version of OIC deals with a specific type of "ought" is that those oughts *are* "oughts" (rather than desirable, but ultimately optional actions) in virtue of certain capabilities of the agents under them.

In the version of OIC I am supporting, and from which I intend to derive PAP in regard to blameworthiness, the answer to the first question is simple: for the purposes of this thesis, the "oughts" in OIC which imply "can" are our genuine, all-things-considered moral obligations. I speak of our "genuine" moral obligations here to distinguish them from the commonly used class of *prima-facie* or *pro-tanto* obligations, which initially appear genuine but may be overridden or extinguished in cases of moral conflict. The sense of OIC I am interested in is one which applies only to our objective moral obligations, once all the normatively relevant facts have been considered, and not for all senses in which we may use the term "ought."[51]

Why then do these obligations require "can", whereas other "oughts" may or may not? To answer this, we must turn to the second of the two questions mentioned above and ask

[51] To be clear, I remain entirely neutral on the question of whether these other "oughts" may imply some sense of "can"- my business here is only with the kind of "ought" I have specified.

what makes obligations such as these morally obligatory. For it is undoubtedly the case that many other types of "ought" are supererogatory, even if social convention or our own desires might make them feel like they are required of us. When for instance we are told "we should leave now if you want to attend the opera tonight" it is perfectly acceptable, morally speaking, to reply that I won't be going because I want to stay home and binge on pizza and ice cream.[52]

This provides a clue to the first inherent quality of obligatory acts. When an action is not morally required of us it is, by definition, optional. Thus, when I examine the range of morally acceptable options which are open to me it does not make a difference whether I go to the opera or stay home to eat pizza - either are fine. But it does make a difference to the number of morally acceptable actions I have if I am *obligated* to do one of those things, because my obligation effectively "rules out" those options which do not involve doing that particular thing, from a moral point of view."[53] An action which is "morally acceptable", in this context, is one which is either morally neutral or morally good, beneficial, or praiseworthy. Morally acceptable actions are those where, as the name implies, the actor has done nothing wrong, and it would be evidently unjust for them to receive any kind of blame or other moral sanction as a result of them. By the same logic, a morally unacceptable action is one that is morally lacking in some way, and whose agents are therefore (I will argue) deserving of some form of blame for having performed them in the circumstances they were.[54]

As such, it seems that a genuine moral obligation inherently *constrains* the number of morally acceptable actions available to us at any given time, ordinarily limiting us to those actions in which we fulfil the obligation in question. That this is a necessary condition of an action's being morally obligatory can be easily demonstrated: if an obligation of mine did not render any course of action as morally "off-limits", as it were, then this would logically imply that every possible action in the current circumstances would be morally permissible for me. Yet this looks conceptually mistaken, because to be under a moral obligation entails by its very nature that I cannot simply do what I like - there is something

[52]Compare the obligatory variant "we should leave now because you promised to take me to the opera tonight", where my preference for junk food makes no difference to what I ought to do.
[53] And vice-versa: with obligations to avoid acting in a certain way, our only morally acceptable options will be those in which I do not perform that action.
[54] Of course, an agent may still mitigate or "excuse" their blameworthiness for such an action, usually by appeal to something relevant and uncontrollable in the surrounding circumstances. On my account, a moral excuse is simply something which justifies an otherwise morally unacceptable course of action.

which *must* be done. In a case such as this where none of my actions are morally constrained, anything I "ought" to do would seem to be one of the other kinds of ought mentioned previously, rather than the type of genuine moral obligation which I am concerned with in this thesis. On the model I am proposing, then, our moral obligations set the bar between the acceptable and the unacceptable: for the most part, those actions which fulfil or do not violate our obligations are morally permissible, whilst those which fall short or conflict with them are not (and will, in most cases, result in blameworthiness for the relevant agents).

Based on the above observation, it may be possible to deduce a further innate property of our genuine moral duties. David Copp has argued that "the point of moral requirements is to affect our decisions, and to lead us to do what is right, by being taken into account in our deliberation."[55] In Copp's view, our moral duties are not only a constraint on our available actions, but a guide to help us determine which of our various available options we should take. I support this perspective because if I am correct that moral duties inherently constrain our morally permissible actions, it is an easy step from there to the claim that they are also inherently action-guiding. For the effect of highlighting a subset of my possible options as morally acceptable (and the rest as unacceptable) is to give me a *reason* to prefer those options which fulfil my obligations when I am in the midst of my moral decision-making. This reason consists, simply, in that they are the morally acceptable (indeed, required) actions, and by taking them I will continue to be a morally upstanding person, avoid committing wrongdoing, and so on and so forth. Thus, we can now see that a genuine moral obligation is also a guide to the actions we ought to take in a given circumstance, as well as a restriction on the number of our morally permissible actions.

From this initial examination of the nature of our genuine moral obligations, it should now be clear why I intend to apply "ought implies can" only to those obligations, and not to any other form of moral "ought". OIC as I envision it is a principle intended to balance the inherent restrictions and demands of our genuine moral duties. Its essential purpose is to ensure the "system" of our moral obligations, for want of a better word, make only fair and reasonable demands in the actions it mandates agents to undertake or avoid performing, given the penalty of blameworthiness which I believe occurs for non-

[55] Copp (2006) p.273.

compliance. Without such a safety measure, either in the form of OIC or at least a highly similar principle, we risk the possibility of unfair, overly demanding, or simply ridiculous moral obligations whose existence, I feel, we should wish to deny outright. By contrast, supererogatory or otherwise non-obligatory "oughts" do not command anything of us or restrict the number of morally acceptable actions in this way - they are at most a suggestion about how best to fulfil our desires or a kind of moral rule of thumb. Therefore, there is simply no need or a principle like OIC when dealing with "oughts" of this kind, because are no corresponding issues with "ought" statements which ask the unreasonable or even impossible in the way there are with our genuine moral duties.

Thus, we now have a picture of the "ought" part of OIC as I seek to defend it. OIC is concerned with our genuine, not prima-facie moral obligations, the distinguishing features of which being that they 1) restrict the number of morally acceptable courses of action available to the obligated party to those which adhere to the obligation in question, and 2) serve as a guide to how we morally ought to act by providing reasons to perform those actions which are morally desirable. With this established, we shall now turn our attention to the next part of "ought implies can", and discuss the nature of the connection between these obligations and the abilities of the agents who hold them.

Section II.II: Linking "Ought" and "Can"

In debates concerning the particulars of "ought implies can", the majority of the arguments tend to revolve around the precise meanings of "ought" and "can." Comparatively little attention has been paid, historically speaking, to the specific type of link that connects the two. This is despite the fact that there are, quite clearly, several viable and mutually exclusive options to choose from in determining what this connection may be. On the one hand it might be believed that "ought", as the phrase suggests, merely implies or suggests the relevant "can" and whilst the two may often or would normally go together this is not necessarily the case. On the other is the view that "ought" does not *imply* "can", properly speaking, but rather *entails* it. With this second line of thought, OIC depicts a necessary and unbreakable connection between "can" and "ought" where every genuine moral obligation we possess, without exception, is within our capabilities at the time it is to be exercised.

The interpretation of OIC which I will defend in this thesis falls into the latter category. I believe that "ought" in the sense of our genuine moral obligations does entail "can", in every case, and so "cannot" entails "not ought." This is to say that for all actions, if I cannot carry out that action then it cannot be truly morally obligatory - though it may, of course, *appear* obligatory from my perspective if my inability is not obvious. Importantly, however, I do not think that this "can" must always simultaneously co-exist with the obligation(s) in question, for some moral obligations relate to actions to be done (and in some cases, which may *only* be done) at a future point in time. It only matters that the obligated agent has the power to perform the action at the point it is due to be executed. Put another way, I believe it is possible in some cases to have an obligation at T to perform some future action at T*, even though I cannot perform said action at T.

This is relevant to consider because it is common to understand the principle of OIC (a la Zimmerman)[56] as having a "double time-index," where if S "ought" [at T] to do A [at T*], then S "can" [at T] do A [at T*]. For instance, having arranged to meet my friend for dinner next week I have now an obligation to meet them at that future time, assuming I can (both now and at that time). Most versions of OIC adopt such an index because it ensures the link between "ought" and "can" is appropriately time-sensitive: statements about what I ought to do and what I can do are always valid at the same time.

I, however, choose to reject this approach. I have concerns about the implications of claiming S can (now) do A (at some future point). I understand the intuition which leads us to this view, but I feel it runs up against a conflicting and equally plausible intuition: that an agent's capacities can be exercised at the time at which they are possessed. This thought, unlike the double time-index variant of OIC, views what one is able to do at any given moment as the alternatives which are available to that agent at that specific moment in time, even if this choice is simply between one action and not acting at all. Not coincidentally, this thought is also central to my preferred interpretation of what the morally relevant "can" in OIC is, which I will explain in the following section. I am drawn to this second intuition due to what Makin once termed the "Necessity of the Present" (NP): "When a specific time comes around, it is by that time too late for A to do anything at that time other than what A is in fact doing at that time."[57] At any given moment in

[56] Zimmerman, M. J. 1996. *The Concept of Moral Obligation*, ed. by Ernest Sosa (Cambridge: Cambridge University Press).
[57] Makin, Stephen. 1996. 'Megarian Possibilities', *Philosophical Studies*, 83, 3: 253-276 (p. 255).

time, it is already past the point where I had the ability to change what I would be doing at this present moment, regardless of what (if any) actions that might be.

NP, of course, creates a problem for the double time-index variant of OIC: if at any given time I am only able to do what I am in fact doing, then it is not the case that "S can [at T] do A [at T*]" since I am evidently not doing "A [at T*]" at T. But this, in turn, means that S *cannot* be obligated at T to do A [at T*], given that they are presently unable to do (A [at T*]), and by extension it becomes impossible for S to be obligated, at T, to do anything at any future point (T* or otherwise). Since most "ought" claims in general, whether they refer to genuine obligations or otherwise, refer to actions to be carried out in the future rather than something to be done at literally the present moment, the double time-index view of OIC would leave us in the enviable but sadly mistaken position of having substantially fewer genuine moral duties than we truly do.

My answer to this problem, as already stated, is to forgo the double time-index categorisation of OIC in favour of a single time-index:

STI: If S "ought" [at T] to do A [at T*], then S "can" [at T*] do A [at T*].

This approach has its own weaknesses, not least of which is the epistemic question of how we are to know exactly what we will be capable of in the future (and thus, which obligations we are to rule out as impossible in the present). However, I find problems of this sort to be more palatable than those faced by the double time-index categorisation. One reason for this decision is that STI better captures the essence of the claim that I currently have an obligation to do something at a future point, as shown by the following analogy:

Suppose I have agreed to present a paper next week, and am currently in the process of writing it. On the double time-index version of OIC, there is now an obvious problem: having promised to present the paper, I am presumably obligated to do so next week. Clearly, though, I cannot "present the paper next week" now, even leaving aside NP, as even if the paper were finished today would not be the right time to do so. This also counteracts the reply that I am obligated to do is present the paper, not "present the paper next week", for again: even if the paper were finished, I could not fulfil my obligation by presenting it now. It is also undeniable that I am indeed under the obligation to "present

the paper next week" at the current moment in time, when it cannot yet be fulfilled, and not merely next week. If I were to abandon working on my paper to go drinking between now and when it is due to be given, that would clearly be morally unacceptable, and the reason I am required to work on my paper is that I accepted an obligation (to present it next week) when I agreed to do so. These considerations indicate that the ability that matters, morally speaking, is not whether I can *now* "give the paper next week", but whether I will be able to give it *next week* - that is, at the time when the obligation must be executed. Hence, the move from a double to single time-index.

This presentation of events also aligns with how would naturally use "ought" in this type of context. If I were wavering between T and T* about whether to give the paper, a friend might remind me of my promise and tell me I ought to do so. This, too, indicates that I possess the obligation to give the paper before the point it can be exercised (i.e. T*). But this reminder would clearly come with an unspoken caveat: that I should give the paper *if*, at the proper time next week, I have not been petrified by stage fright, or run over by a bus, or abducted by kidnappers…and so on and so forth. In short, that I should give the paper next week assuming that I can when the time comes. Were I to protest that I am unable to give the paper (and therefore have no such duty) because next week has not arrived yet, my friend would presumably reply that this is not the point: it makes no difference to my obligation that I cannot complete it *yet*, only whether I can complete it when the time is right.[58]

Having now established what the "ought" of my preferred sense of "ought implies can" refers to, and in what way I believe it is connected to "can", I shall now move on to the final - and by far the most involved - part of the definition of OIC as I understand it - what is exactly meant by that "can", as I intend for it to apply to my chosen interpretations of OIC and PAP alike.

Section II.III: Defining "Can"

When seeking a definition of "can" or "able" for the purposes of OIC, what we are truly asking is which of the obligated agent's various abilities count as the morally relevant

[58] Equally, if it should transpire between T and T* that I will be unable to give the promised paper at T*- perhaps because some jealous rival has crippled me- my obligation to give the paper at T* will vanish even though T* has not arrived yet.

kinds of action when judging whether the obligatory actions are possible for them to perform. It seems helpful, therefore, to begin our search by setting out the broad parameters we are looking for from such a definition.

Firstly, it seems important that whatever eventual sense of "can" we accept as the one that matters to OIC does not yield the conclusion that no agent is capable of anything. This may seem like a somewhat trivial concern, but it is important for the purpose I intend OIC to serve in my wider Derivation. Secondly, I think a functional account of ability in the sense we are interested in must accept the possibility that there are things which we "can" or are "able" to do which we are not presently doing, and perhaps will never actually do. To be clear, this is not to say that such an account must be inconsistent with determinism, as it is possible to define these terms in such a way that this capacity can be maintained in a deterministic world. It is merely a rejection of the Megarian-style position in which the only thing that we are capable of doing at any given moment is the thing which is *in fact* being done by us at that moment (for instance, that we are only capable of building a house during the process of housebuilding, and of nothing else).

Additionally, I think when discussing "able" and "can" in this context we are asking questions about an agent's *powers*, rather than their *dispositions.* One of the natural ways in which we use "can" in our ordinary language is in reference to the behaviour of objects, or groups of objects, under certain stimuli. So we say that wine glasses "can" shatter when hit with the right musical note, or this guitar "can" become untuned if not properly attended to. When we are speaking in this way, we are referring to an object's (or potentially agent's)[59] dispositions - the reactions they are predisposed to experience under a given set of conditions. This is not, however, the sense of "can" I think is the morally relevant one to OIC. When we ask if a moral agent "can" fulfil their genuine obligations, I believe we have in mind the other broad category of "can" and "able" statements, which are sometimes described as statements about an agent's powers. These include, but are not limited to, expressions of the form "Sarah can speak German," "I can play my flute once I've had a nap," and/or "James can drive once he's passed his test."

[59]Whilst this description is not normally associated with moral agents, I think it is appropriate in certain circumstances to describe them as having dispositions as well. A particularly sensitive agent might be disposed to cry if confronted with emotional difficulties, for example, or an aggressive one to become violent after having drunk a lot of alcohol.

This group of claims encompasses the myriad ways in which we might ordinarily refer to an agent's capacities and possible actions, either at a specific time or in general terms. I believe therefore that when we are attempting to understand what an agent is able to do for the purposes of applying OIC, we are essentially asking for a list of that agent's powers at the relevant time(s). Of course, certain statements about an agent's powers do not specify the time at which they are held, and rightly so: "I can speak German" is usually not translated as "I can speak German now, and five seconds from now, and five minutes from now, and…". Thus, we require a second distinction within the realm of such statements, between an agent's *general* abilities and their *specific* abilities.[60]

These two types of ability denote, metaphorically speaking, two "modes" of a particular power (or powers) an agent might possess. To continue a prior example, imagine I am a fluent German speaker who has contracted a bad case of laryngitis. If asked "can I speak German?" there is a genuine sense in which the answer is yes, even at the present moment - I have undergone suitable study, I have a great deal of experience at doing so, and so on and so forth. Yet there is another sense in which I clearly cannot speak German at present, because I currently cannot speak *any* language due to my bad throat. These two senses, both of which are meaningful and accurate, can be said to refer to my general and specific abilities at that moment in time, respectively. In this case, I retain the general ability to speak German but not the specific ability since all my relevant training remains in place, but I am currently being impeded from exercising it due to my bad throat. If I were to be administered a miracle cure for my throat condition, my specific ability to speak German would return (all other things being equal) and if my memory were to be completely wiped by an angry mind-reader, my general ability to speak German would disappear along with my specific ability. One's general abilities can thus be broadly understood as describing what one is capable of under ordinary conditions, or as part of the normal course of events (hence why they are often stated without any temporal qualifiers), whereas one's specific abilities refer to what you are currently "in a position" to do.

For the purposes of determining whether we can meet our obligations at any given time, therefore, it would make sense that an agent's *specific* abilities would be the morally relevant ones to OIC. Suppose, for example, we wish to know if I am obligated to rescue my drowning sibling. When we ask if I "can" or am "able" to rescue them in this case,

[60] See Honoré, A. M. 1964. 'Can and Can't', *Mind*, 73, 292, 463-479 and Mele, Alfred R. 2003. 'Agents' Abilities', *Noûs*, 37, 3: 447–470 for more on this distinction.

we are not enquiring about my usual ability to swim under ordinary circumstances. Rather, what we wish to know is whether I have the ability to swim *now* - is it within my power, at this moment, to act as my moral obligation requires? In most cases of this sort, lacking the specific ability to do what is required will usually be sufficient to defeat one's purported obligation according to OIC, whereas lacking the general ability to act in that way but *not* the specific ability is not. In the case of the drowning sibling, if I normally cannot swim but a helpful stranger tosses me a life preserver it would seem grossly inappropriate for me to claim I have no duty to rescue my sibling because I lack the general ability to swim. By contrast, if I am a perfect swimmer but the same villain who is drowning my sibling has also broken my legs, it is much more plausible that my obligation at that moment would lapse accordingly.

We now have a framework for our definition of "can" or "able" as it will apply to both OIC and PAP. What matters in calculating an agent's abilities when we invoke these principles is, following the above definitions, an agent's specific ability rather than their general ability and their powers as opposed to their dispositions. This answer, however, is not detailed enough. For what, precisely, constitutes a specific ability to do the thing which I am genuinely morally obligated to do (for OIC), or to do otherwise than a particular action (for PAP)? In order to properly understand the scope of the morally relevant "can", we require a more principled means of identifying what one's specific ability in regard to a given action consists in.

In answering this question, I follow (amongst others) Vihvelin,[61] van Inwagen,[62] Vranas,[63] and Lewis[64] in the belief that a great many contributors to the so-called "free will debate" actually agree substantively about the meaning of the morally relevant "can", as it pertains to our moral obligations, blameworthiness and responsibility for our actions, and so on. The famous disagreement between compatibilists and incompatibilists we covered in Chapter I is understood, on this accounting, to be a conflict about not the meaning of the ability necessary for an agent to do otherwise, but whether that agreed-

[61] Vihvelin, Kadri. 2004.'Free will demystified: a dispositional account', *Philosophical Topics*, 32, 1: 427-450.
[62] Van Inwagen, Peter. 'How to Think about the Problem of Free Will', *The Journal of Ethics*, 12 3-4: 327-341.
[63] Vranas, Peter B.M. 2007. 'I ought, therefore I can', *Philosophical Studies*, 136, 2: 167-216.
[64] Lewis, David. 1983. 'Scorekeeping in a Language Game', in *Philosophical Papers* Volume I (Oxford University Press), pp. 233-250.
—1987. 'The Paradoxes of Time Travel', in *Philosophical Papers* Volume II (New York: Oxford University Press). pp. 67-81.

upon sense of ability is inconsistent with the theory of determinism - a disagreement which I am more than content to remain neutral on in this work.

Vihvelin describes this supposedly agreed-upon definition thus: to say "X can do Y" or "X is able to Y", at a specific point in time, is to say that X has the *ability* to do Y and the *opportunity* to do Y. X's "ability" in this context refers to the knowledge, skills and competencies required to do Y, as well as whatever physical and psychological capacities are necessary to apply them - for instance, to have the ability to play Beethoven's Moonlight I will need to know how to read sheet music, have been taught to play a piano, have sufficiently steady hands, and so on. Vihvelin's precise name for these qualities is an agent's "narrow ability", and to avoid ambiguity I will refer to these relevant skills as the *Prerequisites* for Y. Thus, X "can" or is "able" to do Y only if and only if X possesses the Prerequisites for Y and has the Opportunity to do Y.

To have the opportunity to perform an action means, in simple terms, just what it does in everyday language: not only that I have all the relevant skills to perform an action, but also the chance to apply them.[65] Examples of having opportunity in this sense include being seated at (or sufficiently close to) a piano to play Moonlight, having received an offer from the University of Sheffield to study philosophy, or being near to a suitably body of water to go swimming. Exactly what the Prerequisites are for any individual action, and what will constitute having (or lacking) an opportunity to perform that action will therefore vary according to the action in question. As Vihvelin puts it, if S "can" do X "S has what it takes to X…and, moreover, she's got the means and opportunity and nothing external stands in her way."[66] The combination of these, what Vihvelin refers to as "wide" ability, is the sense of ability I believe is morally relevant to, and required for, OIC and PAP as I defend them here.

[65]An important clarification: as discussed in Section II, the opportunity in question does not necessarily need to be possessed at the same time as the obligation is held, due to my single time-index interpretation of OIC. In order for an agent to be counted as able to fulfil their obligation for the purposes of OIC, they must possess this opportunity at the point in time at which the obligation is due to be exercised. In terms of the obligation to present a paper next week, I will be able to fulfil this obligation as long as there is a chance for me to give the paper next week at the appropriate time (and if there is not, perhaps because I have been run over by a bus or abducted by kidnappers in the meantime, I will not be "able" to give the paper and my obligation will lapse).

[66] Vihvelin, Kadri. 2013. *Causes, laws and free will: Why determinism doesn't matter* (New York: Oxford University Press).

This account of "can" also maps neatly into the distinction between general and specific ability with which we began. To possess the Prerequisites for an action but lack the opportunity to perform it is something stronger than the mere general ability to do Y - two classical pianists may share the general ability to play Moonlight, but if one of them is colossally drunk whilst the other is sober, the former will not satisfy the Prerequisites for playing Moonlight at that point in time whilst the latter does. The Prerequisites alone, however, are in turn weaker than the specific ability - in Vihvelin's terms, wide ability - and since we have already established that what matters in determining whether we can fulfil our moral obligations is what we have the specific ability to do, merely satisfying the Prerequisites of an action is not enough to sustain a genuine moral obligation to perform it. For this, we require the presence of an opportunity as well.

Whilst this idea of a universally-agreed upon "can" is appealing (although disputed, as we will see in the next chapter), we require a stronger defence than this. As Haji[67] has observed, even if one grants that philosophers generally are referring to the same concept when they discuss questions of "can" and "able", and further that that concept being referred to is the one suggested above - both certainly nontrivial assumptions - this does not by itself provide a satisfactory definition of the relevant terms. The definition offered by Vihvelin identifies two necessary conditions for being "able" in this morally relevant sense, which in my terms are the Prerequisites and opportunity. The former condition seems broadly unproblematic, since there is an obvious sense in which if one acts, one was therefore by definition able to satisfy the requirements which were necessary for that action (or else it could not have been done). If I spent the evening watching TV I must therefore have owned a television and been able to turn it on, if I swam ten lengths this afternoon I must have been near a swimming pool, capable of getting in it and swimming one length, and so on.

Yet the latter condition, Haji notes, is much more ambiguous. Even if we grant that opportunity as described is a necessary condition for the morally relevant ability to perform an action, "could we not expect a debate among interested parties…on the meaning of "opportunity" or over the analysis of the concept of having the opportunity to do something?"[68] Without a rigorous analysis of the conditions that constitute an opportunity, the argument goes, our questions about the nature of the morally relevant

[67] Haji, Ishtiyaque. 2017. 'The Obligation Dilemma', *The Journal of Ethics*, 21, 1: 37-61.
[68] Haji (2017) p. 51.

"can" only seem to have been partially answered. After all, it is clearly of little value to have a widely agreed-upon of definition of morally relevant ability only to have no such agreement on what its individual commitments amount to. I think Haji is right here to say that there are multiple viable interpretations of what one's opportunities are at any given moment - to take a basic example, even amongst those who accept the sense of "can" mentioned above as the relevant one to OIC and PAP there is considerable disagreement about whether determinism interfere with one's ability to do otherwise. I will therefore simply aim to outline (and in time, motivate) an understanding of "opportunity" here which is in keeping with the principles governing morally relevant ability that I have adopted throughout this chapter.

My favoured understanding of "opportunity" is the one originally put forward by Franklin,[69] which uses a possible-worlds analysis to articulate the various options open to an agent at any given time. For Franklin, to have the opportunity to do something in a given set of circumstances (W) is to have a possible world (W*) which is accessible to me at that time in which I *do* perform the action I currently have the opportunity to perform. Formalised, Franklin's definition of opportunity looks like this:

(O*) S has the opportunity to X (at T) in W if and only if there is a possible world W* in which S X's at T, and everything except S's X-ing and the causal consequences of her X-ing is the same as in W.[70]

I find this definition appealing because it turns on the natural idea that an agent has the chance to do something only if there is nothing which prevents me from doing it. To take a basic example, a lifeguard lacks the opportunity to save a drowning swimmer if they are currently pinned down by a bodybuilder. The need to keep everything in W and W* apart from my action and its consequences identical is, Franklin argues, necessary to sustain this intuitive bit of thinking - "if, in addition to her X-ing, something in her environment must be different, a difference that would not be identical to or a result of her X-ing, then it seems that this required difference is an obstacle: it prevents her from doing otherwise."[71] The line here is that anything else which would need to be different in order for me to access the world in which I perform the action in question is, by

[69] Franklin, Christopher E. 2011. 'The Problem of Enhanced Control', *Australasian Journal of Philosophy*, 89, 4: 687–706.
[70] Franklin (2011) p. 697.
[71] Franklin (2011) p. 698.

definition, an obstacle which currently prevents my performing it. Since we can only have the opportunity to do things which we are not currently being prevented from doing, we therefore do not have the opportunity to perform those actions until such a time when our circumstances have aligned correctly - in other words, when the obstacle to our performing the action is removed.

Though I think Franklin's logic is mostly sound, I do have a concern which will require a slight edit to his definition. I think we have slightly more room for manoeuvre in categorising opportunity than Franklin does here. Specifically, I am not convinced that anything in your circumstances which must be different in order for you to perform a particular action is automatically an obstacle to your performance - at least not in the sense we intuitively have in mind when we talk about *being prevented* from doing something. I am imagining here a type of case in which there is an obstacle of the form Franklin has in mind in the way of my desired action - that is, something else which must be different in W* other than my action and its consequences for me to successfully perform it - which nevertheless clearly does not *prevent* me from performing that action at this moment in time.

To borrow an analogy from Haji,[72] imagine I am playing basketball and wish to score some points. However, an evil demon has sealed the hoops with an invisible forcefield which will deflect even the best-placed shots. Here, Franklin would presumably agree we lack the opportunity to score, as something else must change between W and W* aside from my action first (the forcefield has to be removed). I concur. Now, Haji continues, suppose a benevolent angel intervenes on my behalf and will dispel the forcefield, but only when I attempt to score. Do I have the opportunity to score at this moment, when I attempt the shot? It seems like Franklin's analysis of opportunity, as written, has to say I do not. At the moment I am shooting, there is still something about the world aside from my action that needs to change before W* is accessible to me: the forcefield must still disappear. Nor is the disappearance of the barrier part of the casual consequences of my scoring, since the angel intervenes *before* I score. It is a consequence of my *attempt* to score, which is not the same thing - the barrier would disappear regardless of whether I score with this shot, so long as I try to.

[72] Haji (2017) p. 59.

This is a useful counterexample because I think we do possess the opportunity to score when we take the shot with the angel's blessing, in the intuitive sense of the term I am attempting to maintain. In terms of possible worlds, I believe the W* where I score is accessible to the W I currently inhabit when I shoot with the angel's blessing. In this case, the angel's blessing means that the forcefield is not an obstacle which prevents my scoring, even though it is an external fact that needs to change before I can. There is, I think, a natural instinct that an "obstacle" which vanishes the moment we attempt to test it is no true obstacle at all, as it does not prevent us from the performance of any action(s). Consider again the earlier case of the lifeguard being held down by a bodybuilder: if we know the bodybuilder were to let the lifeguard go the moment they start resisting, it would be illogical for them to conclude that they lack the opportunity to save the drowning person (and therefore have no obligation to do so).

Thankfully, I think a minor tweak of O* will allow us to cover these sorts of cases whilst still maintaining an intuitive sense of opportunity:

Attempt-Opportunity (AO): S has the opportunity to X (at T) in W if and only if there is a possible world W* in which 1) S attempts to X (at T), 2) S's attempt to X succeeds, and 3) everything except S's attempt to X, S's success in X-ing, and the causal consequences of her attempt (and of its success) are the same as in W.

Whilst AO has more moving parts than O*, the only significant difference between the two is that we have expanded the limitations on what is allowed to differ between W and W* to include the process of our attempts to perform those actions. Now AO, unlike O*, will give the correct answer to the question of whether we possess the relevant opportunities in cases of this sort. As we have noted, the angel's intervention is not part of the act of scoring or of the causal consequences of my scoring, but it *is* part of the consequences of my attempt to score - if I had not made the attempt, the angel would not have intervened to dispel the forcefield. As such, it is counted as one of the things which is permitted to differ between W and W*, which was not the case under O*, and hence the angel's presence increases the number of opportunities available to us rather than being a further obstacle to overcome.

This completes my definition of ability in the sense which I believe is the morally relevant one to both OIC and PAP. To summarise, an agent is able to perform an action in the

sense that is necessary to hold a genuine moral obligation to do so if and only if they have the specific ability to act in that way at the current time. To possess the specific ability to do something is to satisfy its Prerequisites - to possess the knowledge, skills and capabilities required for its performance - and to have the opportunity to exercise them at the time the obligation is to be discharged. Opportunity, in turn, is a matter of what actions are open to me in my current circumstances - what I may attempt to do, and succeed at doing, given the world "as it is." At this point, I now wish to turn my attention from setting out the particulars of my analysis of the three components of OIC ("ought", "implies", and "can"), to the task of motivating that definition and offering reasons why we ought to accept it.

Section II.IV: Motivations of Fairness

With a complete understanding of my preferred form of OIC now in place, it is time to set out the attractions of the model I have proposed thus far. Thus far I believe my interpretation of OIC to be intuitively plausible on its own merits, but we have not yet seen substantive reasons why we ought to prefer it to any of the numerous other variants of OIC we might choose - or, indeed, reasons to prefer some version of OIC to none at all. Since this interpretation of OIC serves as the essential starting point for my Derivation of PAP, a robust grounding of the former principle is now required.

Broadly speaking, I believe there are three major categories of reason in favour of OIC: reasons concerning fairness, reasons concerning the function of our genuine moral obligations, and reasons which relate to the world in which OIC and all highly similar principles are false. Furthermore, I believe each of these categories provides a particular incentive to favour my own specific version of the OIC principle, or at least one which is functionally very similar - though I should stress once again that I have no desire to force a competition among different interpretations of OIC. In this and the following two sections, I wish to explore each of these differing types of reason and show how they put meat on the bones of my case in favour of OIC, whilst demonstrating how they draw us toward my own interpretation of the principle in particular.

Perhaps the most basic motivation for OIC and the principles which most commonly accompany it, PAP included, concerns the notion of *fairness* as it relates to our moral dealings. Even outside of specifically moral discussions, it is a popularly-held thought

that it is unfair - not to mention irrational - to expect someone to do something which they cannot do, and to blame someone for failing to do something which is impossible for them. The kind of thinking at work here is, I think, something like this: fairness in this sense consists in being held responsible only for that which you *were* responsible, being blamed only for that which you are blameworthy for, and so on. It is one thing to be expected to do something which (although perhaps difficult) is within my capabilities in the relevant way(s), or to be blamed for failing to do it if the failure was my responsibility in some way. It is quite another, we think, to be told to do something which we know I cannot do, and then be blamed for the inevitable failure. This is because, assuming that I am not somehow responsible for my own inability to succeed, to blame me in such a case appears to be a misattribution of fault: intuitively, we feel that failure here was not *my* doing, because anyone in my position would have failed, just as I did, no matter how hard we may have tried. By the same logic, we also naturally think it is unfair for someone to *expect* me to do something which I cannot do, as we know from the beginning that my failure is inevitable in the same way.

This line of thought provides a strong reason to accept OIC. Although OIC on my reading is strictly speaking a principle about our genuine moral obligations, rather than blameworthiness per se, as I have argued previously I believe blameworthiness inherently arises as a result of performing actions which are morally unacceptable. Since, in turn, an action can only be morally unacceptable if it is ruled out by a genuine moral obligation,[73] the observation that we seemingly do not deserve blame for failing in an impossible moral obligation naturally gives us reason to question whether such impossible obligations are genuine in the first place. This, of course, is exactly the claim made by OIC.

To see how this kind of thinking might lead us to my specific sense of OIC, consider the question in this way: I have argued previously, following Copp, that genuine moral obligations "rule out" certain possible actions of ours as morally unacceptable. In particular, they seem to rule out all actions which do not fulfil or are otherwise inconsistent with the duty (or duties) in question. Copp argues that if it were possible to be under an unachievable, yet genuine moral obligation then it conceivable that a moral agent might end up in a situation where *all* of their available actions are morally ruled out in this way, because none of them are consist with or can satisfy the various moral

[73]These two ideas constitute Premises 2) and 3) of the Derivation, and we shall assess their merits in Chapter IV.

obligations they are under. Irresolvable or "tragic" moral dilemmas would be a reasonably common, if philosophically disputed example of such a phenomenon. But in such a situation, given that we have no morally acceptable courses of action, it appears we will be morally at fault no matter what we do since each of our available actions have been morally ruled out in this way. Even if we were to cite the impossibility of our obligation(s) as a (justified!) excuse for why we would not deserve blame in such instances, it is important to recognise that the obligated agents would still have done something morally wrong - they failed in a genuine moral obligation, and excuses are only applicable to behaviour which needs to be excused (that is, behaviour that would otherwise be morally wrong).

This inevitable moral "demerit", Copp claims, would be deeply unfair on the agent trapped in a situation of this kind. Such a verdict of wrongdoing is unfair for exactly the same reasons as the unfairness of a more everyday, non-moral demand, such as to drive me to work when I know you cannot drive or to memorise the entirety of *War and Peace* in a single night. Moreover, it is unfair to the agent even if, as is likely, no-one would ever *actually* blame them or consider them at fault, because it commits the same misattribution of fault described above. As with an impossible demand made by one agent of another, although the obligated party is (normally) not responsible for their failure to fulfil their moral obligations - given they were impossible - they are still deemed to have acted in a morally unacceptable way.

Hence, if genuine but unachievable moral duties were possible they would appear to give the wrong moral verdict in such cases - they mark us as having done something wrong, if excusable, when we simply have not. If, however, it is apparent that these agents would not have "fallen short" of what is required despite failing in their moral obligations, this indicates the bar for morally acceptable behaviour has been set unfairly high. Since it is nonsensical, in both Copp's eyes and my own, that a genuine moral obligation could ever be unfair to its agent in this way[74] (tantamount to saying that a genuine moral obligation could be itself immoral), we may safely conclude that unachievable moral obligations are *not* genuine and we possess no such duties.

[74] This claim- which I refer to as the "Fairness Principle"- will be the main subject of Chapter V, where I offer a full explanation and defence of the proposal that all genuine moral obligations are necessarily fair to their agents.

The above points serve as more than an argument in favour of OIC in the abstract. They may also, I believe, be used to support the particular variation of the principle I have detailed in the preceding three sections. If one rejects the conception of OIC I am defending in favour of another, there appears to be a resulting explanatory gap. Why, if not for the reasons I have presented, are the consequences of an impossible yet genuine obligation so intuitively unjust? Certainly, it is not difficult to motivate a form of OIC that is framed differently to my own, given the issues involved in rejecting the principle altogether (some of which we shall discuss in Section VI). The concern here is something else: how might one be motivated by these specific concerns about fairness, and yet be led to a significantly different version of OIC than my own? For example, there are numerous compatibilist accounts of "can" which have argued that the morally relevant kind of ability is not a choice between alternatives at any particular moment, but rather the chance to see one's desires and projects realised without obstruction (commonly phrased as "if you wanted to do X, you would have been able to do X").

Whilst these accounts may still object to certain aspects of impossible moral obligations, the specific motivation of *fairness* seems to drop out of the case in favour of such accounts due to the radically different conception of ability they are working with. Alternatively, if we were to take (for instance) Sinnott-Armstrong's view that the relationship between "ought" and "can" is a primarily linguistic one,[75] rather than the logical entailment I have championed, then we will no longer automatically be morally at fault in any sense for failing to clear an impossible obligation. Hence, the unfairness we are worried about in regard to such obligations is simply not present, since there will no longer be an unavoidable moral demerit when their agents fail to complete them. Indeed, it is one of the strengths of Sinnott-Armstrong's account that it manages to defuse these natural suspicions so elegantly, provided one is willing to accept the looser ties between "ought" and "can" which I cannot.

As such, if one is persuaded by the concerns relating to fairness which I have expressed here to accept some form of OIC, it seems one would naturally be led to a formulation of the principle which at least closely resembles the one I am defending here. Whilst there are, of course, many perfectly functional variations on OIC which do not categorise the

[75]See Sinnott-Armstrong, William. 1984. ''Ought' Conversationally Implicates 'Can'', *The Philosophical Review*, 93, 2: 249-261 and:
—1985. ''Ought to Have' and 'Could Have'', *Analysis*, 45, 1: 44-48.

nature of “ought” or the “can” of a moral agent as I have done, these variations seem to be primarily motivated by other objections to impossible moral obligations than the trap of an unavoidable moral demerit which Copp and I have described. Therefore, our thinking about fairness as it applies to our moral obligations, and our blameworthiness for failing to fulfil them, would seem to support not just an abstract version of OIC but one which views our genuine moral obligations as constraining and action-guiding in the way I have argued for.

Section II.V: Motivations of Purpose

Another means by which we can motivate OIC, both in general terms and my own form of it more specifically, arises when we consider the intended purpose of our genuine moral obligations. When I speak of “purpose” here, I do not mean to imply that our obligations have been designed or otherwise constructed in order to produce a certain outcome or arrangement. I am merely suggesting that our moral duties are not arbitrary; that there is a reason why certain, specific actions are required of us in certain conditions rather than others.

Copp believed this idea could be used to support OIC by arguing that our genuine moral obligations have a function - a role of sorts which they are required to carry out. We have already observed that our moral duties appear to provide a certain kind of guidance as a by-product of their innate ability to constrain the number of morally acceptable options open to us. Based on this characteristic, Copp argues that at least part of the purpose of a genuine moral requirement is to help bring it about that the obligatory action(s) are performed by the agents in question. More controversially, he goes on to say that it is *impossible* for an obligation which cannot be fulfilled to serve this purpose for if, under the circumstances, the obligatory action(s) cannot be brought about by its agent, then our obligation to perform that action can never assist in bringing it about that those actions *are* performed in those circumstances (since they never can be).

True obligations, it is suggested, are not like this. We normally consider the fact that we have moral obligations to perform certain actions and/or avoid others to always be relevant to our moral decision-making in the associated circumstances - whilst it does not necessarily settle the agent’s eventual action, depending on their moral fibre, the fact that I possess a moral duty should at least be worth something in the calculus. But if the above

point about unachievable obligations is correct, the only way facts about our moral obligations could be universally relevant in this way is for us to always be able to achieve them - in other words, if OIC is true.

Therefore, on Copp's account an unachievable moral obligation is incapable of serving this purpose, since a morally sound agent who believes they were under a relevant moral obligation will normally[76] take this fact into account when deciding how to act in a given circumstance and be led, as a result, to the morally right thing to do. But if an action were morally required of them that was impossible to perform, then that obligation could not play a role in my moral deliberations - at least, not in the direct way just described. It may well be that such an unachievable moral obligation could be useful to my moral deliberations by serving as a moral example or other indirect insight into what I should do, but - so Copp's argument runs - it could not give me a reason, morally speaking, to favour any one of the actions I *can* carry out over any other, unless some of them should happen to be significantly closer to the unachievable and obligatory action in a relevant way.

I feel, however, that this argument is mistaken. To take an example, suppose a building is burning and I am obligated to call the fire brigade - however, the arsonist is cunning and has cut the phone lines. I can, however, either try to rescue some of those trapped myself or run in search of a working telephone. Both options seem acceptable substitutes for my obligated but impossible action, and my obligation provides no guidance in helping me decide what I should do - which it obviously would if it were possible for me to fulfil it. As far as Copp is concerned, this is because the fact that I know I cannot fulfil my obligation means that I no longer have any reason to take my obligation into account when deciding what to do. Yet this explanation seems wrong. Even though I cannot complete it, my obligation here should *very much* factor into my thinking about what I should do. Indeed, it seems to constrain and guide my behaviour in much the same way it would if it were achievable, because despite being impossible it is still pointing me in the direction of (at least) two morally viable alternatives in rescuing those trapped or going to find help. It also continues to morally rule out certain other courses of action

[76] This excludes, for the sake of argument, truly amoral agents- those for whom moral considerations carry no weight whatsoever.

which are available to me, such as sitting down with a bag of popcorn and watching those inside burn to death.[77]

A simple counterfactual serves to establish why my impossible obligation would function thus. If Copp were correct about how impossible obligations behave in these situations, then all my available options would be morally "ruled out" on the grounds that in none of them achieve what is required of me. As such, all of my options would be morally unacceptable and all I can presumably do is choose the morally bad option that makes me most comfortable, since my impossible obligation is said to play no role in my moral deliberation. But this is not, I think, how this scenario would play out. Although anything I do here will be morally unacceptable - it is, after all, all defying what is still my genuine moral duty - some of my options are still morally worse than others. Once I have recognised that my initial obligation is impossible, it is irrational to suggest that it is somehow discarded when I am weighing up, as I surely would do, what I ought to do *now*. It remains valuable information to me, continuing to constrain my behaviour and pointing me in the direction of what ought to be done given that I cannot, strictly speaking, do what I ought.

Continuing the above example, if I did decide upon learning the phone lines were cut to simply sit and watch the fire, those around me would naturally be outraged and it would hardly make a difference if I were to point out I cannot fulfil my obligation to call the fire department. The only difference my obligation's being impossible seems to make is that if it were possible, my morally acceptable options would be restricted to those which involved doing the obligatory thing. Since it is not, the obligation's constraints and guidance change accordingly: all my potential actions now are unacceptable, but not necessarily equally so. Therefore, the (technically) morally unacceptable action which I ought to perform is the one that is closest to the obligatory thing which I cannot do. And of course, my unachievable obligation is still very helpful to me in working out what that will be, since it allows me to deduce which of my available actions is the closest one.

To briefly digress, one might wonder in this context if my actions are truly "morally unacceptable" in any sense if I am indeed doing the best thing that I can. I agree that in

[77] Nick Hughes also has a strong argument for the conclusion that unachievable moral obligations may still be deliberatively useful. See Hughes, Nick. 2018. 'Guidance, Obligations and Ability: A Close Look at the Action Guidance Argument for Ought-Implies-Can', *Utilitas*, 30, 1: 73-85.

such a scenario as this the closest thing to my supposedly "obligatory" action would in fact be perfectly morally acceptable, because *I* deny that the impossible action is obligatory to begin with. But if we ignore OIC and treat the impossible obligation as truly genuine, as we are assuming for the purposes of this example, to fall short of it would logically be the same as falling short of any other genuine moral duty - a morally wrong action, one which we happen to have a very good excuse for, but wrong nonetheless. This, as we discussed in the previous section, is what creates the unavoidable moral demerit which I believe makes an impossible moral obligation unfair to its agent.

Despite having diverged from this specific part of Copp's argument, I do think there is a way that the inherently action-guiding nature of our moral requirements may be used to argue against the possibility of genuine, yet unachievable moral obligations. We have already noted that our genuine obligations guide our actions through the mechanism of *reasons* - they provide moral guidance by offering incentives to perform certain actions and avoid others. As we have also seen, they seem to perform this function whether or not it is possible for the agent receiving the guidance to complete the obligation in question, which is where Copp's argument goes astray. To properly put down the possibility of genuine and impossible moral duties, we will need to identify a quality which is possessed by all genuine moral obligations which cannot be held by an impossible obligation (or vice-versa).

In my search for this quality, I wish to examine the specific type of reason that is created by a moral obligation. For I suspect our genuine moral duties are more prescriptive in their reasons-granting than has previously been suggested. I think it is true that our obligations provide some guidance whether or not they are presently achievable, but also that *genuine* moral obligations supply additional reasons beyond those explored so far - reasons which an unachievable obligation is unable to provide. It is these additional reasons, I believe, which form the basis of the point which Copp was attempting to make and which I will use to argue in favour of OIC. For if unachievable obligations should turn out to lack one of the integral pieces of any genuine moral duty, then logically our only such duties will be those we can fulfil.

Precisely what reasons, then, do our obligations provide for our moral deliberations? First, I should clarify that I am an externalist regarding the question of moral motivation, which is to say I believe moral truths like obligations provide their own reasons for why

we should abide by them, regardless of whether doing so aligns with our subjective desires. When I consider whether to do what I am obligated to do, there will usually be at least some reasons in favour of each of my available actions. The obligatory action(s) may be better for my health, for example, whilst the various non-obligatory actions might be easier, more pleasurable, and so on. Yet beyond these "contextual" reasons, each of which will have varying degrees of persuasiveness, in such cases there appears to be an additional reason in the obligation's favour: as mentioned in Section I, by denoting a certain action as morally required our obligation gives me a further reason to carry out *that* action rather than any other.[78] The content of this reason is, in essence, that the obligatory action is (morally speaking) the "right" thing to do of those open to me, and what I am required to do for my actions to be morally acceptable.

This additional reason is unique on two grounds: firstly, unlike the contextual reasons both for and against my various non-obligatory options, this reason appears in *all* cases where I have at least one genuine moral obligation. Secondly, it possesses a distinctive authority which overrides any other kind of reason arrayed against it. No matter how many reasons I may have to avoid performing the action which is required of me, when they are set against an action which is genuinely morally obligatory it seems the latter will win out,[79] even if there are no reasons to perform it other than the fact that it is morally required. The only exception to this occurs when I am faced with two options which are both genuinely obligatory, yet mutually exclusive: where each obligation provides this kind of compelling, authoritative reason to fulfil them, yet neither obligation defeats the other. This kind of epistemic deadlock is what drives the phenomenon of "tragic dilemma", and we will examine their implications for OIC in the next chapter.

Next, I wish to examine this version of events as it would hypothetically apply to unachievable moral obligations. If OIC were false and such obligations could therefore be genuine, it seems that when I am faced with the "choice" between attempting to fulfil an impossible moral obligation and doing something else instead, there should (as before) be a correspondingly powerful reason in favour of doing the obligatory, impossible action which overrides whatever reasons may be in favour of the morally optional actions. For if the obligation truly is genuine the same reasons ought to apply as in the previous

[78] For more on this argument that obligations provide reasons for (possible) actions, see Vranas (2007).
[79] "Win out", here, in the sense that it objectively settles the question of what I objectively should do- whether I am ultimately persuaded by this reason will, of course, depend on my moral character.

example, since it is still the “right” thing to do and what I am required to do if my actions are to be morally acceptable. But quite clearly, there would be no such authoritative reason on behalf of the impossible action in such a scenario as this. This is because although in principle there could be reasons to do something which we know we cannot do, what motivates the special kind of reason held by a genuine moral duty is that abiding by it is the morally correct action of those which are available for me to choose from under the circumstances.

Impossible duties, therefore, seem incapable of motivating a reason of this kind in favour of their fulfilment. We would still possess reasons to take them into account in our moral deliberations, as they would continue to steer us toward what we ought to do given our obligation is impossible (and hence, give us reasons to perform *those* actions). But they seem to offer no reason whatsoever in favour of their own performance, since we know we cannot do so. If such obligations could indeed be genuine, this would be highly perplexing. Part of the basic nature of an obligatory action, as we established at the beginning of this chapter, is that there are morally compelling reasons to perform it regardless of our subjective desires. Yet here we seem to have more reasons (albeit merely contextual ones) to do something which is not obligatory in any sense than we do to fulfil what is, supposedly, our genuine moral duty.

That our supposedly genuine moral obligation provides guidance about what we ought to do instead of fulfilling it, as we acknowledged before, does not dispel this confusion. Indeed, it adds to it: the reasons which our impossible obligation provides to perform the closest available - and morally best - action(s) to what is required appear to be reasons with the kind overriding authority that we would normally expect to be provided in favour of the obligatory action(s) themselves since, as we have shown, those reasons favour the best and most moral thing I can do. So we now possess contextual reasons in favour of my possible non-obligatory actions, distinctively authoritative reasons to perform the closest possible action to that which is genuinely obligatory but impossible for me, and seemingly no reason to perform the action which *is* genuinely obligatory but impossible for me. Therefore, it seems that what we have the most reason to do in this case - what we “ought” to do - is the best thing we can do, and not the morally obligatory but impossible action.

However, under such circumstances it is very tempting to wonder whether our genuine obligation in this case might just *be*, as elsewhere, to do the thing which I have such authoritative reasons to do (i.e. the possible "next best" action) and not the thing I cannot do. For it looks very much like that is where this line of thought ends up - I have a "genuine" obligation which I cannot complete, so I must therefore do the closest thing I can.[80] I possess the uniquely authoritative reasons to do that thing which one would otherwise associate with a genuine moral duty, and I do not have those unique reasons to act in the way which is supposedly obligatory for me in this circumstance. On what basis, other than specifically wanting to reject OIC, would we conclude anything here other than that my "next best" action is in fact what is genuinely obligatory for me, and that my original "obligation" is defeated precisely because it is impossible to fulfil?

Thus, we have a means of arguing from the action-guiding nature of moral requirements which supports my characterisation of OIC. If we accept that our moral duties serve to guide our action through the mechanism of reasons, then it should be clear that a moral obligation which is both impossible and genuine would behave very differently to how we would expect our moral obligations to normally function, even though it may still be helpful to us in making moral choices. The nature of this difference, and the apparent inability of such duties to motivate their own completion, gives us cause to question whether these are in fact true obligations as opposed to descriptions of the ideal state of affairs - morally informative, to be sure, but not required. A truly obligatory action, I have argued, provides overwhelming reasons in favour of its performance. One might say our genuine moral obligations are not merely helpful in our moral deliberations - they ought to *settle* our moral deliberations, assuming that one is a rational agent. Simply knowing what would be ideal to do is clearly insufficient for this task, as by definition our moral deliberations are deliberations between possible actions.

These purportedly impossible obligations would seem to be lacking in much the same way, unable to conclusively answer to the question of what I ought to do, and so the nature of our genuine moral duties themselves provides a powerful incentive in favour of my variation of OIC.

[80] We know that I must do this, for if I did **not** do the closest possible thing to my impossible obligation in this circumstance, we have already seen that this would be morally unacceptable (as in my earlier burning building case)

Section II.VI: Motivations of Necessity

So far, the motivations I have marshalled in defence of "ought implies can" have been predominantly concerned with the problems involved with impossible moral requirements and showing, by contrast, how OIC fits into very intuitive and coherent ideas about how our moral duties should behave. In this final section, I wish to expand the scope of my analysis to consider the wider moral picture of our obligations in the world where OIC and its similar principles are entirely absent. I wish now to call attention to various ordinary facts about our moral lives which we commonly take for granted in our discourse, but that we will struggle to justify if we assume that "ought implies can" is false. Depending on exactly how problematic this explanatory gap caused by the absence of OIC turns out to be, we will discover equally strong reason to accept the principle in general terms. As with my other points in support of OIC, I suggest these motivations may be grounded in the same basic moral principles which have led me to the variant of OIC which I defend here, and thus provide us with specific reasons to support it.

The first such assumption I shall examine is the idea that there are some situations in which we fail in our moral duties, and our only possible defence will be to appeal to our inability to complete them. For example, assume that I am morally obligated to visit my grandmother this evening, but am snowed in and cannot leave the house. In this case, the only reason I would not be blameworthy for failing to visit her (all other things being equal) is that I could not do so. We know this because if I were able to visit her and deliberately chose not to, I would clearly deserve blame for my selfishness. Now, whether we accept OIC or not all parties would presumably agree that my inability in this case *does* make a difference to my blameworthiness for failing to see my grandmother, even if it does not defeat my obligation per se. Even the interlocutor who rejects OIC entirely would, I imagine, be prepared to concede that the agent who knows she is obligated to see her grandmother and chooses not to is more at fault than the woman who recognises her duty and is prevented from fulfilling it. Despite their denial of OIC, then, they would still maintain that my inability to complete a moral obligation may still be of significant relevance in determining (or excusing) my blameworthiness for failing in it.

In light of the commentary I have already made about the nature of our moral obligations and their connection to blameworthiness, however, such a critic of OIC now faces a

challenge. Once OIC has been discarded, on what basis can we justify the very common belief that inability can often excuse or in some way mitigate our culpability for failing in a genuine moral obligation? Why does my inability to fulfil my obligation affect my blameworthiness in this way, given that blameworthiness necessarily arises from a failure to fulfil our moral duties,[81] and my inability no longer changes the moral obligation(s) I am under? It is only logical that any blame I might deserve for failing in an impossible obligation lapses once we factor in OIC, since OIC annuls the obligation entirely. But this explanation is obviously not open to our imaginary critic of OIC, and some explanation of the assumption that inability may exempt from moral wrongdoing will still be necessary.

Here, then, is a reason to support at least some formulation of "ought implies can." OIC both acknowledges and provides an explanation for this aspect of our moral thinking in a way that is simply not available to those who reject the principle in all its forms. Not only this, but I believe the underlying intuition of this exculpatory thought can be used to support my preferred form of OIC. The claim that an agent's inability to complete a moral obligation may excuse them from blameworthiness for that failure is grounded in the common belief that once I am unable (in the relevant sense) to carry out my obligations, it is no longer my fault that I do not do so. As such, it is unjust for me to be blamed or held responsible for my failure assuming that I did not sabotage my chances of success. These are essentially concerns about the fairness of our moral obligations and moral judgements and, for reasons which we discussed in Section IV, motivations of this sort are heavily connected to versions of OIC that are structurally very similar, if not functionally the same as the one I have espoused in this thesis.

OIC also underpins other important elements of our standard moral discourse and moral attitudes toward other agents. Frances Howard-Snyder[82] has called attention to another belief which the critic of OIC will not wish to reject, but will have difficulty justifying once the principle is jettisoned. Ordinarily, Snyder argues, we think that in many cases where we cannot do something which would otherwise be genuinely morally obligatory, we are often then subject to an additional genuine duty to perform the closest action that

[81] It is possible, of course, that those who reject OIC will simply reject this premise as well. To help counteract such scepticism, my defence of this principle in Chapter IV is deliberately agnostic regarding the truth of OIC.

[82] Howard-Snyder, Frances. 2006. '"Cannot" Implies "Not Ought"', *Philosophical Studies*, 130, 2: 233-246.

we can - perhaps to *try* to do the impossible action, or to do something which achieves similar results. We have previously seen this type of thinking during our discussion of the arsonist case in Section V, where my obligation to call the fire department was replaced by an obligation to help in some other capacity once the phone lines were cut.

Here, the defender of OIC is free to argue that in this position our genuine moral obligation is simply to do the morally best thing(s) that we can under the circumstances.[83] Our obligations therefore change over time according to what is possible from moment to moment. Such a move is clearly not open to the critic who denies OIC altogether, of course, because the "secondary" obligation(s) described above seem to exist solely in virtue of the fact that our initial obligation is impossible. If the cunning arsonist had not cut the phone lines, I would be morally required to call for the firemen and *not* to seek assistance elsewhere (although doing so may certainly be morally supererogatory), and it is only once this obligation lapses that another such duty replaces it. Once OIC is denied, however, our secondary obligations of this type appear to be entirely redundant - if the obligation to call the fire department persists despite its being impossible for its agent, it is unclear why or by what means a second, equally genuine obligation would be generated.

Once again, therefore, we have an apparently uncontroversial moral principle that both the defender and critic of OIC will wish to retain, but for which the supporters of OIC have a tailor-made explanation. Those who would reject the principle, by contrast, seem to have only three available replies: 1) reject that we have any secondary obligations in cases like this, and thus are only subject to the original and unachievable obligation, 2) suggest I now have two separate obligations, one to perform the original action (A) and another to perform the next best thing (B), or 3) claim that I now ought to do B *instead* of A. Let us now examine each of them in turn.

The first option of simply denying that we have any secondary obligations in cases like this appears to be a non-starter. We have already noted in Section V that regardless of our view of OIC it would be unacceptable if I, upon realising my otherwise obligatory action is impossible, chose to shrug my shoulders and do nothing. If I were under no other

[83]See Makinson, David. 1987. 'Fred Feldman. Doing the best we can. An essay in informal deontic logic.' *Philosophical studies series in philosophy*, vol. 35. D. Reidel Publishing Company, Dordrecht, Boston, Lancaster, and Tokyo, 1986, xiv + 244, *The Journal of Symbolic Logic*, 52, 4: 1050-1051 for an example of such an approach.

genuine moral obligations beyond the one I could not satisfy my doing this would be perfectly morally acceptable, but assuming I am not paralysed by shock or otherwise rendered incapable of any action, it clearly is not. Nor, intriguingly, does the unacceptability of this depend in any way on whether you think I am *still* obligated to do the impossible action in this situation. As Lawford-Smith has highlighted,[84] our moral obligations do not stop at merely telling us what ideally ought to be done, but rather also inform us of what we should do in the event that we do not or cannot act in the ideal way - "all the way down", as it were.

Thus, the critic of OIC is forced to choose between saying we are obligated to do A *and* B, or just to do B given that A is unavailable. For Snyder the former choice is bizarre, as "to say I ought to do A implies that I am [morally] permitted to do A. To say that I am permitted to do A implies that it is not the case that I ought **not** to do A. So the claim I ought to A implies it is not the case I ought not to do A."[85] But since A stands for the best action all-things-considered (e.g. find a phone booth and call the fire department), and B stands for the best thing I can actually accomplish (run screaming for help), A and B are mutually exclusive. This suggests that saying we are genuinely obligated to do both A and B in this scenario is a contradiction - I ought to do A, and hence it is not true that I ought not to do A, but I also ought to B, which by definition means I ought to do ~A. If I ought to do ~A, however, it is true that I ought not to do A. Thus, we find ourselves in explicit contradiction.

What the critic of OIC needs to explain, from Snyder's perspective, is why we are obligated to do B instead of doing A, and it is unclear how they may do so. The obvious reply that "if you cannot do A, you ought to do the next best thing" is not open to the critic of OIC, as said before, since what this really says is "if you cannot do A, you should do the next best thing instead of A" - and that, of course, entails OIC itself. Without another hypothesis which we might present to explain how this transference of our genuine obligations occurs, therefore, we appear to have found a second highly plausible moral principle that we will struggle to defend if we reject OIC.

[84] In Lawford-Smith, Holly. 2013. 'Non-Ideal Accessibility', *Ethical Theory and Moral Practice*, 16, 3: 653-669.
[85] Howard-Snyder (2006) p. 237.

At this point, one may detect a deeper problem which runs through the individual areas where our hypothetical critic of OIC has encountered difficulty. It seems that once we are comfortable, for instance, with the claim that I am morally obligated to save my drowning child even though I cannot reach them in time, or to call the fire department even though I lack access to a phone, it is logical to ask precisely what kind of inability is sufficient to serve as a defeater for my moral obligations, and under what conditions. Why, to compare two entirely unrelated examples, is it acceptable that I could be morally obligated to do something I cannot do where what I am unable to do is swim quickly enough to rescue my drowning child, but not (one presumes) where what I cannot do is travel back in time to prevent my child from drowning in the first place?

To put this point another way, it seems perfectly reasonable to ask the critic of OIC what, if any constraints are set upon our genuine moral obligations by any kind of inability whatsoever. Assuming they would accept that different moral agents may possess different moral obligations at different times - and it would be madness to suggest otherwise - then inability of some form must logically set at least some limits on what our obligations may be. At this point, then, we might also ask why *those* specific inabilities may prevent an agent from being genuinely morally obligated to perform an impossible action, whilst the kind of inability referred to by my definition of OIC (or indeed, any definition of OIC) does not. In any case, once we have accepted that inabilities of at least certain kinds may serve as defeaters for our moral obligations the debate over OIC can be seen as one of degree rather than kind. All parties involved accept that certain kinds of inability defeat obligations, and so the debate then is over which kinds of inability, under what conditions, defeat which instances of moral obligation. This, perhaps, is the best motivation that can be offered for the necessity of some interpretation of OIC.

The aim of this chapter has been to examine the formulation of “ought implies can” which will serve as the grounds on which to build my case for the Derivation of PAP, and to demonstrate that is an intuitive and intellectually rigorous understanding of OIC in its own right. In the process, I have outlined my definitions of each of “ought”, “implies”, and “can”, and sought to base them in a broader picture of how our moral obligations and our blameworthiness associated with those obligations function. In truth, however, this chapter has only provided half of my defence of OIC. Now that my position has been properly laid out, my next task will be to ascertain its viability by considering and replying

to a diverse series of arguments which will challenge the reasoning I have put forward throughout this chapter.

Chapter III - The Case Against "Ought Implies Can"

My argument in support of the Derivation has been primarily concerned, thus far, with setting out my favoured understanding of OIC and the most compelling reasons which underpin it. Over the course of the past two chapters, however, various questions and criticisms will undeniably have been building in regard to each aspect of that understanding. Therefore, I shall now take note of a cross-section of what I consider to be the most powerful and interesting of these objections, and test my chosen interpretation of OIC against as many of them as I can reasonably accommodate. Whilst I suspect the answers I will offer to those questions will not be universally convincing to the Derivation's opponents, they will be consistent with the wider moral principles which I have appealed to throughout this work. By doing this, I hope to provide explanations which will generalise to a wider range of criticism than I have been able to directly address here, and to demonstrate that those explanations are undeniably sourced in, and a direct product of, a more general stance about blameworthiness and moral obligation which is respectable in its own right.[86]

When an objection is made to "ought implies can", either in general terms and/or to a specific formulation of the principle, there are two potential forms it may take. The first is to deny the principle altogether, at least insofar as it proposes a necessary moral connection between "ought" and "can". Such objectors normally have a similar but subtly different moral principle in mind to replace OIC when they do this, for it is rare to find a philosopher who denies the connection between obligation and ability altogether. The second type of strategy is to object to one or more of the particulars of OIC - that is, the specific senses of "ought", "implies", and "can" being put forward - and suggest alternatives to replace them. As I am attempting to derive the principle of alternate possibilities from a specific interpretation of OIC, the objections which I will consider in this chapter will come from both of these categories. I will begin in Section I with a problem for OIC as a whole, regardless of our preferred form of it, and then move to discuss a variety of objections to its component parts in Sections II and III.

[86] Notably, there is one type of criticism of my sense of OIC which I will not yet be discussing. As noted in Chapter II, an important piece of my argument in favour of OIC depends on my so-called "fairness principle", which holds that genuine moral obligations are inherently fair to their subjects. I will explore the arguments in favour of this principle and respond in detail to the (numerous) possible objections to it in Chapter V.

Section III.I: The Problem of Moral Dilemma

The first objection to OIC that I wish to consider concerns the existence of genuine moral dilemma, or "tragic dilemma" as they are often referred to. In a situation of this kind, an agent is caught between two (or more) genuine obligations which are mutually exclusive: whilst the obligated agent can satisfy any of the obligations individually, they are certain to fail in at least one of them. These scenarios are a source of significant debate among moral philosophers, many of whom doubt their existence in the first instance. Mercifully, I am only interested in the specific problem these dilemmas are said to present for OIC, both in general terms and my specific formulation thereof. It has been argued that the existence of "tragic" dilemma would present a significant counterexample to any version of OIC that defends a necessary connection between "ought" and "can". Since there are strong reasons to believe that such dilemma are a legitimate phenomenon, not least that many of us claim to have actually experienced them in our everyday lives, we therefore have correspondingly strong reasons - at least in principle - to reject OIC.

To understand the proposed conflict between the genuine moral dilemma and OIC, we shall now examine a classic "tragic dilemma." Imagine a situation where my twin children are drowning in a lake, and from where I stand I can only reach one of them in time to rescue them. What should I do? Since I have a basic duty of care to my children which compels me to save them if I can, and assuming that my children are morally equivalent, I seem to have two genuine moral obligations in this case: one to save my first child and the other to save my second. Now individually I am perfectly capable of fulfilling each of these obligations, and this is consistent with OIC. However, when one possesses multiple genuine obligations at a certain time we would normally consider them to "stack", such that if I am genuinely obligated to A and also genuinely obligated to B, I am then genuinely obligated to do (A+B). This is known as the *Agglomeration Principle*[87] (AP), an important cornerstone of deontic logic. An irresolvable moral dilemma arises when an agent possesses two (or more) of these genuine yet inconsistent obligations, and therefore cannot complete their agglomerated obligation even though they can satisfy either of its parts.

[87] Williams, Bernard. 1973. *Problems of the Self* (Cambridge; Cambridge University Press).

How does OIC, whether on my interpretation or in general, accommodate these cases? It is unclear that it can. Whilst I naturally would not deny there are situations in which an agent *feels* obligated to do both of an inconsistent pair of actions, I cannot say that they genuinely are. OIC denies that an agent can ever hold an impossible obligation at the time they are called upon to exercise it, so if OIC is true it will be impossible for our obligations to agglomerate in this type of situation. However, since nothing about my duty to my children or their moral status has changed, my inconsistent obligations to each of them remain upon me. This position, taken together with the AP, results in explicit contradiction. If AP is correct, our mutually exclusive obligations will create a "conjoint" obligation, and if OIC is correct they will not - precisely *because* the obligatory actions are mutually exclusive. Since the AP is a point of basic logic and tragic dilemmas are a highly plausible (if intuitively unpleasant) occurrence, it is argued, the best resolution of this contradiction is that OIC is false. This is the position taken by Lemmon,[88] Trigg,[89] and more recently Jacquette.[90]

In order to avoid this conclusion, it will be necessary to reject another part of the inconsistent triad of OIC, the existence of tragic dilemma and the AP. Since abandoning OIC is clearly not an option for me, I am left with the choice between denying that cases of tragic dilemma are possible - and hence being dragged into the messy debate regarding their existence[91] - or rejecting AP and so dispute a seemingly foundational point of formal logic. Given this choice, the natural option would be to reject the possibility of irresolvable dilemma outright, which would also fit neatly alongside the justifications of OIC I have advanced thus far. Arguably, one might think I *must* deny that such dilemma are possible because much of my case in favour of OIC in Chapter II concerned the inherent fairness of moral obligations, which seems a difficult principle to reconcile with the possibility of being caught between genuine but inconsistent moral duties. Additionally, tragic dilemma appear to violate the action-guiding requirement for genuine obligations set out in the previous chapter: the knowledge that I am obliged to save both

[88] Lemmon, E. J. 1962. 'Moral dilemmas', *Philosophical Review*, 71, 2: 139-158.
[89] Trigg, Roger. 1971. 'Moral Conflict', *Mind*, 80, 317: 41-55.
[90] Jacquette, Dale. 1991. 'Moral Dilemmas, Disjunctive Obligations, and Kant's Principle that "Ought Implies Can"', *Synthese*, 88, 1: 43-55.
[91] See Plato. 1992. *Republic*, tr. by G. M. A. Grube, ed. by C. D. C. Reeve (Indianapolis; Hackett Publishing Company), Kant, Immanuel. 2012. *Groundwork of the Metaphysics of Morals*, revised edn., tr. by Mary Gregor, ed. by Jens Timmermann (Cambridge; Cambridge University Press) and Kant, Immanuel 2015. *Critique of Practical Reason*, revised edn., tr. by Mary Gregor (Cambridge; Cambridge University Press). For more recent examples, see Donagon, Conee, Davidson, McConnell et al in Gowans, Christopher W. ed. 1987. *Moral Dilemmas* (New York; Oxford University Press) and Mason, H. E., ed. 1996. *Moral Dilemmas and Moral Theory* (New York; Oxford University Press).

of my children, yet cannot, would seem to provide no guidance in deciding which of my available options I morally ought to take.

This is not my chosen response. Unfortunately for those caught in such dilemma, I think there really can be situations in which we are under two (or more) genuine moral duties which cannot both be satisfied, and their "tragic" nature is no mere illusion. Nor do I believe these obligations behave any differently, in the relevant respects, than a genuine obligation to do anything else. The fact that both obligations cannot be satisfied does not make the reasons in favour of each *individual* obligation any less persuasive, nor each of the obligatory action any less required. Indeed, it is precisely these qualities of the situation which create the tragedy and perceived[92] unfairness of irresolvable moral dilemma.

If I am to maintain both OIC and the above thoughts, there is now but one choice: I must reject AP, at least as a point about all genuine moral obligations. And I do reject it - not completely, as it is quite clear that in "normal" circumstances our ordinary, achievable obligations may often agglomerate in this way, but specifically in regard to those situations where the agglomerated obligation would be unachievable. To put it another way, in those cases where OIC and AP would conflict, OIC takes precedence: genuine moral duties do not agglomerate if the resulting "conjoint" obligation would be impossible for its agent to fulfil, in the sense of morally relevant ability given in Chapter II. In all other circumstances our obligations will agglomerate as we would normally expect. In the case of the drowning twins, I hence possess a genuine moral obligation to save one of my children, and a second and identical obligation to save the other, either of which I can fulfil. What I cannot do, and therefore have no obligation to do, is save *both* my children. If by some good fortune I were to become able to save both my children, however, my obligations would then agglomerate as normal since the resulting moral duty would no longer conflict with OIC.

This explanation neatly resolves the contradiction which threatens OIC, but on the surface it appears blatantly ad hoc. On what basis should we accept the proposed exception to AP that would allow OIC to apply in all cases of genuine obligation, as I suggest it does,

[92] I say "perceived" as I do not think these dilemmas are a counterexample to the idea that moral obligations are intrinsically fair. I will defend this position properly in Chapter V as part of my analysis of my "fairness principle."

rather than the other way around? Why, exactly, should we accept OIC's primary in this regard? Intuitively, it seems very odd that our moral obligations would not agglomerate in one specific kind of situation when they do as a matter of logical necessity everywhere else, especially given that I have already denied that there is anything special about the structure of the moral obligations involved in a case of tragic dilemma.

I acknowledge that my chosen position may appear rather suspicious at this point, and so I offer two methods of grounding this move. The first is to reject our previous starting assumption that agglomeration is a thing which just naturally happens in all situations, which in turn removes the impression that we are engaging in special pleading in regard to agglomerated, yet impossible moral duties. By doing this, I believe it will be possible to offer a superior analysis of how our obligations function in cases of irresolvable moral dilemma (and potentially elsewhere). The objection that this is merely an ad hoc defence to protect my specific form of OIC is mistaken, for limiting the scope of the Agglomeration Principle is neither a wholly unprecedented tactic nor one deployed solely by philosophers interested in defending OIC. Mary Mothershill, for instance, employs a similar restriction to defend the existence of tragic dilemmas at OIC's expense, arguing that "what bearing the axioms of deontic logic have on everyday moral reasoning has…yet to be made out. Why, for example, should we accept the agglomeration principle? At this very moment there are at least five other things I ought to be doing instead of writing this paper, but it does not follow that I ought to be doing all five things and also writing this paper."[93]

The most notable argument for restricting the AP's scope, however, comes from Bernard Williams' "Ethical Consistency."[94] Williams, who formally codified the principle in the first place, notes that "there are certainly many characterisations of actions...for which agglomeration does not hold, and what holds of each action taken separately does not hold for both taken together: thus it may be desirable or advisable or sensible or prudent to do A, and again desirable or advisable to do B, but not to do both A and B."[95] His preferred example is a man torn between which of his two loves he should marry - whilst such a man clearly desires to marry one woman and also to marry the other, Williams

[93] Mothersill, Mary. 1996. 'The Moral Dilemmas Debate', *Moral Dilemmas and Moral Theory*, ed. by H. E. Mason (New York; Oxford University Press), pp. 66-86.
[94] Williams, Bernard and W. F. Atkinson. 1965. 'Symposium: Ethical Consistency', *Proceedings of the Aristotelian Society, Supplementary Volumes*, 39: 103-138.
[95] Williams and Atkinson (1965), pp. 181-2.

argues, he clearly also does not wish to marry both of them. Now, one might object that an action's being genuinely morally obligatory is sufficiently different from its being desirable, useful and so on that this is not a clear counterexample to the universality of AP in regard to our moral duties, and Williams accepts this. The above point is simply intended to show that it is not self-evident that moral obligations agglomerate in the first instance, and if a more plausible picture of how our obligations behave in the case of irresolvable moral dilemma emerges from rejecting the idea that they *always* agglomerate, then that is what we ought to do.

It should come as no surprise that Williams also argues that modifying the scope of AP in the way which both he and I have suggested does, in fact, present the most persuasive (and accurate to our experiences) view of how our moral obligations function in cases of tragic dilemma. Assume that I am obligated to do A and, quite separately, to do B, but to do both A and B is beyond me. Because "ought implies can", I am therefore not morally obligated to do (A+B) at that time. However, it is also true that in this case I retain the power to do A *or* B until the point that I commit to either action, and thus render the other option unachievable. Call this point T. In the drowning twins example, I can choose which of my children to save until after I begin swimming out to one of them. From this, we can deduce that *if* I will not be able to do A (or B), at T, then it will not be the case that I am obligated to do A (or B), at T, even though I was obligated to do A and to do B beforehand. Thus, if I *do* perform action A at T, Williams argues, it is not then the case that I am obligated to do B as well, since I no longer can. The irresolvable dilemma vanishes once an irreversible decision has been taken, since that decision removes the possibility of succeeding in the unchosen obligation, and hence that obligation itself.

I think this description matches the situation of an irresolvable dilemma perfectly, as well as how our obligations change, in general, with our available options over time. If in such a dilemma I choose at T to do A over B, then to do B will no longer be an available option for me. Therefore, according to OIC my prior obligation to do B is now defeated, because to do B *after* A is impossible. Yet at the same time, it is still correct that I was genuinely obligated to do B before I chose which of my obligations to fulfil, when I could have fulfilled it and that I might therefore have potentially acted wrongly in choosing not to do so. All that OIC forbids is the obligation to do B after having done A, as this is the only point at which it is *not* possible, and this arrangement creates no problem for the principle in respect to cases of tragic dilemma.

I find Williams' argument here to be both persuasive and consistent with the moral principles I have advanced throughout this work. I find that modifying the AP in the way he has suggested, so that it does not permit an impossible moral obligation, allows us to avoid the explicit contradiction discussed earlier and - equally importantly, I feel - respects the essential "spirit" of the concept of a tragic dilemma itself. By this, I mean that our solution to the problem of these dilemma does not attempt to retroactively dismiss one (or more) of the original, competing obligations which formed the dilemma in the first place. At the time when I was forced to choose which of my drowning children to save, I really was under two genuine obligations which I could have fulfilled - but not together. Having chosen which of my drowning children to save I knowingly forfeit my chance to save the other, and to discharge the associated obligation to do so. It is from this fact that the "tragedy" of such dilemma arises, for whilst I am never under an obligation which I cannot fulfil, the nature of the circumstances is such that I cannot fulfil all of the obligations which I possess.[96] The waiving of the AP when it would create an impossible obligation therefore results in an account of irresolvable moral dilemma which best describes our experience of a situation of this kind.

The second justification for rejecting an AP which is properly universal in its scope concerns the initial motivation for the Agglomeration Principle itself. Whilst I am unqualified to discuss the justification of the principle within the world of formal logic, I think an exploration of the moral basis behind AP will provide its own explanation of why we should not expect our genuine moral obligations to agglomerate into unachievable duties. I propose that the primary reason we support (or rather, ought to support) AP in regard to questions of moral obligation is that almost everyone accepts a tacit, general obligation to discharge as many of our genuine moral obligations as possible. This additional obligation is expressed by the Agglomeration Principle's ordinary function: when we hold two disparate moral obligations, AP mandates that we are required to carry out both.

[96] An alternative strategy would be to classify one's obligations in a tragic dilemma as a single disjunctive duty: that is, to do (AvB), where A and B are obligatory yet mutually inconsistent options. Whilst consistent with my formulation of OIC, I reject this strategy because I feel it is explanatorily insufficient in regard to the moral conflict felt by the agents in such situations. For more on the problems with this approach, see Jacquette (1991).

If one is inclined to doubt an obligation of this sort, consider this modified example: imagine now that three of my children are drowning in the lake, and I can save only two of them. However, because I wish to maximise my chance of success, I opt only to save my favourite child and leave the other two to die. Despite the fact that I have inarguably fulfilled one of my three moral obligations in this case, my actions are clearly morally unacceptable because I also possess a genuine obligation to save as many as of my children as I can - a distinct duty which agglomerates from the individual obligations I hold to each of them. Importantly, my actions are not morally unacceptable because I failed to save all three of my children, because there can be no such obligation under OIC. The reason why we ought to accept AP in our general discussions of moral obligation is therefore clear: it expresses our general duty in such situations to do everything we can to meet our obligations.

If *this* is the purpose of the AP, however, matters become more complex: in the cases of tragic dilemma we are discussing, it will still be both possible (and required) to fulfil as many of our obligations as possible, even if they do not agglomerate into an impossible obligation. If the normal basis for applying the AP to our genuine moral duties is that we ought to fulfil as many of our obligations as possible, then it is only logical that it would be superseded by OIC when the two clash. For AP to supersede OIC, rather than the other way around, would go beyond the scope of this underlying duty to fulfil as many of our genuine obligations as possible. An AP which applied to all cases of moral obligation, even where the resulting obligation would be unachievable, would seem to express a much more demanding duty: to fulfil all of our of moral obligations, without qualification. Yet we clearly possess no such stronger duty, because as shown in Chapter II ability places at least some constraints upon our obligations regardless of the veracity of OIC. As such, if we accept my proposed rationale for applying AP to cases of moral obligation in general we have good reason to reject the idea that obligations may agglomerate into impossible requirements. Therefore, we can safely deduce that if OIC and AP are both correct, the former should naturally take precedence over the latter in cases where they conflict.

Despite these observations, there is a persistent worry which dogs the idea of a genuine duty to fulfil as many of one's obligations as possible. Even if we restricted such a duty to only apply to our known moral obligations, it will be objected that this obligation is too demanding to take seriously. Given that most moral agents are what we might call

morally flimsy, they will often fail to fulfil all (or even most) of their moral duties at any given time. This is not to imply that these obligation failures are premeditated or the result of deliberate wrongdoing, merely a recognition of the reality that most agents are not morally perfect. If a duty to fulfil all our moral obligations existed, therefore, it would be difficult to avoid the conclusion that a great majority of agents will have acted wrongly in a great many cases, for having failed to maximise the number of obligations completed, and would deserve blame accordingly. But it will obviously be argued that this is an unfairly high bar to set for morally acceptable behaviour - to discharge all of one's moral obligations at any given moment, even those which are general rather than specific - and so clearly, we possess no such duty.

Whilst I sympathise with this concern, I do not think it is a fair objection. I acknowledge that most moral agents have fallen short of at least some of their moral duties because most people are morally imperfect. Combine this point with the Derivation's supporting principle that blameworthiness arises from failure in a genuine obligation, and the result is indeed that most agents are (and have been) blameworthy for various obligation failures, be they major or minor, at various times. Whilst this may be an uncomfortable conclusion, it ought not to be seen as problematic or even abnormal. Recall that the mere fact of being blameworthy for failing in a genuine obligation is not necessarily a serious moral offence. Some moral obligations, whilst genuine, deal with relatively small matters and the level of wrongdoing for failing in them (as well as the amount of blame which is deserved as a result) will therefore not be very significant. You deserve considerably less blame for forgetting to feed your cat today than for poisoning it, because the obligation to feed one's pets is of lesser importance than the obligation to avoid senseless cruelty, and so on and so forth. Even if it is true that most agents are guilty of many obligation failures at any one moment, therefore, the level of blame which they deserve for those failures will often be relatively low, because many of our genuine moral duties concern actions of relatively minor moral importance.

Furthermore, there is an important difference between being blameworthy for an action and it being appropriate for someone to *actually blame* you for that action. Whether an agent deserves blame for an action is a matter of whether that action was morally wrong, and/or whether they have an excuse for its performance. By contrast, whether it is appropriate for another agent to blame them for that action is primarily a matter of where they stand in reference to the blamed party, usually because of some sufficiently close

relationship or because the obligation failure was at their expense in some way. This is important in determining whether our purported obligation to maximise the number of moral duties completed is excessively demanding, because we have not been clear about what, exactly, is unacceptable about the conclusion that most agents will naturally carry a certain degree of blameworthiness for failing in various moral obligations in their day-to-day lives. I doubt this seemingly trivial fact is the cause of such great consternation in and of itself. Rather, I suspect what prompts this objection is a fear of something worse: that on my model, an agent will be under too many and too stringent moral obligations, fail to fulfil as many of them as we judge they should, and hence suffer more actual blame than they fairly deserve for what are essentially reasonable moral imperfections.

This intuition is understandable, but mistaken. What makes it appropriate to serve blame is a matter of one's relationship to the agent who is its recipient, and an overly demanding moral system may make it too easy for agents to have that relationship. But there is nothing about my proposal which modifies the conditions under which such a relationship may be formed, and so there is no reason to think most agents would in fact suffer any more blame if they possessed an obligation to complete as many of their duties as possible than if they held no such duty. To appeal to a genuine obligation of this kind to motivate the primacy of OIC over AP in cases where the two principles conflict, and to provide reasons to accept my model of how our obligations behave in cases of irresolvable moral dilemma, is therefore not overly demanding and we have nothing to fear from such an explanation. As such, cases of tragic dilemma do not provide a counterexample to "ought implies can."

Section III.II: Unfortunate Implications

In discussions concerning which understandings (if any) of OIC are the correct ones, the most attention has generally been paid to the character of the obligations which are bound by the principle and the precise nature of the abilities required to complete them. But potent challenges also await the precise link between "ought" and "can" which I have previously outlined, and this form of argument should not go unaddressed. As mentioned before, objections of this sort rarely deny that there *is* a link of some kind between moral obligation and some form of relevant ability, for fear of the difficulties highlighted in the preceding chapter. Instead, they commonly attempt to undermine the characterisation of

this link as a hard, morally necessary connection, and therefore severely limit the scope of OIC rather than falsify it outright.

A prime example of such an argument can be found in Sinott-Armstrong's case against OIC.[97] There are, he argues, three ways in which "ought" might imply "can": "ought" may *entail* "can", *presuppose* "can", or *conversationally implicate* "can". Which type of implication we interpret OIC as having, Sinott-Armstrong suggests, will determine the truth-value of certain "ought" statements. If "ought" entails "can", then it is false that an agent who cannot perform a given action is obligated to perform it, whereas if "ought" merely presupposes "can" it will be neither true nor false that that agent has this obligation. But if the implication is conversational, as Sinott-Armstrong claims it is, then even if OIC is true as a general (conversational) principle it might still be true that a specific agent is morally obligated to do something which they cannot do.

It is this last categorisation which poses a challenge for my particular understanding of OIC, for Sinott-Armstrong believes that the sense of conversational implicature is the only one in which "ought" can be meaningfully said to imply "can". If this is correct it would prove fatal for my version of OIC and the wider Derivation which depends upon it, since the form of OIC I defend and from which PAP is derived utilises what Sinott-Armstrong would class as an implication of entailment - in every case of genuine moral obligation, its agent is capable of fulfilling it in the relevant sense.[98] Therefore, it is false to suggest we hold any moral obligation which we cannot complete. Whilst he offers an argument against both the entailment and the presupposition models of implication, it is thus only the former which poses a danger to my Derivation of PAP. As such, I shall now investigate that specific part of his case against OIC and judge how best to reply to it.

Sinott-Armstrong's case against "ought" entailing "can" is centred on the objection that an agent cannot exempt themselves from their own moral obligations by making themselves unable to complete them. If "ought" did entail "can", he argues, this is something which would clearly be possible for the obligated agent to do (whether deliberately or by accident) and so we ought to reject any variant of OIC which employs

[97] Sinnott-Armstrong, William. 1984. ''Ought' Conversationally Implicates 'Can'', *The Philosophical Review*, 93, 2: 249-261.

[98] Interestingly, I do not think it would fatal to reframe my OIC as a presupposition, since the Derivation only requires the claim that it is not true that you are morally obligated to perform an unachievable action to function. But this is a discussion for another time.

this type of implication, as mine does. Suppose I have promised to meet a friend at 6pm, but being a poor friend I opt at 5pm to go to the cinema at instead. This, in turn, makes it impossible for me fulfil my obligation to meet my friend at 6pm. Sinnot-Armstrong argues that if "ought" entails "can", I would hold no obligation to fulfil my promise to my friend at 6pm, since it is no longer possible for me to do so. Further, since I no longer hold an obligation to meet my friend at 6pm I cannot then be blameworthy for failing to do so, because both sides agree that blameworthiness occurs only from failing in a genuine obligation and any obligation I had in this case lapses once I have gone to the cinema instead.

Yet this analysis of the case is nonsensical. Our intuitions, it is argued, pull strongly in the direction that my obligation *does* persist after I have gone to the cinema at 5pm and I *am* therefore blameworthy for failing in it, even though I could not have met my obligation at that later point. If my friend were to call me angrily at 6.30 to ask where I am, or at 5.30 to remind me of our upcoming meeting, it would obviously be no defence at all to say "Well I can't fulfil my obligation now, can I? Don't you know that ought implies can?"

This problem appears in more subtle ways, too. Imagine that when I make my promise to meet my friend I know I will not be able to fulfil it at 6pm. If this is true and "ought" entails "can", then it seems there will never be a point at which I am obligated to keep my promise, because I will never be able to. For as we discussed concerning OIC and time-indices, one cannot usually fulfil an obligation which has a particular time-index (i.e. "Do X at T") either before or after that specific point in time, and so there will be seemingly no window in which I can keep my promise to "meet them at 6pm" - it cannot be fulfilled *at 6pm*, nor before or after, much like the obligation from Chapter II to present a paper next week. Yet this, too, is strange because we have at least a prima-facie if not always genuine duty to keep the promises we make, and so the act of promising ought to make a difference to my obligations and to my moral status for failing in them.

If "ought" entails "can", though, this is apparently not the case: whether I make the promise or not, I hold no moral duty to meet my friend at the specified time (since I never can), and therefore cannot be blameworthy for not doing so. Even more strangely, if I *did* promise to meet my friend knowing I could not do so, and then by some means became able to meet them after having made the promise, I would *then* become obligated to fulfil

my promise. This is despite it being entirely irrelevant to my obligations beforehand, due to my prior inability to keep my word. These observations lead Sinott-Armstrong to the conclusion that since this is clearly not how our obligations and blameworthiness function in cases such as this, it is wrong to categorise the "implies" portion of OIC as an entailment and hence any argument which depends on this, as the Derivation does, is false. For my part, however, I believe him to have entirely misunderstood how my version of OIC would assess his counterexamples, and his case against "ought" entailing "can" is unconvincing. Crucially, I believe Sinott-Armstrong to have wrongly equivocated between an agent being released from an obligation and an obligation dissipating because it is no longer achievable. It is this mistake which leads him to the faulty conclusion that if "ought" entails "can", an agent who deliberately sabotages their chances of completing a moral obligation will not be blameworthy for failing to fulfil it.

Sinott-Armstrong's argument is, however, accurate in several other respects. In particular, "ought" entailing "can" does technically allow for the possibility that an agent might defeat their own obligation by rendering themselves unable to complete it. Forms of OIC like the one I favour are exceptionless, and so no agent can ever be under an obligation which they cannot fulfil regardless of the reason why they cannot fulfil it. Additionally, regarding the specific case of a promise you know you will never be able to keep, I also grant that (despite appearances) you will never hold a genuine obligation to fulfil that promise until such a time that you become able to fulfil it. In short, I concur with Sinott-Armstrong about how our moral obligations function in such cases of "self-sabotage". This, however, is not the true danger of the objection to "ought" entailing "can". The real problem is the ability it supposedly grants the unscrupulous to exempt themselves from blameworthiness by forcing their unwanted obligations to lapse - an ability, Sinott-Armstrong claims, they clearly do not have.

If for "ought" to entail "can" *would* allow for the ability to do this, it would be a compelling argument against my understanding of OIC. Mercifully, it does not. Sinott-Armstrong is correct, of course, that sabotaging one's chances to fulfil an obligation (and thereby defeating it) does not free you from any blameworthiness which you would normally deserve for failing to complete it. But "ought" entailing "can" does not permit this sort of ethical sleight-of-hand either. That conclusion arises from a misreading of the relevant portion of my Derivation:

2) If am agent is morally blameworthy for an action, that action is morally wrong.

3) If an action is morally wrong, we are morally obligated not to do it.

Sinott-Armstrong appears to view these premises as amounting to: "for an agent to be morally blameworthy for performing (or failing to perform) an action at T, they must have been under the corresponding obligation(s) *at T*." For blameworthiness only arises from failing in a moral obligation, and if I have no obligation to perform an action at a certain time, I cannot be blameworthy for not meeting it. This interpretation, however, misunderstands the process by which an agent becomes blameworthy for failing in an obligation in the first place. Once an agent has failed in a genuine obligation, assuming it had to be met by a certain point, that obligation will lapse and whatever blameworthiness an agent acquires as a result will attach to them.[99]

This can be clearly seen in the case of promise-keeping: having failed to meet my friend at 6pm, I am not still under that obligation at 10am the next day because the relevant moment has passed. I am, however, still blameworthy at that latter time for failing in my obligation yesterday, even though the obligation itself has expired. This is very different to an obligation lapsing because its agent has *been released* from it - say, if my friend had called me at 5.45 to cancel our dinner plans - wherein no such blameworthiness occurs because the obligated agent is no longer subject to it.

Thus, by attempting to escape a genuine obligation by sabotaging my ability to fulfil it, what I am actually doing is ensuring I will fail in the obligation I currently possess. By going to the cinema at 5pm, I do not exempt myself from my obligation to meet my friend. I simply choose not to fulfil my obligation, which lapses as normal once the failure has occurred. If the sabotaged obligation was genuine, therefore, there is no reason why my self-sabotage should be treated any differently (in terms of the resulting blameworthiness) than if I had failed in my obligation in a more normal way - perhaps by forgetting about the promise or falling asleep. Indeed, one might think we have substantially greater reason to attach blameworthiness to the agent who fails in their obligation due to self-sabotage than to one who fails in the more normal ways, because the former is premeditated. It

[99] Note that if the violated obligation is general rather than specific, such as the duty of kindness or charity, the obligation will not dissipate in this way, but I will still deserve blame for having failed to fulfil it in that particular instance.

should *not* be treated, as Sinott-Armstrong seems to do, as a case wherein the agent has successfully released themselves from their duty. Under the entailment model of OIC, therefore, there is no method by which an agent can cheat their moral duty in the way Sinott-Armstrong suggests, and his argument against that model fails accordingly.

Sinott-Armstrong does, however, have a response to this reply: thus far, my Derivation has been dependent on an inherent link between moral obligations, wrongdoing, and blameworthiness. Certainly, he may agree that those who deliberately sabotage their chances of completing their obligations are more at fault than someone who fails in them as part of the normal course of events. The saboteurs appear to have committed a double offence, being blameworthy not simply for failing in their obligation but also for the attempt at subverting their genuine moral duties. So even if "ought" entailing "can" does allow us to blame agents who sabotage themselves as normal, there is now a further question. If the saboteurs are blameworthy specifically for ensuring their own failure, then according to the Derivation's own internal logic there must be some *additional* obligation which they have also failed in - an obligation to avoid sabotaging one's chances, as it were. This second obligation failure is what accounts for the additional blameworthiness. But whence cometh this additional obligation, and might we not worry about an infinite regress?

Thankfully, I have already answered this question in my earlier discussion of irresolvable moral dilemma. The obligation not to sabotage your chances of completing a particular obligation is a subsidiary of a genuine obligation I have already argued in favour of: to fulfil as many of our moral obligations as we can. For once we grant this latter obligation, which we should if we accept AP as a generally applicable principle, it becomes obvious why you would deserve blame for deliberately ensuring (as the saboteurs do) that you will *not* do that. This reply also has the handy effect of dispelling any potential infinite regress of our obligations in this type of situation. There are only two obligations "at play" in the case of self-sabotage: the obligation to perform the action(s) in question, and my tacit obligation to fulfil as many of my moral duties as possible. Any actions we are required to undertake or avoid are in service only to these obligations. For example, the latter obligation requires me to ensure I can keep my promise to my friend, which in turn requires me to not be doing anything else at 6pm, be in the correct area at 6pm, and so on and so forth.

A useful comparison can be found in the game of cricket. One has various obligations to obey the individual rules of the game - two batsman at the crease, eleven players a side, and so on. But one also has a secondary obligation to *play fairly*: to not merely abide by the technical letter of the rules, which satisfies the first obligation, but also to respect the spirit of those rules and not find clever ways of working around them. So it is with the rules of the moral "game." Avoiding blameworthiness is not merely a legalistic matter of being in violation of no obligation which we hold at the present time. It is also a matter of making, and having made, one's best effort to fulfil as many of one's genuine moral duties as possible. To fail to do this, as the saboteurs clearly have, is to fail in an obligation no less genuine than any other, and to deserve additional blame accordingly.

As such, we now have an excellent explanation for how understanding "ought" to entail "can" explains not just the blameworthiness of those who sabotage their chances of fulfilling their moral duties, but also the common intuition that doing so is morally worse than merely failing in them normally. Far from showing that "ought" does not entail "can", therefore, Sinott-Armstrong's argument shows this account of the "implies" of "ought implies can" to align perfectly with how we should best - and I believe naturally do - understand the relationship between obligation, ability, and blameworthiness.

Section III.III: What I Can (Really) Do

Having defused the challenge to the sense of "implies" in my formulation of OIC, I now wish to discuss the more controversial matter of its chosen "can." In this section I will examine two types of objection to the sense of morally relevant ability I have offered, each of which claims I have set the threshold for being able to complete a moral obligation in the wrong place. One might firstly challenge my definition of the morally relevant sense of "can" and "able" on the basis that its requirements are too strict, permitting agents to escape genuine moral obligations which they should properly be considered to be "able" to meet. Alternatively, the sense of ability which I argue is required to be "able" to complete a moral obligation may be attacked as too lenient, and rejected on the opposite grounds that it will shackle agents with genuine duties which they *cannot*, in the truly relevant sense, fulfil. Whilst these objections cannot both be correct, the essential charge is the same: that the type of ability which I consider to be relevant to both OIC and PAP will give the wrong answer to whether an agent is "able" to fulfil their duty in certain cases of genuine moral obligation.

But before we combat these challenges, we should first reacquaint ourselves with my definition of the relevant "can". The sense of "can" or "able" implied by my understanding of OIC requires a specific, rather than general ability to act in the required way(s), for retaining the general capacity to perform an obligatory action is clearly not always sufficient to be able to fulfil an obligation to perform that action at a particular time. To say X "can do Y" or "is able to do Y" in the sense relevant to OIC is for X to possess the *Prerequisites* and *Opportunity* to do Y. The "Prerequisites" here refer to the various skills, knowledge base, and physical and mental capacities which are required to execute Y. The Prerequisites for a game of tennis, for instance, include a basic understanding of the rules, an opponent, and the ability to run around and hit the ball. Having the "opportunity" to do Y is more complex, requiring an accessible possible world in which I successfully perform the required action(s) at the required time. A possible world is "accessible" in this sense if it is functionally identical to the actual world *except* for my successful attempt to perform the obligatory action, and the consequences thereof. An agent can only possess a genuine moral obligation to act in a certain way, according to OIC, if they will satisfy the Prerequisites and have the opportunity to act in that way at the time they are called upon to fulfil their obligation.

Let us now take these two objections in turn, beginning with the reply that my sense of morally relevant ability is missing certain necessary conditions, and so even an agent that unambiguously possesses it in respect to a given moral obligation will still lack some important component of moral agency. Hence, some additional capacity on the obligated agent's part will be required beyond the two I have specified before they are "able" to complete it in the morally relevant sense. What capacity might this be? One reasonable suggestion has been defended by Bergström[100] and Howard-Snyder[101]: that ability, in the sense relevant to moral duty, requires knowing *that* we have the obligation in question as well as knowing *how* to complete it.

To borrow Alvarez's explanation of this thought, "What about agents who have all the relevant facts in view but fail to meet their obligations because they do not have the right moral beliefs? If their ignorance of their obligations derives from mistaken moral beliefs or from ignorance of the moral significance of the facts they have in view, should they be

[100] 'Bergström, Lars. 1996. 'Reflections on consequentialism', *Theoria*, 62, 1-2: 74-94.
[101] Howard-Snyder, Frances. 1997. 'The Rejection of Objective Consequentialism', *Utilitas*, 9, 2: 241-248.

excused for failing to meet their moral obligations?"[102] In other words, if an agent were to satisfy the requirements I have set out but were in ignorance of - or simply denied - the relevant moral obligation (and/or how to complete it), then it seems the only means by which that agent could complete their duty would essentially be through blind luck.[103] Since an agent that failed in such a duty clearly would not deserve blame for that failure, the argument goes, should this not also lead us to question whether that agent was properly able to accomplish their duty in the first place?

The most defensible version of this additional condition that I can determine runs something like this: to be "able" to fulfil a moral duty, an agent must necessarily either 1) be aware that they possess that duty (and how to complete it), or 2) the evidence available to them must be such that they could reasonably be expected to be aware of that duty (and how to complete it). Even this formulation, however, seems fraught with serious epistemic issues. To take but one example relating to an agent's epistemic responsibility, what exactly constitutes this "sufficient evidence" such that they ought to have recognised they possessed a particular moral duty? After all, even trained moral philosophers may disagree radically about the content of our moral obligations in a specific circumstance, and the average agent generally does not (and should not be expected to) engage in such rigorous moral analysis.

In addition, there seems to be an inherent issue with sourcing an agent's genuine obligations (or lack thereof) in what is, fundamentally, a matter of what they "ought to have known." Suppose an agent is unaware of their moral duty, despite overwhelming evidence in favour of it. Under the proposed additional condition on ability, that agent may still rightly be judged as able to complete their obligation, all other things being equal, and would, I submit, be blameworthy when they fail to do so as a result of their ignorance. However, if (as I have consistently argued) blameworthiness only follows from failure in a genuine obligation, this would indicate that our agent also possessed a genuine duty to have known what their obligations were, and how to complete them, in addition to their obligation(s) to act in a particular way.

[102] Alvarez, Maria and Clayton Littlejohn. Forthcoming. 'When Ignorance is No Excuse', in *Responsibility: The Epistemic Condition*, ed. by P. Robichaud and J. Wieland (Oxford University Press).
[103] We will return to cases of this kind- where an agent may fulfil a moral obligation, but it is not "up to them" whether they will fulfil it- in Chapter V, where they will receive an extensive analysis as part of the defence of my "fairness principle."

By extension, therefore, this additional condition on ability can seemingly only work by assuming a similar duty on the part of *every* agent to be able to recognise their moral duties at any given point in time - and even I, having no objection to large and complex sets of moral obligations, must draw the line at such a suggestion. For this is an obligation which I do not think can be met - whose Prerequisites, to refer back to my own understanding of "can", cannot be consistently satisfied - on any plausible understanding of morally relevant ability whatsoever. Thus, OIC unambiguously defeats the possibility of such a duty. Should we wish to maintain this additional condition of the morally relevant ability, our only alternative is to deny that such an agent is blameworthy for his failure even with such overwhelming evidence pointing to his obligation, as Alvarez considers doing. This appears to be no better, as if this agent is to be released from his moral obligations and/or excused from blameworthiness on account of his ignorance (and there is no other explanation of the fact), then it appears that every agent who fails to recognise their moral obligations for whatever reason, regardless of the amount of evidence in favour of them, should also be excused by that same logic. This, surely, cannot be correct either.

Having now dealt with the objection that my chosen definition of "can" is missing one or more of the necessary conditions of morally relevant ability, let us now examine the more powerful objection that it sets *too many* conditions on such ability. Put simply, this reply holds that the minimum requirements to be capable of performing an action, in the morally relevant sense, are much lower than I have been arguing. It has even been argued in some quarters that certain kinds of obligation have *no* such conditions which define ability, in direct defiance of OIC. If correct, this would indicate an agent that lacks the Prerequisites and/or the opportunity to fulfil their genuine moral obligation might still be "able" to fulfil it in the properly relevant sense. The charge is once again, then, that my understanding of ability will provide the wrong verdict of our ability to fulfil certain genuine moral obligations, and therefore ought to be rejected.

The work of Peter Graham[104] is a particularly interesting example of this school of thought, for although Graham is a staunch critic of OIC he and I agree completely about the nature of morally relevant ability in a wider context. OIC, Graham declares, is a false principle that has persisted because it offers an excellent explanation of a certain pair of

[104] Graham, Peter. 2011. 'Ought and Ability', *Philosophical Review*, 120, 3: 337-382.

moral facts. The first of these is that everyone lacks a genuine obligation to (for example) snap our fingers and end world poverty, halt all wars, and/or generally make the world a perfect place. The second such fact is that if it *were* within our power to accomplish those things by snapping our fingers, we would then undeniably have a genuine obligation to do so. Since the only apparent difference between the two cases is our ability to perform the action in question, we therefore have an excellent basis to assert some form of OIC. Graham, however, argues that although this explanation does hold true for most moral duties, there are clear exceptions. Some types of genuine moral obligation, he maintains, may persist even in the absence of the relevant ability to complete them - and not just in the sense that I cannot *now* fulfil a presently-held obligation to give a paper next week, or meet a friend for dinner tonight.

Graham's chosen example of such an obligation has two parts. First, he invites us to imagine a case called *Transplant*: a surgeon has ten patients who are dying of various organ failures, and no suitable replacements are available. Miraculously, however, two men in the waiting room have, between them, all the organs necessary to treat the ten patients. The surgeon attempts to incapacitate the two with the aim of vivisecting them for parts, but is prevented from doing so by a vigilant hospital janitor who is forced to kill her to save the pair.

Graham identifies two supposedly basic facts about this case:

1) The janitor's killing the surgeon is morally permissible.

because

2) If the janitor had not killed the surgeon, the surgeon would have wrongly killed the two men.

Though we can grant both 1) and 2) without admitting to a link between them, it seems plausible enough that 1) is true in virtue of 2), and Graham offers a thought to support this. If we were to alter *Transplant* so that the surgeon's killing would be morally permissible (perhaps the dying patients are critical to avoiding an impending apocalypse?) this would seem to automatically make the janitor's action *im*permissible in turn. If we grant this, then we have strong grounds to think 1) is justified on the basis of

2). But if we accept this reasoning in respect to *Transplant*, a slight modification of the case will lead to a counterexample to OIC as I seek to interpret it:

Transplant-2: Everything is as it is in *Transplant*, except the surgeon is irresistibly moved to act by some relevant consideration - a powerful mind-controller has taken hold of her, for example.

Here, 1) still seems to hold true. Indeed, the addition of irresistible compulsion is of no consequence to 1)'s truth, Graham suggests, because what explains 1) here is the same thing that explains it in the original *Transplant* - that is, that 2) is true. But if we grant that the surgeon truly could not have refrained from attempting to kill the two men, and yet maintain that it was still wrong for her to do so - as 2) claims - then we have a counterexample to OIC. For 2) mandates that the surgeon's action is morally wrong and therefore that she is obligated not to do it, even though she cannot (in the relevant sense) stop herself from doing it. If she possesses this obligation even in *Transplant-2*, therefore, an agent can possess at least a certain kind of genuine obligation even whilst being unable to fulfil it, and so my preferred understanding of OIC is false. The Derivation of PAP thus collapses accordingly.

Perhaps not surprisingly, I have a significantly different analysis of *Transplant-2* to Graham's. I think Graham is right that 1) holds in both this case and in the original *Transplant*, since the janitor has a genuine obligation to stop the surgeon and fulfils it by the only means available. However, I deny that in *Transplant-2* 1) holds because of 2), because I think 2) is false in *Transplant-2*. I do not deny that if the janitor had not acted the surgeon would have killed the two men, merely that the surgeon's action in this case is morally wrong or impermissible. As such, the surgeon possesses no impossible obligation in *Transplant-2* to refrain from trying to kill the pair, and Graham's objection to OIC is nullified. Two questions, however, immediately rear their heads: first, how can we be so sure that the surgeon's killing of the two men would be morally permissible? Second, even if we grant that the surgeon's action would not be morally wrong in *Transplant-2*, what then justifies 1)? Once the janitor's action involves killing a woman who is not technically doing anything wrong, even in the defence of others, we might naturally wonder if that action is as morally clean as has hitherto been assumed.

Taking these questions in order, what makes the surgeon's action morally permissible lies in her inability to avoid performing it. This is not, however, because of anything to do with OIC. Rather, if the forces acting upon the surgeon in *Transplant-2* really are such that she cannot refrain from attempting to kill the two men, then she is not (presently) the kind of being that is capable of holding genuine moral obligations or committing morally wrong actions. Nor may she be an appropriate target for any negative moral attitudes such as blame or resentment as a result of her behaviour. As Vranas says of such individuals "it is plausible to say that during her fits she is in a certain respect akin to a malfunctioning robot: the concept of obligation does not apply to her."[105] In this context, describing the surgeon's action as "morally permissible" may be a somewhat misleading term, insofar as it implies a normal moral functioning on the surgeon's part which she does not possess. Yet if the surgeon's action was not morally wrong, as argued previously, she therefore possesses no genuine yet impossible obligation to avoid performing it and hence there is no threat to my interpretation of OIC.

Now for the trickier question. Suppose we accept the surgeon's behaviour is an action with no moral content due to her complete lack of control over her actions, and therefore is not morally wrong. How can we now maintain that 1) is true - that the janitor was morally correct, and quite possibly morally required, to kill her? To do so is to hold that the janitor was morally required to kill a woman who had done nothing wrong, and that seems ethically repugnant.

I think, however, this is precisely the correct analysis of *Transplant-2*, because your genuine moral obligations are sometimes affected by facts about you, and available to you, at the time at which you act. In the janitor's case, his killing of the surgeon is not morally wrong specifically because he knows what the surgeon intends to do, and that this is the only way to stop it. For him to kill the surgeon unnecessarily, or to do so randomly having no inkling of her plan, would definitely be morally outrageous.[106] However, it would be equally morally wrong for the janitor to allow the surgeon to murder the two men, given what the janitor knows, even though killing her is the only alternative. Thus, the janitor has a genuine obligation to stop the surgeon as a result of his knowledge of the surgeon's intentions, which tragically involves the murder of an innocent woman.

[105] Vranas, Peter B.M. 2007. 'I ought, therefore I can', *Philosophical Studies*, 136, 2: 167-216.
[106] Keep this observation in mind. It will become important in our discussion of a similar case in the following chapter.

None of this, however, changes the lack of wrongness of the action of the surgeon, who is still not an acceptable target of moral appraisal.

We will further discuss this topic of how our genuine moral obligations may be affected by our knowledge (of lack thereof) when I come to defend Premise 2) of the Derivation in the following chapter. For now, I wish to finally move on from the many and fascinating arguments concerning OIC to examine what I call the "connecting premises" of my Derivation. These premises, which supply the link between my version of OIC and the Principle of Alternate Possibilities, are long overdue for a proper analysis and so I will now let the defence of OIC rest in order to take another important step forward in my pursuit of PAP.

Chapter IV - The Connecting Premises

Having now thoroughly explored and replied to several objections against my preferred formulation of OIC, I now wish to examine what I have called the "connecting premises" of my Derivation - those which, once we have granted that "ought implies can", allow us to deduce PAP. Whilst individually these premises are relatively uncontroversial, there is sufficient disagreement regarding them that it would be unwise for me to take them as mere background assumptions in my argument for PAP. It is really the case, we may wonder, that blameworthiness can only arise from actions which are morally wrong, or that we are always genuinely morally obligated to refrain from such actions? In order to properly ground these premises, a detailed defence of each of them will be required. For despite the comparatively little attention I have paid to justifying these premises so far, their truth is no less important than OIC's to the overall success of my project since we will be unable to derive PAP in regard to blameworthiness from OIC if any of them are false.

Here is a reminder of the complete Derivation of PAP, which the connecting premises highlighted for clarity:

1) "Ought implies can": an agent may only be morally obligated to perform (or refrain from performing) an action if they are able to perform (or refrain from performing) that action.

2) If an agent is morally blameworthy for an action, that action is objectively morally wrong.

3) If an action is objectively morally wrong, we are morally obligated not to do it.

4) Thus, an agent is morally blameworthy for an action only if they were able not to do it (from 1 - 3).

5) If an agent who performs an action was able not to do it, they were able to act otherwise than they did (even if this is just to do nothing).

6) Thus, an agent is morally blameworthy for an action only if they could have acted otherwise than they did (from 4, 5).

Conclusion: PAP in regard to blameworthiness is true.

I have already offered my argument for Premise 1) in the preceding two chapters, and Premises 4) and 6) are strictly deductive. Therefore, three premises bridge the gap between OIC as I have defended it and PAP in regard to blameworthiness. Premises 2) and 3) are related claims which serve to tie the concepts of genuine moral obligations, moral wrongdoing and blameworthiness together: if an agent is blameworthy for an action (or for failing to act), that action is morally wrong and they must therefore have had a genuine moral duty to avoid performing it. This broader normative picture has grounded several of the arguments which I have made thus far, yet neither 2) nor 3) are without critics in their own right. In particular, a significant body of opinion has surfaced in recent years that 2) is false - that it is possible for an agent to be blameworthy for an action which is *not* morally wrong. If this is correct, it would peel apart the concepts of genuine moral obligation and blameworthiness that I have laboured to merge (casting significant doubt on 3) in the process), and the Derivation will fail.

Premise 5) may appear initially trivial, as a statement of the seemingly banal point that the ability to avoid performing an action requires the ability to act otherwise, if only in the technical sense of being "able" not to act at all. But this, along with the related claim that an obligation to avoid performing a certain action is therefore equivalent to an obligation to *act otherwise* than that action, have been denied by Yaffe. Yaffe offers an argument for an alternative vision of the morally relevant sense of ability, one that if correct would prevent my Derivation from establishing its desired conclusion. If, however, my connecting premises can be successfully defended against this range of objections we will be left in a strong, if still incomplete position from which to motivate the claims and underlying assumptions of the Derivation. Having sketched the objections to these premises in general terms, then, we shall now examine the specific arguments that have been made against each of them.

Section IV.I: A Distinction with a Difference

Let us begin with the argument against Premise 5). The simplest and most basic starting point for disputing this claim, and indeed scepticism about my chosen sense of "able to" more generally, was put succinctly myself in the following point from Chapter I:

"At first glance, it might seem odd to suggest that PAP is a subsidiary principle of "ought implies can." The Kantian principle is fundamentally concerned with the nature of our moral obligations and PAP, as I want to argue for it, is a principle about blameworthiness."

Whilst my earlier remark may now sound somewhat dismissive in light of the work I have done to establish the relationship between these two principles since, it nevertheless picks out an important point about how we should understand "can" and "able" in regard to each of them. At their heart, OIC and PAP are two distinct, albeit in my view inherently related, principles with two different subject matters. Even I must acknowledge that OIC and PAP are not, for all their similarities, expressions of the *same* moral claim. This is why the Derivation requires connecting premises in order to deduce the latter principle from the former. From this starting point, it would not be altogether surprising if OIC and PAP happened to be concerned with two different senses of ability - that is to say, if there were a particular set of capacities which were the relevant ones when determining whether an agent could fulfil a given moral obligation, and a discernibly different set that was more appropriate to consider when assessing whether that agent was able to act otherwise than they did.

This line of thought may lead us to be suspicious of 5), because for my Derivation to be logically valid requires an additional, unspoken premise: that however one chooses to define the normatively relevant kind of ability that an agent must possess in order to satisfy OIC and PAP, that same sense of ability is the relevant one to apply for both principles. In other words, whatever we think it means that I "can" fulfil my genuine moral obligations for the purposes of OIC, that same meaning is also what we mean when we say I "can" act otherwise than I in fact do (and vice-versa). In order for my Derivation to function as intended, the threshold for an agent to be "able" to fulfil a genuine moral obligation or to be "able" to act otherwise must not be meaningfully different, and our

uses of the term when discussing this threshold in regard to either OIC or PAP must not be referring to distinct powers or capabilities of the agent in question.

It is important to be clear, however, that this unspoken premise does not require that the agent who satisfies OIC in respect to a given moral obligation will always satisfy the conditions of blameworthiness set by PAP (or vice-versa). It is easy to imagine corner cases in which an agent is able to clear their moral duty, in the relevant sense, but is for some reason unable to do anything else. Thus, their obligation satisfies OIC but the agent in question could not be deserving of blame according to PAP.[107] Such cases are insufficient to show that two different senses of morally relevant ability are at work in OIC and PAP. They merely demonstrate that, in my terms, the prerequisites and/or opportunity required to act otherwise may differ from those which are necessary to fulfil the relevant moral obligation(s), and so the two may not always be satisfied, or satisfiable, at the same time.

What has all of this to do with premise 5)? In my Derivation, the purpose of 5) is to serve as the bridge between the ability *not to do* something (which is necessary for holding a genuine obligation to refrain from performing a certain action) and the ability *to do otherwise* than that action (which is necessary to be blameworthy for having performed it). Until now, the assumption that these two abilities were the same thing has been asserted rather than argued for. In light of the above points, however, it is now clear that 5) may reasonably be disputed without having to challenge the particular definition of morally relevant ability which have I set out and defended in the preceding two chapters.

As has been noted variously by Kittle,[108] Franklin,[109] and Schwan,[110] to make the type of move employed in 5) without explicitly calling out what is meant by "able" in each use of the term is rife with the potential for equivocation over differing senses of ability. Franklin in particular has opined that what is often missed in discussions about the relationship between ability and moral responsibility (or blameworthiness) is that there is

[107] Cases in which one may be blameworthy for having fulfilled a genuine moral obligation may be difficult to imagine, but are not incoherent. In the case of competing moral obligations, for instance, one might theoretically choose to satisfy the "wrong" moral obligation (such as choosing to keep a promise to return a borrowed axe) and deserve blame on the basis of their choice.
[108] Kittle, Simon. 2015. 'Abilities to do otherwise', *Philosophical Studies*, 122, 11: 3017-3035.
[109] Franklin, Christopher E. 2015. 'Everyone thinks that an ability to do otherwise is necessary for free will and moral responsibility', *Philosophical Studies*, 172, 8: 2091-2107.
[110] Schwan, Ben. 2018. 'What ability can do', *Philosophical Studies*, 175, 3: 703-723.

not one singular ability to do otherwise but several, as is reflected in our varied linguistic usage of "can" and "able." For example, when we say that Anne can play Beethoven's Moonlight, we are sometimes referring to her physical and mental capabilities - what Vihvelin would call her "skills" - and sometimes talking about what she is capable of doing at this particular moment[111] ("She can play for you if you like, the piano's just over there"). This ambiguity often leads philosophers into accidentally talking past one another by using different senses of "able" interchangeably, without having properly determined what *sort* of ability is the morally relevant one in each case.

The key question for my argument now, therefore, is which of these various abilities to do otherwise are the necessary ones for an agent to be blameworthy for what they do. Specifically, if it should transpire that the sense of "able to do otherwise" which is relevant to PAP is one which is *not* synonymous with the ability to avoid performing a particular action, as Premise 5) holds, then it will fact be meaningfully different from the normatively relevant "able" that is required by OIC.[112] In turn, this would mean that 5) is straightforwardly guilty of exactly the kind of equivocation Franklin describes, and so my Derivation would be clearly invalid.

This is the position taken by Gideon Yaffe,[113] who argues that to have the power to avoid performing a specific action is indeed *not*, in and of itself, to have the power to act otherwise than that action. As a result, both 5) and its related assumption that an obligation to refrain from a certain action is equivalent to an obligation to act otherwise than that action are false, because even granting OIC it would then be possible for us to be able to complete the first of these obligations but not necessarily the second. If these observations are correct, then the Derivation will fail: the most it could conclude would be Premise 4), which if Yaffe is correct is meaningfully weaker than PAP. There are numerous cases in which an agent may refrain from performing a given action by doing nothing at all, but lacks the ability to perform a "replacement" action instead. 4) allows for the possibility that such an agent might be blameworthy for their action, whereas PAP does not.

[111] See Vihvelin (2000, 2013).
[112] It would also mean that my own preferred interpretation of that "able" as presented in Chapter II and III would be obviously false, at least in regard to PAP, but that is not particularly relevant here.
[113] Yaffe, Gideon. 1999. ''Ought' implies 'can' and the principle of alternate possibilities', *Analysis*, 59, 3: 218–22.

Assuming Yaffe's distinction between the ability to *refrain*[114] and the ability to *do otherwise* is correct, I would reject that verdict of blameworthiness. I agree that an agent whose only choices are literally between performing a certain action and not performing it may be blameworthy for that choice, but this is only because I view the refusal to act as a morally relevant alternative in its own right, no different in this sense from the power to undertake another action instead. If we are not to necessarily count this as such an alternative, as Yaffe proposes, then I believe it would be unfair to blame such an agent for exactly the same reasons I have previously given for why it would be unfair to blame an agent who possessed absolutely no relevant alternatives to their action (the ability to refrain from it included). Hence, 4) alone is an insufficient conclusion for the argument I am attempting to set out in this project, and so Yaffe's argument against 5) requires a response.

To support Yaffe's distinction between the ability to refrain and the ability to act otherwise, he offers the following example. Suppose I am in a *Sticky Situation* where I have only two available options: I can perform a morally forbidden action, thus violating my genuine obligation not to do so, or I can do nothing. Since all my obligation requires is that I avoid acting in a certain way, doing nothing will be sufficient to fulfil it and so OIC does not rule out such an obligation. But the Derivation assumes that there is a stronger obligation in this case, one I have previously been assuming is functionally identical to the obligation in *Sticky Situation* as written - to act otherwise than the forbidden action. Yaffe argues that this stronger obligation would *not* be genuine for the agent in *Sticky Situation,* because to choose to do nothing is not to act otherwise but rather the simple absence of any action whatsoever. In *Sticky Situation,* thus, there is no other action I could have performed instead of my forbidden one, and - according to OIC - I therefore cannot possess an obligation to do otherwise. Despite this, I do remain under the genuine obligation to refrain from the forbidden action, and hence we can see that the two obligations are distinct. Moreover, since OIC rules out one of these obligations but not the other, the abilities required to complete them must logically be different in some way as well.

For Yaffe this case demonstrates that contrary to 5) the ability to not perform an action and the ability to act otherwise than it can potentially come apart, as may an obligation to

[114] To "refrain", here and after, is merely the ability to not perform the action in question- it does not require that the ability be consciously exercised.

refrain and an obligation to act otherwise. As a result, he concludes that my Derivation is founded on a faulty assumption about the nature of action, and fails as a proof of PAP. Yaffe defends this distinction by appealing to an interesting belief about moral obligation in general: "morality doesn't require that I consider, reject and take steps to prevent the occurrence of all the things I am obligated not to do…it's good enough to simply not know about them and not do anything to support them." By contrast, if I have an obligation to act otherwise in some way and I simply do nothing, Yaffe argues that I have *not* fulfilled this obligation and should be judged accordingly: "if I did nothing at all…I surely didn't fulfil any moral obligation to act in particular ways."[115]

This view of action is intuitive, after a fashion. For whether a moral agent can act otherwise than they in fact do certainly seems to be a matter of their potential courses of action at a particular moment in time, and so it sounds correct that the sense of "able" which should be relevant to PAP is one that is concerned with the *actions* that agent could undertake. And it is often true that this same "able" is the one which is relevant to OIC, as I have argued in Chapter II and Chapter III: whether an agent is "able" to fulfil a purported moral obligation is frequently a matter of whether or not they can perform some action which satisfies that obligation at the time in question. Consider the obligations to pick up my grandmother from the hospital this evening or rescue my drowning child from the lake. However, one might object that this is not necessarily the case, as when it comes to what we might call "negative" moral duties - obligations where we are required not to do something, as in *Sticky Situation* - such obligations can often be satisfied without taking any action whatsoever.

For example, take what I presume is my highly genuine moral obligation not to murder my housemates. Sometimes - when they are playing music very loudly at 3am, say - it will take a supreme act of will for me to abide by this obligation, and it would be correct to say I chose to act in a way which did. Yet there are also clearly many points at which I am abiding by my obligation without this being the case, such as when I am sleeping or my mind is fully occupied with writing my thesis. It would be quite natural to argue that my passively being in accordance with my obligation not to kill my housemates in this way surely cannot be classified as an action (or a set of actions) in the way that my choosing to fulfil an obligation like picking up my grandmother or rescuing my child

[115] Yaffe (1999) p. 220.

would be. If this is so, then the sense of "able" which is relevant to OIC is not inherently the same as the one which matters to PAP, since to satisfy the latter principle's requirement that the agent is able to act otherwise (in Yaffe's sense) necessarily requires an action to be able to be performed, and this is clearly stronger than the "ability" to merely refrain from acting which is sufficient to satisfy OIC in certain cases.

As Kittle[116] puts it, what matters at this point is how to determine what kinds of abilities we should be interested in. In this specific case, which of these different abilities to do otherwise - the power to refrain from X, and the power to act otherwise than X in Yaffe's sense - ought we to regard as the relevant one to apply when discussing PAP? As Schwan once remarked: "there is no consensus on the right way to analyse ability...different meta-ethical accounts take different senses of ability to constrain obligation." The issue, he perceives, is that "so long as each party to a metaethical dispute has an internally consistent story to tell about the normatively relevant sense of ability, substantive ability claims will have trouble gaining traction absent some explanation of why *that* sense of ability is the normatively relevant one."[117]

The true problem with 5), then, goes beyond the specific challenge raised by Yaffe's account of action. What we have here are two perfectly reasonable accounts of the relevant sense in which an agent must be "able" to act otherwise before they deserve blameworthiness for their actions - the ability to refrain, and what we might call the ability to do differently. No doubt, given time, we could think of more. What we must have in order to defend 5), therefore, is not simply a rebuttal to Yaffe's rival conception of the ability to do otherwise, but an independent reason to understand the sense of "able to" which is relevant to PAP (and OIC) in the way that my Derivation requires. How might we go about providing such a reason? In this, I am sympathetic to Franklin's view that questions about the nature of moral responsibility and blameworthiness are conceptually prior to questions about which abilities are required for an agent to be responsible or blameworthy.[118] If, then, there is something about the character of blameworthiness and its relationship to our genuine moral obligations that points to the ability to avoid performing an action as the morally relevant ability to do otherwise, then we will have

[116] Kittle (2015) p. 3034.
[117] Schwan (2018) pp. 718-19.
[118] Franklin (2015) p. 2101-02.

correspondingly strong grounds on which to reject Yaffe's argument and accept Premise 5).

Before we investigate that question, however, I would first draw attention to a particular facet of Yaffe's argument. It is worth keeping in mind that when we ask whether the power to refrain from an action is equivalent to the power to act otherwise than it, we do so within a specific kind of moral context. Specifically, we wish to know whether the acting agent possessed alternatives, in the morally relevant sense of the term, to the action which they performed. By denying that the power to refrain is a facet of the more general ability to do otherwise, then, Yaffe must also be denying that inaction *can* be a relevant alternative for the purposes of determining our genuine obligations and/or blameworthiness. Indeed, this would be consistent with his earlier claim that to do nothing would be insufficient for fulfilling an obligation to act in a particular way, whether in the earlier case of the sleeping man who does not murder his housemates or more generally.

If we understand Yaffe correctly in this regard, then I believe we now run into a competing intuition to those which support his view of what counts as an action - one which, in my opinion, is more strongly held than those on which his argument leans. It seems intuitively plausible that we may sometimes fulfil our obligations passively, without necessarily trying to do so or indeed without *doing* (in Yaffe's sense) anything at all. As Yaffe himself said above, it is not as we fulfil each of our moral duties by considering and rejecting all of the possibilities opportunities to fail in them. I can, and usually do, fulfil them by just not acting in certain ways, usually without even thinking about it. This by his own admission is sufficient for us to be able to fulfil our genuine moral obligations, so it is unclear why it would not also be sufficient for grounding our blameworthiness given that (as I shall defend later in this chapter), blameworthiness inherently arises from failing in moral duties of this kind.

Furthermore, there seem to be clear examples in which the power to simply do nothing constitutes both an action in its own right and a morally relevant alternative - one whose existence justifies quite logical judgements of responsibility and blameworthiness, which would otherwise be inappropriate. Imagine I am a principled man who believes in a genuine obligation to do otherwise than vote for a corrupt candidate in the upcoming

election.[119] However, each candidate happens to be sufficiently corrupt that supporting them would violate my obligation. To spoil my ballot is punishable by the death of my family - clearly, also a violation of my moral obligations - and is therefore not an option. Call this case *Corruption.*

Here is, I think, the central question. Am I capable of meeting my moral obligations in this situation? Can I *act otherwise,* in the relevant sense, than vote for a corrupt candidate? I think most certainly I can. I can choose to do nothing and stay home rather than go and vote. Is this alternative morally relevant in determining whether I can fulfil my obligation, and to the question of whether I deserve blame if I do not? Absolutely. If I decided to violate my obligation and vote for one of the candidates, it would be preposterous for me to argue I am not blameworthy for that decision because I could not have done otherwise - as if I had been frogmarched to the polling station and forced to cast a ballot. By the same token, it would be equally asinine for me to invoke OIC as a means of defending myself, as I am quite clearly still able to fulfil my obligation not to vote for a corrupt candidate and so that obligation remains genuine. Since what holds for *Corruption* also seems to hold (by extension) to any such case where one's inaction is an available and morally relevant option, we now have a large subset of examples in which the ability to refrain is seemingly sufficient to grant the ability to act otherwise required by PAP.

Unfortunately, there is now a problem. At this point, we might reasonably conclude that there are some cases in which the power to refrain and the power to act otherwise are as one (as in *Corruption*) and others where they are not (as in *Sticky Situation*), with our obligations to refrain and to act otherwise behaving in much the same way. That, however, would still mean that my Derivation would fail as a proof of PAP, since it requires 5) as a truth about all cases of the ability to not perform an action in order to successfully deduce PAP from OIC. Merely highlighting that 5) is true in many, but not all situations is therefore not enough.

What we must do, as I mentioned before, is show that the considerations at work in *Corruption* can be applied universally - that what makes it unacceptable for to say I could not have done otherwise there, where I retained the ability to not perform the relevant action, is not some quirk about that specific case but an observation about morally

[119] For simplicity's sake, leave aside the possibility to that vote is a moral obligation in and of itself.

relevant ability and blameworthiness in general. We would also need to show that this extends to those cases in which refraining from a certain action is not a conscious decision, as it is in *Corruption*, but the seeming absence of any action. To demonstrate this conclusively would not only show Yaffe's distinction to be meaningless and destroy his argument against 5), but would also provide the independent reason to support the ability to refrain from (that is, to avoid performing) a given action as the morally relevant one to PAP which we have been seeking.

To see how this *is* in fact the case, let us consider *Sticky Situation* once more. Yaffe splits apart the ability to avoid performing an action and the ability to act otherwise on the basis that one may sometimes possess the former but not the latter, and so can complete an obligation to do the first but not the second. Yet it is far from clear to me that an agent in the type of situation Yaffe describes in *Sticky Situation* ought to be exempted from the obligation to act otherwise (specifically, as opposed to one to simply refrain) on the grounds of inability to fulfil it. For although it is true that Yaffe's agent cannot act otherwise than his forbidden action, strictly speaking - no other actions are open for him to perform - it is not true that he cannot *do* otherwise than perform it. He may do otherwise, in a morally relevant way, than perform the action which is forbidden by doing nothing, and his level of blameworthiness will change depending on whether he does so. The same, of course, holds for any situation of this sort regardless of whether the inaction is a deliberate choice or a passive occurrence.

This insight about the nature of blameworthiness leads me to conclude that the kind of ability which is required to possess an obligation under my OIC, and to be blameworthy under the PAP that derives from it, is the latter and *not* the former. As Alvarez says of PAP, "I cannot see any reason to interpret the Principle thus [in the former sense], nor do I think that this is how the Principle has been traditionally interpreted."[120] Copp also observes that although there is no indisputable argument in favour of there being only one sense of ability involved in OIC and PAP, both principles are motivated by the same beliefs about the role of fairness in determining our moral obligations and blameworthiness, and the relevance of both concepts to our process of moral decision-making -"the Maxim and PAP have similar intuitive grounding in arguments about fairness and decision-making, arguments that invoke the familiar intuitive sense of an

[120] Alvarez, Maria. 2009. 'Actions, thought-experiments and the 'Principle of alternate possibilities'', *Australasian Journal of Philosophy*, 87, 1: 61-81 (p. 63).

ability to do something. This fact strongly suggests that the doctrines are both concerned with this sense of the ability to act."[121]

Crucially, therefore, to be able to act otherwise in the sense Yaffe describes is not a necessary condition for possessing an obligation to act otherwise in the morally relevant sense of "act otherwise." Hence, OIC does not defeat an obligation to that effect even if the obligated agent can only refrain from (that is, avoid performing) a given action and cannot perform another action instead. As such, Yaffe's case and others like it fail as counterexamples to 5) because they fail to show that an agent that can only refrain from acting is a certain way is incapable of acting otherwise than it in the sense which is relevant to PAP. Meanwhile, cases like *Corruption* show that an agent who can refrain from a morally unacceptable action, but can perform no other action instead, does seem to be able to fulfil an obligation to act otherwise in the morally relevant sense of ability with which I believe both OIC and PAP are concerned.

This gives us a strong case in favour of the idea that to avoid performing a certain action *is* to act otherwise than it for the purposes of determining our genuine moral obligations and blameworthiness, as I have understood these concepts throughout this thesis. Therefore, I believe we should accept 5) even if Yaffe is right that an agent who is able to refrain from a certain course of action may not necessarily be able to act otherwise in the stronger sense of being able to perform a different action instead.

Section IV.II: No Good Deed Goes Unblamed

Two premises of the Derivation remain, and it is best to tackle them in concert. The belief that if an agent is blameworthy for a particular action, that action was morally wrong (Premise 2) and that, given it is morally wrong, that agent therefore had a genuine moral obligation not to perform it (Premise 3) are technically distinct statements and one may accept either or both. When taken together, however, they form a claim about our moral obligations and blameworthiness which is opposed by a sizeable number of philosophers,

[121] Copp (2006). "The Maxim" is Copp's term for OIC.

including Capes,[122] Haji,[123] and Zimmerman.[124] If this claim is false, as this group of objectors argue it is, the Derivation of PAP would be unsound. Before we examine the specific counterexamples to 2) and 3), however, it will be helpful to first set out the problem in the abstract.

Many of the objections to these two premises are based on the concept of "suberogatory" actions - actions which, whilst not objectively morally wrong, are still what Capes calls "morally untoward." Whilst we hold no genuine obligation to avoid performing such actions, they nevertheless possess qualities which prompt justified moral suspicion of their performers. A morally suberogatory action in the game of football, for example, would be one clearly within the rules of the game but not the common social norms associated with it.[125] The question that is of interest to my work here is this: could we be blameworthy for performing a suberogatory action? If so, Premise 2) and by extension 4) and 6) of the Derivation are false. Not only that, but - as we shall see later - if this notion of a suberogatory action is a substantive one we shall also have reason to doubt Premise 3), albeit on very different grounds to the attack on Premise 2) we shall discuss imminently.

Before tackling this argument, I should clarify that I do not wish to deny the existence of suberogatory actions. I am happy to accept that some actions are not objectively morally wrong but are sufficiently close to morally wrong so as to prompt justified discomfort. All I must deny to protect 2) is that these actions are close *enough* to being morally wrong to make their agents blameworthy for having performed them. I therefore wish to examine some potential examples of suberogatory yet blameworthy actions which have been put forward by my opponents, and show that they fail to establish beyond reasonable doubt that the agents in them are blameworthy as a result of acting in these ways. Having done this, I will then discuss how a different understanding of suberogation might allow these cases threaten 3) instead of 2), and respond to an argument to that effect.

[122] Capes, Justin A. 2012. 'Blameworthiness without wrongdoing', *Pacific Philosophical Quarterly*, 93, 3: 417-437.
[123] Haji, Ishtiyaque. 1998. *Moral Appraisability; Puzzles, Proposals, and Perplexities* (New York: Oxford University Press).
[124] Zimmerman, M. J. 1997. 'A Plea for Accuses', *American Philosophical Quarterly*, 34, 2: 229-243.
[125] Such as not kicking the ball out of play if an opposing player is injured, for example.

Let us begin with Haji's case of *Dr Deadly* and the failed assassination. In brief: Deadly despises an ill patient and plots to murder him with an overdose of morphine. However, a sharp-eyed nurse has discovered Deadly's plan and switches the syringes at the last moment; thus, what Deadly believes to be a lethal injection is now precisely the treatment the patient requires, and they make a full recovery. It is inarguably that Deadly's action is morally suspicious, and yet it is not morally wrong. A doctor has a genuine obligation to cure their patients if they can,[126] and Deadly has fulfilled that obligation, albeit in a highly unusual manner. Having done so, her action cannot be morally wrong since it is not possible for the same action to be both morally required and forbidden at the same time. But is Deadly *blameworthy* for her action? Haji argues she is, on the grounds of attempted murder and a gross breach of doctor-patient ethics. Though Deadly technically fulfilled her obligation, her motive and the intended goal of her action was appalling and thus she warrants blame on that basis.

This idea that blame may be a deserved response to an action performed for impure reasons is appealing, but I must reject it. *Dr Deadly* is not a situation in which an agent is blameworthy for a morally permissible action, nor one where they had no obligation to avoid acting in a certain way. To be sure Deadly is blameworthy for something here, but it is not the thing Haji identifies. The act of injecting the patient with the cure, being morally obligatory, cannot be morally wrong and so Deadly is not blameworthy for that. She does, however, deserve blame for *trying* to murder her patient, and the poor effectiveness of the weapon is of no consequence. For doctors also possess a genuine moral obligation not to try to kill their patients, and Deadly is in clear violation of that obligation and thus acts wrongly in doing so. Her attempt on her patient's life, therefore, is both a morally unacceptable action and one for which she is blameworthy, no different from the average act of evil, and hence fails as a counterexample to 2).

To his credit, Haji offers a response to this sort of move. He claims that Deadly's attempt to kill her patient is morally indistinguishable from the actual act of injection. In other words, Deadly's attempt is *also* not morally wrong, carries no associated obligation to avoid performing it, and is still (decidedly) morally fishy for the same reasons as the actual action of injection is. Now suppose we grant, as my previous point does, that the

[126] Importantly, Haji argues- and I agree- that her obligation is stronger than one merely to try and cure patients, or to act in a way as they think will cure the patient. When a doctor errs in an operation and the patient dies, for instance, they have failed in their obligations to them and may deserve blame accordingly.

actual act of injection is not morally wrong - it is, in fact, morally obligatory given it will save the patient's life. It is also a commonly accepted deontic principle that if I am obligated to A, and doing A requires B, then I am also obligated to do B. I have relied on this principle myself in our discussions of an obligation to fulfil as many of my duties as possible in Chapter III.[127] In Deadly's case, to fulfil an obligation to inject the patient will obviously require her to attempt to inject the patient, which by the same logic makes the latter morally obligatory for her as well. But my last response held that the attempt to inject the patient, being attempted murder, was the morally forbidden action for which Deadly is blameworthy. How can this be the case if Deadly is in fact obligated to try and inject her patient, and the same action cannot be both morally wrong and genuinely obligatory?

This move is clever, but rests on a false equivocation between being obligated to attempt to inject the patient, which I agree Deadly is, and being obliged to attempt to *murder* the patient, which she certainly is not. Haji's mistake is to identify the (failed) attempt to murder the patient with the (successful) attempt to inject them with the needle. Deadly's obligation is to save her patient's life, which she fulfils unknowingly through the act of injection. This obligation requires her to attempt to inject the patient with the cure, but it certainly does not require her to attempt to kill them. In *Dr Deadly*, the attempt to inject them happens to also *involve* an attempt to kill them, because Deadly believes the injection is lethal. Yet the murderous trappings of the case are clearly not necessary for our doctor to meet her obligations, as she could have just given up on her plan to kill the patient and given them the cure normally. As such, we have little cause to abandon our initial response that Deadly's attempt to kill her patient is both morally wrong and an act for which she deserves blame, and so *Dr Deadly* fails as a counterexample to the view that blameworthiness only results from a morally wrong action.

A second example of a suberogatory action for which its agent is blameworthy comes from McKenna.[128] Imagine that I could do a great deal of good by acting in a way which comes at no cost to me and requires very little effort. That action, however, is not morally required of me and thus to refrain from doing it would not be morally wrong. McKenna argues that if I chose not to perform such an action despite it being so easy and painless,

[127] That is: if I am obligated to fulfil as many of my moral duties as I can (A), and doing that requires fulfilling each of my individual moral duties (B), then I am obligated to fulfil each of those duties.
[128] McKenna, Michael. 2012. *Conversation & Responsibility* (New York: Oxford University Press).

I might then be blameworthy for that refusal even if my motives were not particularly evil or selfish. Bill Gates refusing to give £10 to a homeless man would be one such example of this: having given millions to charity already, he might justifiably regard his obligation to the poor as satisfied even though the £10 would be meaningless to him. Yet to refuse so little a sum having already given so much would naturally raise suspicions about Gates' motives, which could in certain cases justify his blameworthiness in the same way as Deadly's reasons for acting did for her.

Unlike in *Dr Deadly*, however, to suggest there is any blameworthiness on Gates' part here goes too far. As Nelkin[129] observes, upon greater examination of such cases it becomes increasingly difficult to defend both the view that the agent is blameworthy for their refusal to act in a certain way *and* that the action in question is merely praiseworthy rather than morally required. In this example, we are inclined to say Gates has no obligation to give £10 because he has already met his obligations to the poor. But I think we are also inclined not to blame Gates for his refusal for precisely that same reason. Even though it would cost him nothing and do much good, he has already done far more than his share of this particular moral task and is permitted to stop without deserving blame for doing so. We can also imagine potential cases of this sort where the lines blur in the other direction: suppose yet another child is drowning in a lake and I, a capable swimmer, could easily rescue them. Certainly, it would be very morally suspicious if I did not do this and many would think I am blameworthy if I refused to do so. But this feeling is sourced, I think, in the natural intuition that I *am* obligated to rescue the child, and so the situation is not really a case of suberogation at all.

Essentially, the problem with this set of cases is that as each supposedly suberogatory-yet-blameworthy action tends closer to the morally obligatory, it becomes increasingly less plausible that its performer would not be acting wrong by refusing to do it. In turn, the more morally acceptable the refusal appears, the greater the difficulty in claiming the agent who *does* refuse is actually blameworthy for doing so, as in the Gates case. McKenna's argument depends on a "sweet spot" at which the lack of a relevant obligation and sufficient proximity to blameworthiness are perfectly balanced, but I struggle to find an example of this kind in which both the lack of a moral obligation and presence of blameworthiness are unambiguously clear.

[129]Nelkin, Dana Kay. 2011. *Making Sense of Freedom and Responsibility* (New York: Oxford University Press).

The final and most dangerous argument for the existence of blameworthy, yet suberogatory actions which I will consider comes from Justin Capes. Like his predecessors, Capes rejects my Derivation's 2) and therefore the broader claim that blameworthiness only follows from failing in your genuine moral obligations. In his mind, the deciding factor of blameworthiness is not a failure in a genuine obligation or even just having behaved morally wrongly, but rather what he calls a "morally objectionable quality of will" on the part of the acting agent. In other words, an action which comes about through an agent's morally unjustified mental states, beliefs, and motivations is an action for which that agent can be justifiably deemed blameworthy, even if that action is not objectively wrong and they violate no obligation by performing it. Thus, 2) is false.

If this logic holds, then Haji's *Dr Deadly* will be an example of a suberogatory and blameworthy action after all. For although Deadly's curing of the patient is morally required of her and hence not morally wrong, it is clearly motivated by her hatred for her patient and is therefore an act for which she deserves blame. This account of blameworthiness also accounts cleanly for the intuition that we do not deserve blame for suberogatory acts performed with a noble motive, such as blackmailing a Nazi to ensure the safety of Jews, because it is the quality of the will which motivated the action that matters to determining blameworthiness and not the action itself.

Before we analyse this new proposal, we shall look at Capes' own example of a suberogatory-yet-blameworthy action:

"Beatrix freely shot and killed Bill. She did this despite believing that it was objectively wrong to kill Bill, that it was within her power to avoid killing him and, indeed, that it was within her power to avoid wrongdoing altogether. Why, then, did Beatrix kill Bill? A significant role was played by her hatred of Bill and her (no doubt morally unjustified) desire to rid the world of him…unbeknownst to Beatrix, however, Bill was just about to torture and kill her daughter, and the only way she could have prevented him from doing this was to shoot and kill him. Call this case 'Kill Bill."[130]

[130] Capes (2012) p.18.

Kill Bill, and Capes' account of blameworthiness more generally, is a trickier case to answer than the previous efforts. Like Deadly, Beatrix seems to be acting out of an undeniably objectionable will toward Bill, and her action invokes the moral suspicion typical of suberogation. But Capes argues we might justifiably think Beatrix's action is also morally permissible, considering Bill's intentions and appealing to common principles concerning self-defence and the defence of innocents.[131] The result is a morally acceptable, albeit suspicious action for which Beatrix deserves blame. Unlike in *Dr Deadly*, however, we cannot escape this problem by suggesting Beatrix is only truly blameworthy for something which *is* morally wrong (attempting, successfully, to murder Bill). For if we grant that common principles of self-defence and defence of the innocent exempt Beatrix from wrongdoing for the killing of Bill, it seems logical that those principles would also clear her of wrongdoing in regard to the attempt to kill him as well. Haji's earlier point about an obligation to A entailing an obligation to do what is necessary for A seems relevant here: to shoot and kill Bill, Beatrix must obviously attempt to shoot Bill. If the former is morally justified, therefore, it seems the latter should also be.

Capes next considers various possible responses that would exempt Beatrix from blameworthiness for killing Bill, but we need not consider those here. For I wish to take the other track and deny that Beatrix's shooting of Bill is morally permissible in the first place. I dispute none of Capes' observations about Beatrix's blameworthiness for the shooting - we agree about the malicious nature of her will, the facts of the case, and so on. But my explanation of Beatrix's blameworthiness is simpler than Capes'. She is blameworthy because shooting Bill dead, and the attempt to shoot him which it entailed, was a morally wrong action in *Kill Bill* even taking into account its happy consequences. Capes' stated justification for the permissibility of Beatrix's action (the use of principles governing the defence of others) strikes me as deeply unconvincing, and I can think of no other principle which would satisfactorily suffice for this purpose.

My issue with Capes' defence is that the principles he appeals to justify Bill's death simply do not operate in the manner suggested. If I gun down a random passer-by because of a paranoid belief that they will kill me when my back is turned, and it later turns out that they really *were* planning to kill me when my back was turned, I cannot retroactively

[131] Importantly, Beatrix has no obligation to shoot Bill. It is merely morally acceptable for her to do so. This is the difference between Kill Bill and Dr Deadly, where Deadly does have an obligation to inject the patient with the cure. Her obligation comes from the duty of care she has to her patient, whereas Beatrix has no such duty to Bill.

invoke the principle of self-defence to justify my paranoia. Nor does this become any more acceptable if the perceived target is not myself but someone close to me, or the motive is not paranoia but loathing. This matters because *Kill Bill* is clear that Beatrix did not know, or even suspect, her child was in any danger when she shot Bill. As such, the danger to her child did not and could not have played a part in the motivations which led Beatrix to kill him. Were we to discover later what Bill had intended, perhaps this would make it inappropriate to actually blame Beatrix for his murder. Capes, however, has firmly stressed that such considerations should not change our view that she is blameworthy for the action, given all the relevant facts about her will. In this, I quite agree with him. I merely go one stage further and argue that those relevant facts, both about Beatrix and the circumstances of the situation as a whole, make her action morally wrong *as well as* one for which she is blameworthy.

These considerations indicate to me that Beatrix is in violation of a genuine obligation not to murder Bill, one which persists even accounting for Bill's future intentions and the principles of self-defence and defending the innocents which Capes cites. Hence, I see no reason why we ought to regard *Kill Bill* as an example of an action that is not morally wrong, but for which its agent is blameworthy regardless. Indeed, I think I have now shown beyond reasonable doubt that the types of cases we have discussed here do not serve as compelling counterexamples to Premise 2) of the Derivation in general, and/or to the wider view that blameworthiness occurs as a result of failing in one's genuine moral obligations.

The challenges to the Derivation as a whole from these cases do not end here, however. Until now, I have assessed these suberogatory actions as direct counterexamples to 2). There is, however, a way of using these counterexamples to attack Premise 3) of the Derivation as well, and I wish to consider it briefly in the interests of completeness. Until now, suberogatory actions have been classified as actions which are not morally wrong, but are nevertheless morally suspicious or questionable in some important way. It is this latter quality which is said by critics of 2) to be sufficient for blameworthiness, a position which I have comprehensively rebutted.

What if, however, there were another way of classifying suberogatory actions? What if, rather than being actions which are morally permissible yet blameworthy, we presented them as actions which are objectively morally wrong, but which we hold no genuine

obligation to avoid? Such an approach would give up the attack on 2) in favour of denying 3), which states that every morally wrong action has a corresponding moral obligation not to perform it. The advantage of such a strategy is that it arguably offers a more compelling picture of the "suberogatory-yet-blameworthy" actions than in their original formulations, one which is not vulnerable to the objections I have previously raised. It also provides a better explanation for precisely *why* such actions appear so morally suspicious or untoward, despite not being morally wrong - because on this revised definition they *are* morally wrong.

Consider again *Dr Deadly:* is Deadly's stabbing[132] the patient with a cure she thinks will kill them morally wrong? Yes, because attempted murder is wrong, and it violates a doctor's various obligations to their patient. Does Deadly have an obligation not to stab the patient with the cure? No, because injecting the cure *fulfils* Deadly's obligation to treat her patient, and we cannot be both obligated and obligated not to perform the same action at the same time. Bill Gates refusing to give £10 to the homeless man is wrong, because it would cost him so little and help them so much. Yet it also violates no obligation of his because Gates has already discharged his obligation to the poor, possibly for life, by giving so much to them already. In *Kill Bill*, it is wrong for Beatrix to shoot Bill for all the reasons I have previously highlighted, but unlike in regard to the other (actually) innocent bystanders she holds no duty to avoid shooting him. This is because, like Deadly, to do so is necessary to discharge an obligation she *does* have, in her case to keep her child safe. So it goes, in one way or another, for many if not all legitimate cases of suberogation we might suggest, and so what failed to disprove 2) succeeds as an attack on 3).

Without delving again into each of the different cases of suberogation we have considered, there is a simple response to such a proposal in general terms. On what basis, exactly, does an objectively morally wrong action not generate an obligation not to perform it? It is arguably conceptually necessary that morally wrong actions carry such obligations - that "morally wrong" *means* "you ought not to do it", in the sense of being morally required to avoid performing them. Even without going that far, I believe this objection mischaracterises the nature of moral wrongdoing in an important way: without a corresponding obligation to avoid performing a morally wrong action, it is unclear on

[132] For clarity, I mean the actual injection of the patient here as opposed to the attempt to murder them in the more abstract sense.

what basis we should view that action as morally wrong as opposed to morally neutral or even praiseworthy. If Deadly, for instance, has done nothing she was morally required to avoid doing, and succeeded in the obligation she did possess (by unknowingly curing the patient), it seems unclear *what exactly* is supposed to be wrong about her action at all.

To use intuitions of blameworthiness as a guide to moral wrongdoing in this case would put the cart before the horse: before we can judge blameworthiness, it is first necessary to determine what constitutes wrongdoing because the former is grounded in the latter (Premise 2). On the Derivation's model, the answer to this question is clear - an action is morally wrong if it violates a genuine moral obligation (Premise 3), a rationale that is unavailable to this account of suberogatory actions by definition. Without such an explanation of the moral wrongness of these actions it seems we have little reason to accept that they constitute a counterexample to 3), and therefore the Derivation as a whole stands.

Before we conclude, there is a final point about the nature of our moral obligations which I would like to highlight. What an agent is genuinely morally obligated to do, or to refrain from doing, may depend significantly on one's knowledge and/or evidential situation at the time of acting. This is because the evidence available to us may make it morally unacceptable (or permissible) for us to perform certain courses of action under the circumstances, and as 3) states, if an action is morally wrong we have a genuine obligation not to perform it. When Beatrix shot Bill in *Kill Bill*, she had no idea of what Bill intended, and I suggest was therefore genuinely obligated not to kill him in the same way as she was genuinely obligated not to kill any other innocent bystander. Having violated this obligation without any justification, she has acted wrongly and becomes blameworthy for her action, again as if she had killed anyone else.

How could it possibly be otherwise? Do we say that given Bill's horrible plan for her daughter, Beatrix has a genuine obligation to kill him, and so it is morally permissible for her to do so? Even when we take into account that she does not know, and seemingly cannot know, what Bill intends at the time she kills him - *and* given everything about her selfish reasons for killing Bill which the case has specified? Surely not. If this makes Beatrix's action morally permissible, then it seems to also justify the shooting of any other random civilian in Bill's position, and that would clearly justify too much. Beatrix

has no more evidence to suspect them of imminent evil than she did Bill, and just as much reason against killing them.

This point highlights an important moral difference between the case of *Kill Bill* and that of *Transplant-2* in the preceding chapter. One might wonder why my model of moral obligation deems the janitor to have a genuine obligation to stop the murderous surgeon, whilst Beatrix has a duty to not kill Bill. The answer lies in my analysis of what was wrong with Graham's challenge to OIC in Chapter III. Unlike the janitor, who knows the surgeon's plan, Beatrix was entirely ignorant of Bill's intentions. When she killed Bill, Beatrix had no reason to believe he needed to die, and for her not to do so would have been perfectly morally acceptable (even though it would have led to her child's torture and death) and in violation of no obligation.

The opposite is true in *Transplant-2*. Unlike Beatrix the janitor knows what the surgeon is about to do, and therefore possesses substantive reasons to prevent that action by killing her. It is this knowledge, and the reasons to act it provides, which creates the janitor's obligation to act which is otherwise absent. For it is his knowledge of the surgeon's intentions which makes it objectively wrong for him to *not* intervene to stop the surgeon, which it would clearly not be if the janitor were ignorant of her. But if an action is morally wrong we have an obligation not to perform it, as 3) holds, and so given that the janitor *does* know what the surgeon intends, he now has a moral obligation to stop her - which can only be fulfilled in *Transplant-2* by shooting her dead. By the same logic, *if* Beatrix had somehow known what Bill intended she would have been morally obligated to prevent him from hurting her daughter, which *Kill Bill* also specifies would require killing him. Our discussion of suberogatory actions leaves us, then, with an observation about moral duty in general: that what one is genuinely morally obligated to do is often determined by facts about (and available to) the acting agent, rather than facts about others or the external world. *Kill Bill* is a clear example of this phenomenon, where the wrongness of Beatrix's action does not turn on anything about *Bill* but rather on the matter of her feelings toward and (lack of) knowledge concerning him.

In this chapter, I have sought to defend the individual premises which, in conjunction with my form of OIC, allow us to derive PAP. As with the previous chapter, I do not claim to have dealt with every possible charge against my position here. Rather, my answers are intended to demonstrate how my account of moral obligation, wrongdoing

and blameworthiness would answer a diverse array of views which oppose my Derivation's connecting premises. Where possible, I have answered my objectors in ways which are applicable to a much wider range of potential criticisms than I have had the opportunity to answer directly. If I have succeeded, it should now be clear what role each premise of the Derivation is supposed to play in both arguing for PAP and the wider set of moral principles which I have defended throughout this work.

With this in mind, in the final chapter of this thesis I will turn my attention away from the formal premises of the Derivation itself to argue for one final and critical point which I have offered in support of it: the somewhat radical position that a genuine moral obligation is intrinsically *fair* to the agent that holds it. This claim played a large role in my motivation and defence of OIC in Chapters II and III, and it is only once this idea, too, has been properly defended that the Derivation can truly stand as a sound, well-grounded proof of PAP in its own right.

It is to this last task, therefore, that I shall now devote my energies.

Chapter V - A Question of Fairness

The question of whether an agent's moral obligations must be fair to them, in the relevant sense, may initially seem of little relevance to PAP. Even if we were to grant the link between moral obligation and blameworthiness which I have advocated in previous chapters, nothing about such a connection presupposes anything about the *fairness* of those obligations in and of itself. Nor is such an appeal to moral fairness strictly necessary to defend OIC, whether understood in broad terms or the specific formulation of the principle which I have put forward in this thesis.

Yet the goal of this work is not simply the Derivation of PAP from a certain understanding of OIC, challenging enough though that task has proven to be. I have also been seeking to place the Principle of Alternate Possibilities, and the Derivation which supports it, as the logical conclusion of a certain set of moral principles. One of these principles - hitherto undefended - is the claim that all genuine moral obligations must be fair to the agents that possess them. This "Fairness Principle" is not an essential part of the formal Derivation, but I have appealed to it throughout this work as a foundation for various arguments that I have offered in favour of the Derivation's premises. In many respects, then, the intellectual foundation of this project as a whole depends on the plausibility of this principle. In this final chapter, therefore, I now wish to examine whether it truly stands up to scrutiny, and defend the idea that genuine moral obligations are inherently fair in the same way as I have defended the individual premises of the Derivation itself.

My argument for this position consists of multiple steps. First, I will clarify exactly what I mean when I claim a genuine moral obligation must be fair to its agent, as well as those properties which are necessary for this quality (Sections I-II). Next, I explain why I believe those same conditions are also necessary for a moral obligation to be genuine (Section III). In Section IV, I then present an argument that the qualities that are required for both a fair obligation and a genuine moral obligation are also *sufficient* to constitute a fair moral obligation. Thus, we may conclude that any genuine moral obligation will also be fair, in the sense relevant to us, by default. This done, I devote Sections V, VI and VIII to combating three distinct challenges to the Fairness Principle and the necessity and sufficiency of the specific properties I have championed: what I refer to as the problem of Hindsight, the problem of Helplessness, and the problem of Ascription. In addition to these three challenges, I diverge from my overall argument in this chapter in Section VII

to resolve an as yet unresolved question concerning the impact of so-called "tragic" moral dilemmas on my account of a fair moral obligation.

Whilst this is not an exhaustive analysis of the potential challenges to my position, I hope my responses to these problems will serve to establish a solid foundation for the claim that genuine moral obligations are intrinsically fair in a meaningful sense of the term - and provide firm ground upon which to base my previous arguments for the wider Derivation by doing so.

Section V.I: The Nature of a Fair Obligation

Let us begin by answering the most basic question of this discussion: when I say a genuine moral obligation must be "fair" to its agent, what am I referring to?

The first thing to understand is that I am only interested in the definition of a fair moral obligation, and not a full and complete analysis of fairness in general. Whilst the two lines of inquiry will undoubtedly overlap, the scope of this chapter only extends to the necessary and sufficient properties of an *obligation* that is fair to its agent(s). What this means in practice is that I am only exploring the contours of one respectable sense of fairness, not conducting an exhaustive analysis of each and every way in which we might legitimately apply the concept. As such, although I believe these properties are also necessary to all cases of genuine moral obligations (hence my conclusion that such duties are inherently fair), there may be other viable understandings of fairness under which a genuine moral obligation does *not* have to be fair to its agent in the manner I suggest.[133]

Secondly, my investigation in this chapter deals with a sense of fairness that differs somewhat from our everyday use of the term. Ordinarily, when speaking of fairness or unfairness we are usually referring to a quality possessed by agents and/or by an agent's behaviour. For example, we may describe a man's treatment of his employees as fair or a parent's favouritism for one of their children as unfair. It may seem quite strange, therefore, to talk as I do about "fair obligations" as if fairness were a property of those obligations in and of themselves. Such an approach might seem to anthropomorphise morality, which might make sense in a more theistic context but fits very uneasily into a

[133] This will become apparent in Section VIII, where we discuss the Fairness Principle in the context of so-called "distributive" senses of unfairness.

secular work like this one. This is the main charge of the Problem of Ascription, which holds that fairness is simply the wrong concept to apply in discussions of this sort. Whilst I will cover this objection properly in Section VIII, I suspect there will be enough natural scepticism about the concept of a "fair" moral obligation in the first place that we should begin with a deeper explanation of such a duty, before we begin to discuss its precise components.

It is true that most non-philosophical talk about fairness centres on an agent and/or their conduct in a given situation. It is also accurate that fairness understood in those terms makes little sense to apply to one's moral obligations. It would clearly be inaccurate to say that just as my employer acts unfairly in setting me extremely difficult or arbitrary tasks, morality or moral duty may "act" unfairly by assigning me certain moral requirements. Morality is not a thing which acts, in any sense of the term, and so fairness conceived strictly as a quality of agents or their actions is obviously inapplicable to moral duty.

Importantly, however, I believe such an "action-focused" conception of fairness is not the only legitimate way in which a thing may be fair or unfair. Consider that fairness is often a quality not of agents or individual actions but of *circumstances* or *states of affairs.* Certain events are unfair, in a meaningful sense of the term, not in virtue of any agent's actions or motives but as a by-product of the way particular systems or structures are arranged. In most societies, for example, there are those suffer unjustly simply because of how their institutions function - the innocent person who falls afoul of poor laws, the deserving poor who cannot get support, the qualified man who cannot find a job - but this clearly does not require there to be an *agent* or *action* somewhere that has mistreated them. Perhaps the support services for the poor have been cut due to needed austerity, or the unfairness of the legal system is an unintended consequence of badly drafted laws. We should let go, then, of the intuitive but I think mistaken idea that fairness is solely a concept that relates to how agents treat one another - at least, for the moment.

It is in one of these broader senses of fairness that I believe our genuine moral obligations must be necessarily fair to their agents. For fairness is also often a matter of *desert.* When an agent is made to face unreasonable challenges through no fault of their own, or suffers excessive punishment or consequences for their actions, it is both natural and right to decry this as unfair. It is unfair in this sense of the term because they do not deserve to be

in the circumstances that they are in, even though no other agent may be responsible for what has happened to them. Having said that, this sense of unfairness also commonly applies to cases where agents act unfairly toward each other - the reason their behaviour is unfair because the wronged party does not deserve such treatment. Imagine I am suddenly threatened with losing my job unless I can flip ten heads in a row with a coin. Even if I am skilled enough in the art of coin-flipping that I might be able to do this, our instinct is that the task is obviously unfair because I have done nothing wrong, and therefore certainly do not deserve to have my job risked in such an outrageous fashion.

Note, however, that when we say the requirement to flip ten heads in a row is unfair we are not simply making a claim about the unreasonableness of my employer. That begins in the wrong place. Rather, I think we are saying that the requirement is itself unfair, and *therefore* my employer is wrong to ask it of me. This in turn explains neatly why I would not deserve any negative consequences as a result of failing in it. In this specific regard, I think there is a parallel between non-moral cases like this one and unfair moral obligations. I have argued before that genuine moral obligations set the baseline for what constitutes morally acceptable behaviour, and so an agent's failure to fulfil them is itself a morally wrong act. There may be justifications for their failure, of course, or excuses that exempt them from any negative moral judgement. But as discussed in Chapter II, failure in a genuine obligation negatively affects my moral status even if no other agent would think it appropriate to blame or judge me for that failure because my actions did fall short of what was morally required. This moral "demerit", as I have previously called it, is justified on the basis that the obligated agent was responsible for their failure, in the relevant sense, and so *deserves* the change in their moral status which follows.

These considerations prompt the following base for a fair obligation in the sense we are concerned with: a "fair" obligation is one that can be fulfilled with the right level of moral awareness, and the proper effort, on its agent's part. If this is not the case - if an agent who knows what they must do and makes their best effort toward it still cannot succeed - the obligation is unfair, because it places the agent in a position they do not deserve to be in. Another way of putting this thought is that with a fair obligation, an agent's success or failure in it will be tied in the appropriate ways to their own skills, knowledge, and decisions. In other words, whether they will meet their moral duty will depend in the relevant ways "on them." This ensures they will deserve whatever moral credit they are entitled to for completing that obligation, or any blameworthiness or other negative

changes in their moral status which attach to them for failing in it. An obligation is unfair when this is not the case - when it is *not* up to the obligated agent whether they will succeed or fail, and therefore do not deserve the resulting change in their moral status (be it for good or ill).

The claim that all genuine moral obligations must be fair, then, is essentially to say that their agents' success or failure in those obligations is always sufficiently in their own hands.

Section V.II: The Properties of a Fair Obligation

This is a good starting point for our understanding of fairness, but we must now be more precise. In virtue of which properties, specifically, are moral obligations fair or unfair to their agents? Here matters become more difficult. Fairness is an inherently nebulous concept in any philosophical discussion - though no less important for that - and as stated it is beyond the scope of this work to offer an exhaustive account of it in all its forms.

That being said, it is possible to develop the sense of fairness I believe is intrinsic to genuine moral duty in a more indirect sense. For the presence of *un*fairness is often a far easier concept to recognise and categorise, both as it pertains to our moral obligations and elsewhere, than its opposite. I think we can identify several qualities which would each, if possessed, be sufficient to render an obligation unfair to its agent(s) in the sense we have been developing thus far. If this is the case, as I now hope to demonstrate, we could then work backward and establish the necessary conditions for a fair obligation accordingly. Naturally, it may be that there are other qualities sufficient for unfairness than those I will highlight here, and therefore more necessary conditions for a fair obligation than I will suggest. This poses no threat to my argument, as I do not require these qualities to be the *only* necessary conditions of fairness for my claims about the nature of genuine moral obligations to hold, and to support the arguments I have made elsewhere in this thesis.

Not coincidentally, the qualities that I will soon identify as sufficient to make an obligation unfair to its agent in the sense I am invested in are the same whether that obligation is a specifically moral one or an everyday non-moral duty (such as expectations between friends, work-related duties, and so on). This is, of course, not to say that *every*

matter of fairness is alike between the moral and the non-moral, simply that what makes an obligation fair or unfair to its agent is in this sense the same regardless of the type of obligation that it is.[134]

The essence of the sense of fairness we are working with is that an agent's success or failure in their obligation should depend primarily on them, rather than factors outside their control. Therefore, there are at least three properties which, if absent, would be clearly sufficient to make an obligation unfair to its agent. In turn, a fair obligation in the sense I argue is intrinsic to all genuine moral duties must therefore possess at least three necessary conditions, each corresponding to one of the sufficient conditions for unfairness. These are the conditions of *Achievability*, *Discernibility,* and *Equitability*. The absence of any one of these properties is sufficient for an obligation to be unfair in this sense, even if the other two are met: for instance, if an obligation is achievable but not discernible by its agent, it is unfair to them. By the same reasoning, *if* it should transpire that there are further necessary conditions for a fair obligation than the three I identify, the absence of any of *those* properties would also be sufficient to render an obligation unfair in this way even if my stated conditions were also satisfied. This means that whether a moral obligation is fair (and, if my wider argument is correct, genuine) may vary from moment to moment, and will depend significantly on the current circumstances of the agent that holds it.

At this point, two logical questions emerge - precisely what do these necessary conditions entail, and why should we think they are needed for an obligation to be fair to its agent? I shall endeavour to answer those questions now.

To say a fair obligation must be necessarily achievable, first, is to say that the obligated agent(s) must be able to complete it at the time(s) they are required to do so, in the sense of "able" I have advanced and defended in previous chapters. This means they must possess the Prerequisites for the obligatory course of action, as well as the opportunity to use them at the required time.[135] This condition does not require that success in one's obligations be guaranteed, nor that it is even easy - after all, an obligation is not unfair merely because we might fail in it. Notably, it also does not necessarily require that we

[134] I am unsure of the precise scope of this claim. It may be that the same conditions I set out in this chapter extend to other sorts of genuine obligation than those specified (e.g. our epistemic or legal duties). Alas, this is a question for another project.
[135] See Chapter II, Section III for more on this definition.

can fulfil the obligation in question in concert with all (or any) of our other obligations.[136] It might be the case that fulfilling a particularly important obligation requires all my attention and effort, thus causing me to fail in other less important duties. If I must miss my promised dinner date with you because I need to rush my dying grandmother to the hospital, that evidently does not make my obligation to save her somehow unfair to me.

Whilst the above may appear as a simple restatement of OIC, this is somewhat misleading. Since I have previously justified OIC as I understand it by appeal to my Fairness Principle, to now do the reverse would be viciously circular. Instead, the relationship I propose is more nuanced. Here achievability is one of several necessary conditions required for an obligation (genuine or not) to be fair to its agent, and the requirement that specifically genuine obligations *must* be fair is one piece of supporting evidence for the principles of OIC and PAP. Whilst I do believe a genuine moral obligation must ultimately be both fair to its agent and achievable, an obligation being genuine is not necessarily the same as it being fair to its agent. There are many potential obligations one could think of which would be fair in the sense I am interested in, but are nevertheless *not* genuine - a purported obligation to give 99.9% of my income to charity might pass the tests I have set out in certain circumstances, but I would clearly have no such obligation in virtue of those facts alone. Thus, fairness is only a necessary condition for genuine moral duty and not a sufficient one.

But why, exactly, does a fair obligation require achievability? We have already begun to answer this question in our earlier discussion of desert. If, as I have argued, an action which violates a genuine moral duty is (all other things being equal) morally wrong, it is not difficult to see how an unachievable obligation would be unfair. If there is an intrinsic link between failing in one's obligations and having acted wrongly by doing so, any agent who possessed an unachievable duty would face what I have called an "unavoidable moral demerit," whereby they will inevitably fail in what is required of them regardless of what they do, and their moral record will suffer accordingly. Whilst certain agents in this position might still deserve this demerit,[137] most would not - assuming the agent in question was not responsible for their obligation being impossible, it is evidently not "up to them" in the relevant sense whether they will succeed or fail in it because their failure

[136] We will return to this point in Section VII, where it will form part of my explanation for why I believe cases of tragic dilemma do not involve unfair obligations.
[137] I am thinking here of the lazy, perhaps, or the incompetent- those who would have failed in the duty whether it was possible for them to achieve it or not.

is guaranteed. If an event would occur regardless of what we choose to do, we cannot be at fault for the fact that it happens, only the particular time and circumstances in which it does so. As such, it is apparent why an unachievable moral obligation would almost always be unfair to its agent in the sense I am outlining, for their inability to affect the outcome of their unachievable obligations means they will never deserve the moral demerit which inevitably follows.

Whilst the achievability condition is perhaps the most basic requirement of a fair obligation, it is by no means the only relevant one. We will now speak, therefore, of the condition of discernibility. To be able to "discern" an obligation for my purposes, first and foremost, is to be capable of identifying what that obligation requires of us as well as the particular time (if any)[138] that we are required to act. The specific knowledge required to meet this condition will therefore vary according to the precise nature and complexity of the obligation in question. However, it seems clear as a starting point that an obligation which its agent never has the chance to discern is unfair to them, in the sense relevant to us. From this we can deduce that a fair obligation must be discernible by its agent at some point before the obligation is to be discharged, or else they would have no power over whether they succeeded or failed in it (of which more later). The *exact* point at or before which the obligation must be discernible is a much more ambiguous matter, but I am mercifully able to leave that particular controversy to one side. My argument here is only concerned with the unfairness of indiscernible duties - how we define the precise boundaries of those duties is a question I am content to remain neutral on.

I should also emphasise that the requirement of discernibility only requires that the obligated agent has the capacity to discern the details of their obligation, and not that they actually do so. There are many normal cases in which we fail to discern our obligations due to ignorance or forgetfulness or a myriad of other factors, most commonly our mistaken moral beliefs and/or lack of care in thinking through the relevant issues. In those instances, it is right that our obligations stand as normal - as noted with the achievability condition, an obligation is not unfair to its agent merely because they fall short of it, so long as the failure is properly theirs. The unfairness comes specifically from the obligated agent's inability to discern their duty at a time when that knowledge may aid them in completing it (that is, a time before that obligation is to be exercised).

[138] Some moral obligations, of course, do not need to be exercised at a particular time, like the general moral duties of kindness or charity.

To understand the necessity of discernibility to a fair obligation as I understand it, assume for a moment that it is not necessary. If that were the case, it seems that an agent might naturally find themselves with a perfectly achievable obligation of which they were not aware and could not, for whatever reason, discern. Maybe the required information is no longer available, like an unseen wedding invitation eaten by their dog, or the agent is so lacking in moral sense that despite their best efforts they simply cannot work out what are they required to do. Whilst such a duty violates neither OIC nor our initial condition that a fair obligation should be achievable, the only way in which such an agent could succeed in their obligation would be to "get lucky" - perhaps they go for a walk and miraculously end up at the wedding at the right time, for example. Such cases may or may not be examples of genuine obligations (I suggest not, but that argument will come later) but they are certainly not examples of *fair* obligations, in the sense of interest to me. Like the earlier requirement to flip ten heads in a row, the success or failure of these agents is determined by essentially random guesswork. The outcome, then, doubtlessly does not turn on anything within their control in a morally meaningful sense, and so they do not deserve to have their moral record suffer for it. The possibility of fair yet indiscernible duties can be ruled out accordingly.

The final property I have identified as integral to a fair obligation is that of "equitability." I choose this term, which refers to impartiality and even-handedness, to pick out an important dimension of fairness which we have not yet accounted for. Our previous necessary conditions focused on where an agent stands in relation to the tasks set by his duty. Does he possess the right skills and opportunities? Has he access to the correct knowledge? Although these are undoubtedly core components of my sense of fairness, I believe it should take account of more than merely an individual agent and their duty. Fairness must also concern the distribution of the moral burden: how an agent's obligations compare to those of others in the same (or similar) circumstances as them. Just as two people doing the same job ought to possess the same level of responsibility in their work, and the fair law of the land should apply equally to its citizens, so too should agents in the same *moral* circumstances possess the same moral obligations. Hence, the requirement of equitability holds that a fair duty should be consistent and impartial: that anyone in the same position as an agent under that obligation, accounting for all of the morally relevant factors, would possess that same obligation and for the same reasons.

Any obligation which fails this test may be safely regarded as unfair, as we would intuitively do, on the grounds that it is fundamentally arbitrary.

Once again, we can see how this condition is necessary to a fair obligation as I seek to understand it by simply imagining a definition without it. Even putting aside the question of whether an unequitable obligation could be genuine, if such obligations could be fair we would be left defending a patently absurd state of affairs. Picture a straightforward case of moral duty where two parents are caring for their severely ill child. Normally we would think that, all other things being equal, each parent possesses the same level of responsibility for their child's health and thus the same obligations to care for them. The fact that *they* possess these obligations to the child, whereas other agents do not, is fair to them because of their special (familial) connection to the child in question - if those other agents had ill children of their own, it would be their duty to care for them in just the same way.

Yet now imagine that, somehow, one of the parents holds stronger or more numerous obligations to their child for no morally relevant reason - it is just a quirk of the system. This seems obviously unfair to the parent that happens to have the greater burden, even if they are both aware of it and capable of discharging it. Having already established there is nothing morally relevant behind the additional duties, we can therefore ascertain that they have done nothing to deserve being in this situation. Furthermore, even if the morally encumbered parent fell short of their arbitrarily greater responsibilities through their own faults - and hence seem to satisfy our core principle that moral success or failure should be one's own doing, rather than the product of external forces - we would *still* have cause to question whether their additional obligations were fair. This, too, is because of their arbitrariness. For once we have determined that there is no relevant reason why one parent should have these extra duties, it no longer seems warranted that their moral record should be downgraded for failing in them, even if that failure was the agent's own doing.

Once again, therefore, we see the concepts of fairness and desert go together. An unequitable obligation, like an unachievable or indiscernible obligation, is unfair because it asks *too much* of its agent - this time in the sense of expecting more of them than of others in the same position, as opposed to expecting more than they can possibly do or expecting them to know things which they cannot be aware of. Intuitively, we push strongly against the possibility of such a scenario. This, we think, is not how actual moral

duties are supposed to function - one parent could not just randomly have more obligations to their child than the other. If there really were such a difference in their genuine duties, the thinking goes, there would *have to be* some relevant moral difference in their circumstances to explain that fact. I agree with this assessment, and so I deny that scenarios like the one I have just presented are possible.

My explanation for this claim is a simple one: the three properties I have highlighted here are not merely necessary for a moral obligation to be fair to its agent, but are also necessary conditions of any genuine moral obligation at all. My next task, now that my chosen understanding of fairness has been properly defined, will be to demonstrate the truth of this point - to at last begin arguing for the claim that our genuine moral obligations are intrinsically fair.

Section V.III: The Necessity of Fairness

Having now categorised what I consider to be the necessary components of a "fair" obligation, the time has come to establish my second claim that all genuine moral duties necessarily possess that same set of properties. By this, I mean that any genuine moral obligation is (at least) achievable, discernible, and equitable, and in a wider sense that success or failure in that obligation should be primarily determined by its agent as opposed to the product of external forces. To be clear, these criteria are to be understood as necessary conditions for a moral duty to be genuine rather than sufficient for it. It would be obviously incorrect to suggest that achievability, discernibility and equitability are the *only* requirements for a course of action to be genuinely obligatory, and what constitutes the remainder of those requirements is a question beyond the scope of this thesis. My intent, for the moment, is merely to highlight the link between a fair moral obligation and a genuine one by showing that the three properties I have highlighted are necessary for both, and to argue that this necessity arises for the same reasons in each case.

To this end, my strategy will be to develop an argument along the same lines as the preceding discussion about the nature of fairness. Just as we were able to discern the necessary qualities of a fair obligation by determining the sufficient conditions for an unfair one, so too can we discern the necessity of certain qualities to a genuine obligation by identifying circumstances that are sufficient for an obligation to not (or no longer) be

genuine. Unsurprisingly, I believe that impossibility, indiscernibility, and lacking equity are examples of both kinds of sufficient condition, and for the same reasons. If this could be shown, as I now intend to do, it would confirm the opposite conditions (of achievability, discernibility and equitability) as necessary for any moral obligation to be genuine as well as to be fair.

Before I explore the specific problems associated with the concept of a genuine obligation which is impossible, indiscernible, or un-equitable, however, I wish to say a little more about an interesting similarity between our moral and non-moral obligations. There appears to be a curious symmetry in how the two kinds of obligation maintain and lose their authority over the agents that hold them. As discussed in Chapter II, an obligation's "authority" in this context refers to the justified pull or force it exerts over its agent to abide by it. With everyday non-moral duties like the legal, social, or work-related, there is an unspoken yet clearly present implication of reasonableness that attaches to them. In other words, although these duties are usually genuine there is an implied understating that they remain so only for as long as they involve reasonable expectations[139] of the obligated agent. If and when this is no longer the case, the genuine duty the agent was under lapses and any authority it had over them is lost. Naturally, it then becomes unreasonable of us to continue to expect that duty to be fulfilled.

I draw attention to this phenomenon because it parallels, to an intriguing extent, the corresponding situation with our genuine moral obligations. Although the circumstances required for a genuine moral duty to lose its authority and lapse might be significantly stronger than "ordinary" unreasonableness, it is apparent that there *are* qualities which are sufficient for this to occur (of which I claim to have identified three already). As with one's non-moral obligations, there are countless everyday cases where moral duties lapse and cease to be genuine under certain conditions - if OIC is correct, for one, when they cannot be completed. And once a moral obligation has lapsed, as with a non-moral duty, it would be wrong for of us to still require the obligated agent to abide by it.

[139] Quite what constitutes "reasonable" in any given circumstance is clearly ill-defined, but there are countless instances of this process in everyday life. The courts are hesitant to press charges against the battered wife who snaps and murders her husband, my obligation to go to work is temporarily waived when a family member is injured, the expectation to attend a dinner party naturally vanishes when I am caught in a blizzard…

These considerations suggest there is a connection of some sort between the authority of an obligation (of any kind) and whether it makes reasonable demands of its agent(s). This is not direct evidence for my position that a genuine moral obligation must be fair in the specific sense I have outlined, but it is a useful starting point. In particular, the sense of "reasonableness" at work here in the cases of non-moral obligations seems to closely resemble the broader sense of fairness with which we began in Section I. For once an agent's circumstances are such that they cannot be reasonably expected to complete their obligation, they no longer deserve to be in a position where they are required to do so or to suffer any consequences (of any sort) for *not* doing so. Furthermore, what determines whether it is reasonable - in other words, whether it is *fair* - to expect them to complete their non-moral obligation is, I suggest, whether their success or failure in that obligation will depend in the right way on their own skills and capabilities. In other words, the expectation is only fair to them if it is "up to them" whether they will succeed at the obligation in question, just as it is with our genuine moral duties.

If we accept the above parallel, we now have the beginnings of an explanation why the same considerations would both make an obligation unfair to its agent and render it non-genuine. The hallmark of a genuine obligation is the authority it wields over its agent(s) - if there were no such obligatory "force" in favour of its completion and the obligated agent could ignore or act upon their duty at their leisure, the action(s) in question would evidently no longer *be* obligatory. We have already touched on this topic in Chapter II concerning the inability of impossible duties to provide compelling reasons for action, but now it will become the centre of our attention. If, as I have suggested, both moral and non-moral obligations sustain and lose their authority in the same way - by placing unfair requirements on their agents, as we have previously defined "unfair" - it will then be apparent why the specific necessary conditions for a fair moral obligation are also required for a moral obligation to be genuine. For if an obligation were to lack any one of these properties, it would be unfair to its agent and its authority would lapse accordingly - and thus, it would no longer be a genuine moral duty.

The core of my argument for why genuine moral obligations must be fair, then, goes like this: we have observed a noticeable parallel concerning *when* and *how* our genuine obligations (moral or otherwise) lapse and lose authority over us. I posit that there is, in the same way, a parallel regarding *why* our obligations might fail to hold power over their agents in this way. They lose their authority because their demands are unfair to the

agent(s) under them, which will occur once the agent lacks sufficient influence over whether they will be able to complete the obligation in question.[140] We have already determined in Section II some of the qualities I believe to be sufficient for this lack of this authority, through our discussion of the properties sufficient for unfairness in my chosen sense. Impossibility, indiscernibility, and the absence of equitability dispel the authority of any obligation, moral or otherwise. Since we know a genuine moral obligation is authoritative by definition, we can therefore conclude that any such duty must, at least, not possess any of those qualities.

From this position, we appear well placed to hold that a genuine obligation is necessarily achievable, discernible, and equitable. One part of my argument thus far, however, requires reinforcement. We may be understandably suspicious of the claim that the three conditions I have identified inherently weaken or destroy a genuine obligation's authority, rather than merely being sufficient to make a genuine obligation unfair. In other words, on what basis should we assume that these conditions are sufficient to render an obligation no longer genuine, as opposed to that obligation being both genuine *and* unfair to us? In what remains of this section, I wish to address this criticism by examining our three criteria once more, and demonstrating precisely why the absence of any or all of these properties appears sufficient for a moral obligation to lose authority over its agent (and thus, cease to be genuine).

At this point, there is little new to be said concerning the necessity of achievability to an obligation's authority that we have not already covered elsewhere. To restate the case for clarity, however: it is an observable fact that an obligation certainly appears to lose any authority over its agent(s) once they know they cannot complete it, insofar as they will no longer feel any requirement or "pull" to perform the impossible task. Any further demands to do what we know to be impossible will naturally be dismissed out of hand. They may certainly acknowledge an obligation to undertake the required action(s) if and when they *become* possible, of course, but that is clearly not to say they *currently* feel any obligation to abide in that particular way, under their current circumstances. Additionally, there is a very common feeling that it is wrong and unfair to be expected to complete an impossible task or suffer consequences, of whatever form, when we fail to

[140] Notably, this is not the only way in which an obligation might lack authority over its agent and therefore not be genuine. Sometimes this occurs because the obligation itself is simply baseless, and so the agent has no relevant reason to abide by it- as is the case with unequitable moral duties. We shall return to this topic shortly.

do so. In the specific case of moral obligations, as I have argued elsewhere, these consequences need not be from other agents: it would be sufficiently unfair for there to be a negative effect on one's moral record as a result of an inevitable failure to do what one ought, even if no other agent would blame them. I have previously described this phenomenon as a "black mark" or "moral demerit."

On my account of obligations and their authority, it is not coincidental that these two reactions should so commonly go together. Genuine moral obligations wield authority over their agents, and impossibility (whether we are aware of it or not) erodes that authority in two important ways. Firstly, as discussed in Chapter II part of the function of a genuine moral duty is to provide distinctly powerful reasons in favour of performing the action(s) it requires. Whilst unachievable duties might still be morally useful and informative, I argued, the fact that we cannot fulfil them means there cannot be this type of unique reason - one which overrides all other sorts of reason set against it - in favour of them. This is crucial to an obligation's authority because a genuine moral duty is supposed to provide a definite answer to the question of what, morally, one ought to do. This is what Schwan has called the "central deliberative question."[141] Without the ability to provide assuredly stronger reasons than those in favour of any non-obligatory actions which may be available, it is no longer possible to provide such a definitive answer to this question - one might well have stronger reasons to do something which is achievable but morally optional than to fulfil an impossible duty. The authority of our hypothetical impossible obligations is therefore undermined, as it is no longer clear that we ought automatically to fulfil them.

Secondly, just as it is unacceptable for a fair moral obligation to land its agent(s) with an unavoidable moral demerit - by setting them a requirement which cannot be fulfilled - I suggest it would be equally unacceptable for a *genuine* moral obligation to do so either. Why might this be?

Recall that a genuine moral obligation sets the bar for what constitutes morally acceptable behaviour, and so actions which violate our duties are morally unacceptable by default.[142]

[141] Schwan, Ben. 2018. 'What ability can do', *Philosophical Studies*, 175, 3: 703-723 (p.719).

[142] This is not to say that failure in one's genuine obligations is always to have done something wrong. There can, of course be legitimate excuses or justifications for having fallen short of one's obligations. However, excuses are only relevant for behaviour which would otherwise be morally wrong. We have an excellent excuse to use if we were to fail in a genuine, yet impossible moral duty- but it is important to recognise that if one had no such excuse, their actions would still be morally wrong.

Setting aside the scenario where their own actions have rendered an obligation impossible to complete, an agent's failure in an unachievable duty will occur regardless of what they choose to do. Hence, they cannot be properly judged as responsible for having fallen short of what is morally required. For their moral record to be blackened anyway, as it would be if the impossible obligation really were genuine, would damage the credibility of that obligation and weaken its authority. This is because it results in an incorrect moral verdict: it judges an agent's failure in an impossible obligation to be a morally wrong action for which they are responsible and at fault, when this is most certainly not the case. But if the agent commits no moral error by failing to act in a certain way (failing to fulfil their obligation, in this case) we are then lead to the natural thought that they hold no genuine moral obligation *to* act in that way, and in turn to the conclusion that an unachievable "obligation" is not a genuine one.

An unachievable moral obligation therefore lacks authority on two fronts. It cannot provide the kind of overwhelming reason in favour of its completion that is intrinsic to a genuine obligation, since any such reasons are diminished if not defeated entirely by the impossibility of the task. More seriously, it would often unjustifiably worsen the moral status of its agent(s) by incorrectly counting them as having acted wrongly when they are not at fault for their failure. These considerations indicate that achievability is a cornerstone of an obligation's credentials, since its absence leads to notable moral errors and impacts the ability of that duty to function as a genuine obligation inherently ought to do. In such a case the authority of the impossible obligation over its agent would therefore be lost, and it would simply be one optional action among many for them to choose from.

The absence of discernibility, too, seems sufficient for a moral obligation to no longer be genuine. My reasoning here, as elsewhere, turns on very similar grounds as those which established that indiscernibility was sufficient for an obligation to be unfair. Earlier I made the point that if legitimate obligations could be indiscernible by their agents, a clear majority of them would fail in those obligations and have acted wrongly by doing so. This would occur because if a moral obligation is truly indiscernible the only means by which its agent may complete it would essentially be through good fortune, such as being in the right place in the right time or choosing the morally obligatory path without realising it.

As Lord has rightly argued,[143] this is not how obligation works: it surely cannot be the case that the *only* way to act as you are required to is through a stroke of luck. In my terms, such an arrangement would be unfair to those agents because they would effectively be sleepwalking into morally unacceptable courses of action, and the resulting negative changes in their moral status, through no fault of their own.

I believe this logic can also be used to show that indiscernible duties are undeserving of being called genuine, because it undermines the authority of these obligations as well as their fairness. As we saw previously with unachievable "obligations", the agents in this kind of situation would not deserve the moral demerit they would receive if the indiscernible obligations they failed in really were genuine. The hidden nature of the obligation ensures that it is not within its agent's control in the morally relevant sense whether they will succeed or fail, and therefore they ought not to be counted as responsible for the outcome of their duty.[144] Once an agent is forced to act in ignorance of their duty, they would seem to deserve neither whatever moral profit they would normally be entitled to from their success, nor any losses they may incur for their failure. Yet once again, if an agent has done nothing wrong by failing to act in a certain way - not merely that they have an excuse for their offence, but have committed no offence that requires excusing - it is logical to deny that those actions were genuinely obligatory for them in those circumstances. The authority of the indiscernible moral obligation is thus called into question, since it demands action(s) which are seemingly not truly morally required.

In other words, if an indiscernible moral obligation could be genuine there would seem to be an error concerning its agent's wrongdoing (or lack thereof): because success in one's genuine obligations is the minimum that is morally required, to fail in such an indiscernible duty would be morally below par and therefore naturally incurs a moral demerit. This account seems obviously incorrect. For without any influence over whether they succeeded in their "duty" the obligated agent appears to commit no offence, and so their failure to do what they could not have known was required ought *not* to be counted

[143]Lord, Errol. 2015. 'Acting for the right reasons, abilities, and obligation', in *Oxford Studies in Metaethics*, 10, ed. by Russ Shafer-Landau (Oxford: Oxford University Press), pp. 26-52.

[144]A clarification: whilst the agent's individual actions are relevantly within their control, the obligation's indiscernibility means that it is no longer "in their hands" whether they will succeed or fail in it. Compare a man betting at a roulette table- which number he bets on is completely his decision, but without knowing which number will eventually come up it would be wrong to say he is in control of whether he wins the bet. The same logic applies here.

as morally unacceptable. Since an indiscernible moral obligation may, thus, be unfulfilled whilst still acting as one morally ought, and without committing any moral error in the process, we may deduce that such “obligations” are clearly not authoritative and therefore that they are not genuine.

We have already discussed the subtle importance of equitability to the fairness of our obligations. To conclude, I will now explore the parallel that I believe exists, as with achievability and discernibility, between the necessity of an even moral burden to a genuine moral obligation and to a fair one. To recap, to say a fair or genuine obligation must be equitable is to say that it would apply to any agent that happened to be in the same or relevantly similar moral position as the one currently under it. This means that anyone in the same circumstances, once all the morally relevant facts have been taken into account, will possess the same genuine moral obligations. An agent in relevantly different moral circumstances would possess different moral obligations accordingly, as we would intuitively expect.

I argued in Section II that if this proved to not be the case, and an agent could magically have more genuine moral obligations than another despite there being nothing morally relevant to distinguish them, those additional obligations would be unfair to the agent in question. The unfairness arises from the fact that they have clearly done nothing to warrant the additional burden, as there is no relevant reason why they have it, nor any moral consequences which may logically result from their failure to discharge it. Crucially, this unfairness would persist even if the agent with the greater obligation were both aware and capable of fulfilling it, as they still would not deserve a moral demerit for failing in it *even if* that failure were their own doing in the way I have suggested typifies a fair and genuine moral obligation.

The reason for this apparent discrepancy is to be found in my suggestion that an unequitable obligation inherently lacks authority over its agent, and therefore cannot be genuine. As with the prior qualities of impossibility and/or indiscernibility, a lack of equity in a moral obligation guts any authority it purports to have over its agent(s). At a very basic level, for a moral obligation to possess authority over its agent requires the actions it commands to actually be morally required of them, and this in turn comes about through specific facts about the world and our relation to it. I am genuinely obligated to have dinner with you become I promised that I would, I have a duty to feed my child

because they are *my* child, and so on and so forth. If all the morally relevant facts about two (or more) agents are identical, therefore, there appears to be no room for additional genuine duties to arise for either of them. In the absence of any such morally relevant reason behind the expectation that one agent should do more than another in a relevantly identical position, there is simply nothing for the inequitable obligation(s) to derive its authority *from*.

If this thinking is correct, any additional genuine moral requirements held by one agent but not another would have to be generated by some change or difference in the relevant moral facts. In the case of inequitable moral obligations, we know such facts to be absent owing to the very nature of such obligations. As a result it seems a genuine moral duty cannot possibly be inequitable, for any authority it wields over its agent would also have to stem from this difference in the important moral considerations…a difference which we have just ruled out. Yet a genuine obligation *must* inherently wield authority over its agent, for it is the inability to be avoided whilst acting "rightly" that marks a course of action as truly obligatory rather than desirable but ultimately supererogatory. It is therefore only right that an obligation which lacks any of the precise conditions of fairness which we have identified should automatically be discounted as a genuine duty, since it has forfeited any authority it might have possessed.

This concludes my argument for the conclusion that achievability, discernibility, and equitability are necessary conditions of both a fair moral obligation and a genuine one. Yet this alone is not enough to establish the claim made by my Fairness Principle, which holds that every genuine moral obligation is also inherently fair to the agent that holds it. To prove that this is the case, it will be required to show that the notions of fairness and true obligation are tied together in a stronger form than I have previously established, such that every genuine moral obligation is guaranteed meet the standard required of a fair duty in my sense. In the following section, I will pursue my chosen strategy for this task: to argue that the conditions we have already shown to be *necessary* to a fair moral obligation are, when taken together, also *sufficient* to make an obligation fair. As we have already determined that any genuine moral obligation must necessarily possess those same conditions, we may then safely conclude that all such obligations are intrinsically fair in the sense we are concerned with.

Section V.IV: From Necessity to Sufficiency

Before I put forth my argument that achievability, discernibility, and equitability are collectively sufficient for a moral obligation to be fair to its agent, I wish to briefly reiterate the background assumptions within which that argument is being made. First, when I refer to a moral obligation being "fair" in this context I mean that it is fair in at least one specific and legitimate sense, and not in every applicable way that we might use the term. Second, the precise understanding of fairness I believe these three properties to be sufficient for is the naturally intuitive one which I have detailed in the initial sections of this chapter. Understood in this way, an obligation is "fair" if success or failure in that obligation is sufficiently within the obligated agent's control (in the relevant respects) that they can rightly be said to deserve whatever moral credit or demerit they would be due as a result.

I clarify this from the beginning because there are philosophically rigorous senses of fairness according to which the conditions of achievability, discernibility and equitability are clearly insufficient for an obligation to be properly fair, and I wish to note that acknowledging their existence has no bearing on my overall argument. The conclusion I seek to defend here is one with a defined and limited scope, and my arguments for the central importance of fairness to our everyday moral thinking will not be structurally sound (nor, I suspect, be particularly persuasive) if one understands "fairness" in a broader and more nebulous sense than I intend it.[145] For instance, when I have previously stated that one of the strongest points in favour of OIC is the intuition that morality must be fair, or that tragic dilemmas are not a counterexample to that intuition (of which more later), these claims are clearly only true within the context of a certain meaning of "fairness." It is that particular meaning which I have been attempting to understand and clarify throughout this chapter, and my observations do not and are not intended to function outside of it.

Here, one might object that I have stacked the philosophical deck. Having narrowed my argument about "ought implies can" and fairness to a single definition (it may be protested) and having gone on to define that sense of fairness in the obscure terms I have, it is entirely unsurprising that the conditions we previously identified as necessary to that

[145] As we often do, for instance, in our ordinary non-philosophical use of the term.

sense of fairness would also end up being sufficient for it. There will undoubtedly be a suspicion that I have *engineered* this chosen definition of fairness for the express purpose of supporting my understanding of OIC and PAP beyond it. If so, the overall conclusion I am seeking in support of my Derivation - that all genuine moral duties are inherently fair to their agents - would turn out to be merely trivially true. Quite apart from the question of whether my three conditions really are collectively sufficient for a fair obligation as I have defined it, therefore, one might question the merits of that definition in its own right.

That thought process, however, inverts the truth of the matter. It is not that I have chosen a particular sense of fairness in order to use it as an artificial foundation for my argument in favour of OIC, and from there my wider Derivation. Rather, my inquiries throughout this work have proceeded in the reverse order. The defence of "ought implies can" I offered in Chapter II and III highlighted the importance of the intuitive sense that morality is fair to the most persuasive arguments in favour of OIC. My attempts to codify that intuitive sense have led us to the broader formulation of a fair obligation which I offered at the beginning of this chapter, and in turn to the specific three qualities which I believe are both individually necessary and jointly sufficient for it. Therefore, it is only natural that my arguments in favour of OIC and the importance of control in determining our moral standing would ultimately come to depend upon the sense of fairness I have been slowly developing here, because it is that sense which strongly inclines us toward the view that OIC and PAP are true in the first place.

On what grounds, then, should we think the properties of achievability, discernibility, and equitability are not merely necessary, but collectively sufficient for a moral obligation to be fair? Well, to make any argument that X is sufficient for Y is to claim "necessarily, if X then Y", but there are several means by which this relation might be achieved. X might be sufficient for Y because X *constitutes* Y, in the way that being an unmarried man is sufficient for being a bachelor. Alternatively, X could be sufficient for Y in the sense of *meeting the threshold* for Y, as with minimum requirements for an objective. A score of ABB in one's A-levels may be sufficient for getting into Sheffield Hallam, in this sense, by satisfying the university's entry conditions. It might also be that X is sufficient for Y because X is a *type* of Y, as is the case with herons being a species of bird.

Of these kinds of sufficient relation, I think the correct one for my purposes here is closest to the last type. It is certainly not that this trio of conditions *constitutes* fairness, as I have already conceded the possibility of other meaningful senses of the term that are not reducible to my three chosen properties. It may be technically correct that achievability, discernibility, and equitability together satisfy the threshold for a fair obligation in the sense we are investigating, but I feel this is slightly misleading. It implies that there is a single threshold of fairness which may be met in a variety of different ways - of which my three conditions are but one - and therefore hints at a single, overarching sense of fairness which I have previously disavowed. Hence, I suspect the best approach will be to categorise achievability, discernibility and equitability as sufficient for fairness because they (taken together) encompass a specific *type* of fairness, just as herons and eagles exemplify specific types of bird. Just as being an eagle or being a heron is sufficient for being a single kind of bird, therefore, so too is an obligation's being achievable, discernible, and equitable sufficient for it to be a one kind of "fair" among many.

Even accepting this, however, and taking it together with my earlier analysis of this sense of fairness in the preceding sections, it is still not obvious that my trinity of properties do guarantee that a moral obligation will be fair to its agent in the sense of interest to us. I will attempt to show now, therefore, that an obligation which possesses these properties grants the obligated agent sufficient power over the outcome of their duty[146] that they will, at least *prima facie*, deserve whatever changes to their moral record may occur should they succeed or fail in it. I will then explain how this level of agency satisfies my requirements for an obligation to be fair to its agent (as discussed in Sections I and II) and how this, in conjunction with the arguments I have already made, yields my desired conclusion that all genuine moral duties are fair to their agents. In the process I will speak a little more of what "sufficient power" might actually look like in real cases of moral obligation, this being an area which has hitherto been somewhat opaque. Once this is done, I will turn my attention to three dedicated challenges which aim to show that even if we accept the legitimacy of my understanding of fairness, it is *not* the case that an obligation which is achievable, discernible, and equitable grants this required level of control over its outcome, and thus my wider claim that all genuine moral obligations are inherently fair in this sense is false.

[146] That is to say, whether or not they will fulfil it at the time required of them.

As we have seen, the sense of fairness which I have been steadily developing begins from the principle that a "fair" obligation involves a connection between the abilities and decisions of the obligated agent and the eventual result of that obligation. This is to say that when a moral obligation is fair to its agent, whether that agent will ultimately succeed or fail in their obligation should depend upon their actions and decisions in a meaningful way. If the outcome of a certain duty is entirely divorced from one's own situation, most likely because you hold no power over whether you will fulfil it, I submit that that obligation is unfair - it would clearly be wrong for you to be held responsible for that outcome or for your moral status to change accordingly, as it would ordinarily do if the obligation were genuine.[147] It is for this reason that a fair obligation in this sense must be not merely achievable, but also discernible - if a moral obligation were to be somehow achievable but indiscernible, as mentioned before, it would be impossible for its agent to knowingly act upon it. Any such connection between their agency and their success or failure in that obligation would therefore be entirely accidental, and so any resulting change to that agent's moral status would be entirely undeserved.

The question before us now, then, is essentially whether an obligation's possessing the properties of achievability, discernibility and equitability is, by itself, sufficient to create a connection of this type. Put another way, if the obligated agent will necessarily have the right level of influence over the outcome of their duty so long as that duty possesses these three properties, then those qualities will indeed be collectively sufficient for a fair obligation in my sense. But what does having this "right level of influence" or "sufficient power" amount to? This language is still too vague to be properly informative. At the same time, there are a great many genuine moral obligations and even greater variety of situations in which they are held. Therefore, it is impossible for me to detail the exact requirements for each obligation in each individual circumstance, and I will not even attempt to do so. Instead, I shall focus my explanation on the specific nature and level of influence which the obligated agent must possess in order to deserve a change in their moral status for failing in a given obligation.

It is my belief that whether an agent is responsible or "at fault" for their success or failure in a given moral duty is determined by whether they possessed the ability to affect the outcome of that duty. Therefore, the primary factor in deciding whether that agent

[147]Note, however, that you may still be responsible for failing to try to complete their moral duty, and deserve a moral demerit on that specific basis.

deserves a change in their moral record as a result of such a failure - and thus, whether their obligation is fair to them - will be whether their actions have *efficacy* in regard to the outcome of that obligation. In simpler terms, they must be able to succeed in their obligation through their own agency, regardless of what they are morally required to do. This is not simply the claim that they must be able to fulfil their obligation, as that would be to simply restate the condition of achievability. Equally, it is not to require that the obligated agent must be able to ensure or guarantee that they will succeed in their obligation, for then no genuine moral obligation could ever be counted as fair. Neither the mere fact of an agent's failure in their obligation, nor the relative difficulty of a given moral duty indicate anything about the *fairness* of that duty in and of themselves. What matters in determining the fairness of an obligation in this sense is the relation its agent's choices and qualities have (or do not have) to the potential outcomes of that obligation.

To understand the importance of efficacy to a fair obligation and what distinguishes it from an agent simply being able to act according to their duty, consider: there may clearly be cases in which an agent can technically "do their duty", but are unable to achieve the required outcome through their own agency. Imagine a hypothetical genuine obligation to win at roulette by betting on exactly 13 black. The average player could theoretically complete this duty by placing the necessary bet and getting extremely lucky, but they would not deserve any moral credit for having done so if they did. This is true for the same reason that, perhaps more obviously, they would certainly not a deserve a moral demerit if they were to fail in it: they lacked the required influence over the outcome of their duty, since success or failure was literally a matter of random chance.[148] Let us call this first case *Layman's Roulette.* Conversely, a man with this obligation who was a professional cheat, capable of rigging the odds of the game in his favour might well deserve a moral credit or demerit for succeeding or failing in such an obligation as long as his cheating contributed to the eventual result (as opposed to it being the product of factors beyond his control, as in *Layman's Roulette).* In such a case, it appears that his obligation would be fair to him. Call this second case *Skilled Roulette.*

What is missing in *Layman's Roulette* and present in *Skilled Roulette* is that the skills and decisions of the professional cheat are relevant to the result of his obligation in a way that

[148] It is for this reason, of course, that on the terms of my argument an obligation such as this would be neither genuine nor, in fact, fair to its agent- even though it satisfies my three conditions of achievability, discernibility, and equitability. This apparent discrepancy will be examined in Section VI.

the layman's are not. In *Skilled Roulette,* the obligation to win on 13 black may be fair to its agent because the capacities of the cheat could, in certain circumstances, make a difference to the outcome of that obligation. His abilities do not guarantee that he will succeed, of course, since he might rig the table incorrectly or be caught whilst doing so. But they enable him to affect the outcome of his duty in a way that is clearly not available to the layman. No matter how diligently you or I may bet on 13 black and attempt to sway things in our favour, the only thing we can do to bring it about that our obligation is completed is to place the bet,[149] and being able to place the bet patently does not provide enough influence over the outcome to justify a moral demerit if we fail. Since the duty is achievable we may ultimately complete it or we may not, but our choices and capabilities can make no difference either way. Such is not the case in *Skilled Roulette,* and it is this *efficacy,* for want of a better word, which justifies the change in his moral status which occurs once his obligation has concluded. The obligation of the cheat is fair to him because it can (though not necessarily will) be completed in virtue of his agency and the other factors over which he has control.

I set the threshold for a fair moral obligation here because, speaking in general terms, an agent that fails in an obligation which was both achievable and discernible and to which they hold efficacy in regard to its outcome will manifestly fall short of what is morally expected of them. If, say, the obligated agent were to discern their moral duty but fail to complete it despite having the opportunity to do so and satisfying the prerequisites in this efficacious way, or failed to discern an obligation which could have been identified if they had acted differently (by giving more thought to the issues, perhaps) it is clear that their actions would be morally lacking. Having been capable of knowing what was required of them and of meeting that requirement, not merely in terms of technical achievability but as a result of their own agency, the obligated agent is - in principle - suitably equipped to influence, albeit not fully control the outcome of their obligation. Hence, if such an agent fails to complete it they will deserve to suffer the moral demerit that would be due.

Section V.V: The Problem of Hindsight

[149]And, most assuredly, if we did not place the bet in such a circumstance we would certainly deserve a moral demerit on that basis, for we certainly possess sufficient efficacy and influence in regard to that.

This having been said, it appears there may be certain circumstances in which the kind of failure just mentioned may occur with no fault on the part of its agent. In such cases, a moral downgrade does not appear to be warranted even though the agent(s) in question seems to have had the "right" kind of influence over the outcome of their obligation. If I am obligated to return ten pounds which I borrowed from you last week, and at the very instant I extend my hand to give the banknote to you the wind intervenes and snatches it away, I self-evidently do not deserve any sort of moral demerit for failing to return your money. This is true even though I failed in a genuine obligation which was achievable, discernible, and equitable and over which I clearly had efficacy - after all, I was just about to pass the money back to you. Interestingly, this kind of accident does not intuitively appear to render the obligation itself unfair *in hindsight*, since it continues to satisfy all the relevant conditions to be fair to its agent, even though I am no more deserving of a moral demerit for my failure to fulfil my duty here than I was in the obviously unfair case of *Layman's Roulette.*

Nor is this the only example of this type. Consider again the case of *Skilled Roulette.* Earlier I drew a contrast between the efficacy possessed by the professional cheat in regard to his obligation to win on exactly 13 black and the relative lack of influence possessed by the agent in *Layman's Roulette.* And it is certainly true that there are several variations of *Skilled Roulette* in which the cheat does possess this greater degree of efficacy and therefore appears deserving of a moral demerit in the cases where he fails in his duty. This, as said before, led us to think the obligation involved in this case is a fair one. The problem arises when one considers that there are also multiple potential variants of *Skilled Roulette* under which the facts of the case remain the same, including the cheat's increased efficacy vis a vis the result of his duty, but in which he intuitively *does not* deserve a moral demerit if he fails. This is despite his superior skills compared to the layman and his obligation still being achievable, discernible, and equitable. What if our professional cheater is picked up by the casino's security before he can rig the roulette table, or is caught in the crossfire of a mob shooting as he walks across the floor? Call these new variations *Unlucky Roulette*.

As with the wind blowing away the money I am due to return, it appears that the cheat's obligation is not retroactively unfair to him just because he happened to be fatally misfortunate, yet at the same time that he evidently deserves no moral demerit for having failed in that obligation in *these ways.* Neither does this conflict with my prior observation

about the original version of *Skilled Roulette*, that if the cheater were to make some error in rigging the game or did not devote sufficient attention to the task (and so on and so forth…) he would certainly deserve a moral demerit in that specific case. It is unclear how my account of a fair obligation deals with this kind of situation, for until now it has depended heavily on the deservedness - or lack thereof - of the eventual demerit that would be received for failure as the litmus test for a fair moral obligation. Now the story becomes more complex, since it appears that a fair obligation may occasionally result in an undeserved demerit - albeit one for which its agent has an ironclad excuse - even though that obligation undeniably has the characteristics I argued are sufficient for providing efficacy.

A different kind of example may serve to illustrate the problem better. Suppose I am in the finals of a fencing tournament, and it is my genuine moral obligation to defeat my opponent - perhaps innocent lives rest upon the outcome, or my opponent is a flagrant cheat who brings dishonour to my sport. Suppose further that my opponent and I are equally matched in terms of skill level and so either of us might realistically triumph. Such an obligation satisfies the conditions of achievability, discernibility, and equitability. Not only can I discern what my obligation is, I also have a chance of completing it - a good one, no less - and anyone who was in the same position that I am would face the same obligation to win. The requirement of efficacy is also satisfied, for my swordsmanship, how I choose to approach the fight and various other skills of mine are obviously relevant in determining the eventual result of the bout. To use the prior vernacular, I can influence but not guarantee the required outcome of my duty, for it is equally likely that my opponent might outfight me or that I might make a mistake “on the day” (and so on and so forth). Each of the conditions I have set forth on a “fair” moral obligation therefore appear to be satisfied. However, now imagine that I were to fail in this kind of moral duty, or for that matter to succeed. Would I truly deserve the change in my moral status that would occur if the duty were genuine?

The answer to this question is not as simple as it may appear. As with *Skilled Roulette* we can certainly imagine versions of this story in which my skills and preparation were insufficient, and I fail in my obligation as a result. What if my opponent was simply the better fighter when the time came, or I made the mistake of indulging in my favourite tipple before the match? Surely then the demerit would be deserved. Yet we have also clearly established that the fact that I lost when I could have won does not justify a moral

demerit in and of itself, or else the layman without roulette skills would also be deserving of one. And just as we saw with *Unlucky Roulette,* it seems easy to imagine cases in which the obligation is identical and yet the manner of the defeat means that a moral demerit is clearly *un*deserved. What if I were forced to forfeit the match because my opponent threatened my family? What if the serial cheat I am pitted against cheats against me as well? It seems that with an obligation of this sort - one which possesses my required three properties and therefore grants efficacy, but where the outcome is partially but not completely within one's control - the negative change in moral status will only be deserved when my failure is of a certain sort. I will only deserve the demerit, it seems, if it occurred because *I* was lacking in my skills, decisions or efforts, because of *my* shortcomings in some way.

These anomalous cases present my account with a significant problem. I have argued throughout this chapter that the defining test of whether a moral obligation is fair, in my terms, is whether an agent deserves the moral demerit that would occur if they were to fail in it. The highlighting of the necessity of achievability, discernibility, and equitability and the importance of efficacy have essentially been a protracted attempt at establishing the powers the obligated agent requires for this to be the case, and the various qualities a fair obligation must possess to ensure those powers are present. It now appears, however, that whether an agent deserves a moral demerit for failing in a given obligation has nothing to do with the *properties* of that obligation, and everything with to do with the circumstances of their failure - whether the failure in this particular instance was in fact the result of their shortcomings in some way, or was essentially forced on them as the product of external factors they could not influence. Worse, this looks to be the case regardless of the presence or absence of efficacy in that agent's actions regarding the result of their obligation, making it at best a necessary rather than sufficient condition for fairness as I understand it.

This indicates that whether a given moral obligation is truly fair to its agent, in the sense of interest to us, could only ever be determined in hindsight - that is, after the outcome of the obligation has been decided and the precise circumstances of the agent's success or failure are established. It is therefore difficult to see how a "fair" obligation with the character I have argued for could possibly ensure that its agent would deserve a moral demerit for failing in it, when whether that demerit is warranted depends entirely on the actual manner and circumstances of that failure. This "hindsight problem" is not merely

epistemic (that is, that it would be impossible for us to *know* the fairness of the obligation in advance) but also a metaphysical problem. It implies that before the agent's failure there is no fact of the matter about whether or not the obligation is fair to its agent, since the moral demerit they would receive for failure cannot yet be correctly said to be either deserved or undeserved.

A final example will highlight the lunacy of this position: imagine I am offered a contract for teaching work in the coming term, and I wish to know - quite reasonably - whether what is asked of me is fair before I accept the moral obligations involved. If whether a moral obligation is fair to me is a matter of whether I would deserve the moral debit I incur for failing in it, and whether I deserve this kind of moral debt for failure is a matter of the circumstances of that specific failure, then it looks as though the correct answer to the question "is this a fair contract?" according to my account, will be "We cannot say. Best of luck with your new duties, and if you fail - *then* we'll be able to tell you." Yet it is surely inaccurate to claim that the legitimacy and authority of any given obligation can only be determined retroactively, as is seemingly the consequence of holding (as I do) that 1) the fairness of an obligation is determined by whether any change in moral status that would accompany its resolution is deserved, 2) all genuine moral obligations are necessarily fair in this sense, and 3) whether a change in moral status of this kind is deserved is dependent on the reasons why the obligated agent succeeded or failed in that obligation.

If this "hindsight problem" were a sound criticism, it would prove fatal to my account of fairness. Not only would it undermine all of my arguments in previous chapters about the role that I believe fairness plays in determining the nature and scope of our moral duties, it would also destroy the wider foundation that my account is intended to provide, through my Fairness Principle, to both OIC and my wider Derivation of PAP. It is therefore very good that it is not, in fact, a sound criticism. This objection mischaracterises both how my account would understand the seemingly problematic examples we have discussed, as well as the "blind spot" which they supposedly create. I shall now explain why this is the case and offer a clarification of how to correctly interpret such cases correctly according to the principles I have laid down thus far.

To begin, I would note again in simple terms the definition of a fair moral obligation I have offered: an obligation that is fair to its agent is one in which that agent will deserve

whatever moral demerits would naturally be incurred if they fail to complete it. The previous statement of the problem interprets this to mean that a moral duty can therefore only be fair once there is a fact of the matter about whether the demerit which would be incurred is deserved, which (so it is claimed) could only be determined in hindsight, once the circumstances of the failure in question are known. This is an intuitively tempting reading of my position, but it is inaccurate. For there is a fact of the matter about whether the moral demerit for failure would be deserved at an earlier point than this - the various points at which the obligated agent possesses, but has not yet "resolved" their obligation. This is true even though the specific circumstances of that failure are, obviously, not yet established. It is therefore entirely possible on my account for an obligation to be genuinely fair or unfair to its agent before they ultimately succeed or fail in it, and so the problem of hindsight is a phantom.

To see how this can be the case, consider once more the *Unlucky Roulette:* our professional cheater, being under the obligation to win on exactly 13 black, enters the casino fully intending to rig the table and fulfil his duty. However, he is mistakenly and unceremoniously gunned down by a drunken mob boss before he can complete his task, and dies with his obligation unfulfilled. The proponent of the hindsight problem suggests that on my account this ought to be an unfair obligation, since our cheat clearly does not deserve a moral demerit on the basis of the mob boss' poor aim and the obligation was "fairness-neutral" before that point.

The issue here is fundamentally one of tenses. For at the point that our cheat acquires his obligation to win on 13 black - call this T1 - provided his obligation is achievable, discernible, and equitable, and that his actions have efficacy in regard to its outcome (as all parties agree they do in the cases of *Skilled Roulette* and *Unlucky Roulette*), he is suitably equipped that it is correct to say that at T1, *prima facie*, he will deserve a moral demerit if he fails in his obligation.[150] As such, it is also correct to say that at T1 the obligation to win on 13 black is *fair* to our cheater, in the sense we are interested in. This is because the conditions for the obligation to be fair to its agent at that point in time have been satisfied at that point in time, since there is a fact of the matter at that moment about

[150]This logic also applies in reverse: if, at this point in time, the obligation to win lacked the conditions of ADE and/or did not grant sufficient efficacy in regard to the outcome, as it fails to do in the case of Layman's Roulette, then it would correct to say at T1 that the demerit would not be deserved and so the obligation is not fair at that time.

whether his duty possesses the required trio of properties and whether he presently has efficacy regarding it.

But what then is occurring at the moment that our cheat fails in his obligation, or the wind whisks away the money to be returned? How can it be at the same time that the obligation remains fair, and was previously so, if it issues its agent with a moral demerit that is so obviously undeserved at the time it is received? This question, too, mischaracterises what is taking place in such cases. Since the obligation involved here is genuine and fair to its agent, a moral demerit would normally be due (and deserved) if the obligated agent were to fail to complete it. However, if the agent's failure is of the kind suggested, where it is clearly no fault of theirs and occurs due to forces completely beyond their control, any demerit that they would normally suffer will be excused on that basis. This does not render the obligation they are under retroactively unfair, nor imply that it was not previously genuine before the specific circumstances which excuse their failure occurred. It is not, then, that these cases form a special exception to the "rules" of how obligation and fault are supposed to work on my account of fairness, merely a recognition that there are circumstances in which we do not accumulate the moral debit we would normally be due.

In light of this discussion, I shall now offer a more precise definition of my account of a "fair" moral obligation in the sense I am interested in:

A moral obligation is fair to the agent(s) under it *if* they will deserve any moral demerit which they would normally receive as a result of failing in it, *unless* that failure is excused by specific external circumstances and/or qualities of the failure in question - circumstances and qualities which were outside of that agent's control at the relevant time(s).

To parse this more easily, it may help it consider it in light of an earlier example. Is the moral obligation to return the ten pounds I have borrowed from you fair to me, even though the wind may snatch away the money at the last moment and prevent me from doing so? Both intuitively and on this account, the answer is yes - ordinarily, as part of the "natural" course of events, I would clearly be at fault if I could return your money and did not, and so deserve the demerit would be due for failing in my moral obligation. In this specific case however, my failure is simply *excused*, as is the demerit I would

normally be subject to, because the failure is the direct result of external circumstances that I had no power over at the time (the force of the wind, it blowing at that particular moment and so on). The same verdict applies, by extension, to the various other cases of this sort that we discussed earlier.

Section V.VI: The Problem of Helplessness

Now that the problem of hindsight has been clarified, we can return to my argument that the conditions of achievability, discernibility, and equitability are not merely necessary, but also *sufficient* for an obligation to be fair to its agent(s). For quite apart from the subset of cases we have just discussed, there is another, perhaps more basic problem with my account that must be addressed. I argued earlier that whether or not an obligation is fair is determined by whether the obligated agent(s) can meaningfully influence the outcome of that duty through their own skills, decisions and actions (for if they cannot, they will not deserve a moral demerit for that failure and the obligation will not be fair). But is it *truly* the case that a moral obligation which is achievable, discernible and equitable necessarily imbues its agent's actions with this required degree of efficacy? This now appears to be the critical question, for if this trio of properties should fail to grant enough efficacy in even a single case of genuine moral duty, the conclusion that they are collectively sufficient for a moral obligation to be fair to its agent would be false according to my own logic.

The initial prognosis for this claim does not look positive. Even when we restrict ourselves to just the kinds of hypothetical cases which I myself have brought up whilst categorising a "fair" moral obligation, there appears to be at least one clear counterexample to this proposal. Earlier we discussed the case of *Layman's Roulette*, where a man with no particular gambling ability is morally obligated to win on exactly 13 black. When I originally discussed this example, I noted that such an obligation could not be counted as fair to its agent under those conditions and thus - if the wider argument of this chapter is correct - could not be genuine either. I offered this verdict for reasons which should by now be very familiar: that the titular layman does not possess any kind of efficacy with regard to the outcome of his duty, having no way to influence the game, and therefore clearly does not deserve any sort of moral demerit for failing in it.

I stand by this verdict and the rationale behind it, but the problem should now be self-evident: *Layman's Roulette* involves an obligation that possesses my three properties of achievability, discernibility, and equitability. Therefore, if its agent does indeed lack the requisite level of efficacy regarding its eventual result and deserves no moral demerit in the event of failure, it seems that there are at least some cases in which achievability, discernibility, and equitability are insufficient to constitute a fair moral obligation. This invalidates not merely the specific premise I defended earlier in Section IV, but also the formal argument that I have been making throughout this chapter for my conclusion that all genuine moral obligations are fair to their agents.[151] Due to the lack of influence these agents possess in this set of cases, I shall call this the Problem of Helplessness.

To combat this line of criticism will require more of a shift in my argument than was the case with the Problem of Hindsight. Despite appearances, though, I do not think my ultimate conclusion is substantially weakened by admitting that my earlier claim concerning the sufficiency of my three properties is not a blanket and entirely exceptionless principle. The first reason for this is a simple point of order: as I mentioned when it was first introduced, the moral obligation involved in *Layman's Roulette* is not one which my account, nor the vast majority of other such accounts would recognise as being either fair or genuine. Hence even if one completely rejects my explanation of *why* this is the case, *Layman's Roulette* (and other such scenarios) are not counterexamples to my Fairness Principle per se, even though they are sufficient to disprove my earlier claim that the conditions of achievability, discernibility, and equitability are always sufficient for a moral obligation to be fair to its agent. Consequently, one move available to me would be to amend that premise to hold that achievability, discernibility, and equitability are sufficient for a moral obligation to be fair, *provided that obligation is genuine.* This would avoid the problematic exceptions whilst allowing me to preserve the existing structure of my argument as a whole.

Depending on your point of view, this shift may be all that is needed to account for the problem posed by *Layman's Roulette.* I, however, am unwilling to narrow the scope of my account of fairness in this way (i.e. to only concern itself with genuine duties), because I believe it is intelligible and useful to talk about the fairness of prima-facie duties in

[151] For if achievability, discernibility and equitability are necessary for a genuine moral obligation, but insufficient for a fair moral obligation, the conclusion that all genuine moral obligations are necessarily fair would be clearly false.

various contexts and that in many cases - although not all, as *Layman's Roulette* proves - the conditions which are sufficient to make such obligations fair are the same as those for genuine moral duties. I feel that any adjustment I make to my earlier argument ought to take account of these points, which only concerning myself with genuine duties would not.

In light of the problem of helplessness, I shall indeed alter my premise concerning the sufficient conditions for a fair moral obligation slightly: in the case of almost all moral obligations (whether prima-facie or genuine) that obligation's being achievable, discernible, and equitable will be collectively sufficient for it to be fair to its agent. For clarity, the relation of "almost all" here is akin to that between the crying of a hungry child and its being fed. In most cases where a child is hungry, that child's crying is sufficient for their parents to feed it, and this statement is true even though it is clearly not sufficient in *every* case - some parents are highly negligent, and some children do not have parents or carers. This new position, of course, does not formally yield my intended conclusion that all genuine moral obligations are fair even in conjunction with the other premises of my argument. But that need not necessarily be a great problem, as if there is a common thread which groups together the exceptions to the original version of my premise in a principled way, it should be possible to account for them in a manner that is consistent with my overall argument.

What might this common thread be? To understand this, we should first consider some other examples of the same kind as *Layman's Roulette* where although my chosen three properties are present, the obligation is unquestionably still unfair and the absence of the agent's efficacy in regard to its outcome is indisputable. Imagine, for instance, a royal consort's obligation to produce a male heir. Such a duty is achievable, discernible, and equitable, and yet this clearly could be neither a true moral obligation nor a fair one on my account because, as in *Layman's Roulette*, the obligated agent lacks any sort of influence over the outcome of their obligation beyond their most basic efforts (there, the placing of the bet; here, the act of becoming pregnant). For this reason, here again, no moral demerit for failing to bear a son is warranted. Further cases of this ilk might include the "obligation" to pick the correct lottery numbers, or to guess the President of the United States in 2076. It is no coincidence that these duties are all alike in the manner of their unfairness, and I submit that there is, likewise, a shared explanation for the insufficiency

of my conditions to grant efficacy (and therefore a deserved moral demerit) in this kind of case.

The reason why the properties which would under more "normal" conditions grant sufficient influence to ground the fairness of the obligation do not do so here is, I think, quite straightforward. With each of these purported obligations, including the original *Layman's Roulette*, their very content mandates that whether their agents succeed or fail in them will inevitably be dictated by factors outside of those agents' control, given the circumstances in which they are being held. In other words, the only reason achievability, discernibility, and equitability are insufficient to grant efficacy to the agents in these cases is because it is *impossible* (in those circumstances) for those agents to have such efficacy over the outcome of these obligations, owing to the very nature of what is required of them. They are, as the problem suggests, completely helpless. It therefore should not be surprising that my account counts such moral obligations as unfair despite their clearly satisfying the three conditions I originally set out, since there are no means by which they *could* be fair in the conditions the obligation is to be exercised.

Moreover, if the content of these obligations, or the circumstances in which they were held were such that this was not the case, I believe that the three properties I have been championing *would* be sufficient to grant efficacy over their outcome as normal, and in turn to justify the moral demerit that would be received for failure and render the obligation fair to its agent. If this claim appears implausible, I invite you to consider the *Layman's Roulette*-esque set of cases once again. Both *Layman's Roulette* and *Skilled Roulette* involve an identical obligation and are differentiated only by the skills of the obligated agent, yet this is enough to grant the requisite efficacy in the latter case and not the former. This phenomenon can also be observed in the other obligations of that set: if there were some means available to the royal consort to notably influence the sex of her child the unfairness of the moral duty to have a male child would seem to vanish, and being asked to predict the winning lottery numbers is substantially more reasonable if you know the means by which the machine selects them. Once again, the only scenario where the three properties I have highlighted fail to be collectively sufficient for a fair moral obligation is in cases where the obligation cannot possibly be fair in the sense I have in mind. When that impossibility disappears, so too does the exception.

I speak of the "content" of these duties to distinguish this point from our earlier discussion of excuses. Unlike there, with cases such as *Layman's Roulette* it does not simply "turn out" that the moral demerit is undeserved because of the specific, token circumstances of the obligation failure - rather, the undeserved-ness of the demerit is apparent from the moment the duty is imposed. This is because the very nature of what is being asked, given the conditions in which the obligation is held, prohibits the obligated agent from possessing the efficacy necessary for it to be fair. By contrast, this is neither the case in *Unlucky Roulette* nor the obligation to return the borrowed ten pounds, where we can easily imagine how the obligated agent might have succeeded in their obligation because they did wield influence over its outcome. That is why *those* obligations were ultimately judged to be fair to their agents despite the inappropriateness of any moral demerit for failure, whilst the *Layman's Roulette*-like cases can be safely dismissed as unfair and false obligations from the outset.

It should now be clear how examples of this latter kind are consistent with my argument that achievability, discernibility and equitability are collectively sufficient for an obligation to be fair to its agent. Together, these properties grant the quality of fairness to an obligation because they permit the agent sufficient agency in respect to its outcome that they will deserve to receive a moral demerit if they should fail to complete it. As we have seen, it is the presence of this efficacy (and not merely of the three conditions) which is truly critical to determining the fairness of a given moral obligation. In those cases where such efficacy cannot, in the circumstances, be acquired due to the nature of the obligation itself, the only possible verdict is therefore that the obligated agent will not deserve a moral demerit for failing to complete it, and hence it is unfair to them. Far from being a fatal blow to my account, however, this conclusion aligns perfectly with our natural understanding that moral obligations such as these, in circumstances such as these, could not be genuine. My reasoning here provides the link between these two strongly held intuitions by showing the concepts of a fair moral obligation and a genuine one to be fundamentally inseparable.

This completes my account of a fair moral obligation and my arguments for why the properties of achievability, discernibility, and equitability are necessary to a genuine moral obligation and both necessary and sufficient for a fair moral obligation. At this point, I maintain with confidence my conclusion that, at least in all cases where it is *possible* for a moral obligation to be fair to its agent in my chosen sense, *all* genuine moral

duties are necessarily fair to their agents in that sense. Before I proceed to the final challenge to the account of fairness I have presented in this chapter, however, I must beg a momentary diversion from these questions. An outstanding issue remains to be addressed from an earlier chapter, and only now that my account of fairness is complete are we in a position to do so correctly.

Section V.VII: A Word on "Tragic" Dilemma

In Chapter III, we discussed the potential problems of irresolvable or so-called "tragic" moral dilemma - where one is caught between genuine, but mutually exclusive moral obligations - for my account of "ought implies can." In the course of that dialectic, an interesting point was raised: is not the very idea of these situations inconsistent with my claim that genuine moral duties are intrinsically fair, and ought I not therefore to deny the existence of such dilemma outright? Previously, I claimed that there was no such inconsistency and chose a line of argument to defend my account of OIC which was essentially agnostic regarding the question of whether these "tragic" dilemma are a legitimate phenomenon. I did not provide a justification for denying this objection to my Fairness Principle at that time, as we had yet to properly explore the contours of the relevant sense of fairness. Now that we have done so in the proper detail, the time is right for me to provide an explanation for my previous claim.

Fortunately for the health of my wider case in favour of the Derivation, by this point in my argument it should already be apparent to alert readers how I would come to the conclusion that I did regarding genuine or "tragic" moral dilemma. Even though I think it is correct to say that it is "unfair" to an agent, in a certain sense, for them to be caught in a situation where they cannot fulfil all of their moral obligations, that sense is neither the one I am concerned with in this thesis nor, crucially, is it the sense which I believe is necessary to all genuine moral obligations. As noted in Sections I and II my account of fairness is concerned primarily with the properties of individual moral duties, rather than how groups of genuine moral duties may behave when taken together. The simple fact that certain moral obligations cannot be satisfied "in concert" with any or all of one's other duties is not itself sufficient to make any of those obligations unfair to their agent(s). At the risk of repeating myself, the critical factor in whether a given moral duty is fair in my terms is whether its agent holds efficacy in regard to its outcome, and it does appear

that this condition is satisfied in the paradigmatic cases of moral dilemma which we are concerned with.

To see how this is the case, consider a straightforward case of "tragic" dilemma: two men are dying in front of me and I only have time to resuscitate one of them. Here, I clearly hold efficacy in regard to my obligation to save one of the men and my obligation to save the other. If I were to attempt to save either of them I would likely (but not necessarily) succeed, and my success would be directly tied to my decision and efforts to save *that* man over the other one. In other words, if an agent finds themselves in a tragic dilemma and, after much hesitation, chooses which of their mutually exclusive obligations to fulfil, it seems they cannot rightly complain that success or failure in those obligations was not rightly "up to them." For it seems they could legitimately have chosen differently in terms of which of their obligations they did fulfil, and if they had done so what they were morally responsible for in this case (including any changes to their moral status that may occur and any praise or blame they would be entitled to) would have differed accordingly.

This is also the reason why, despite how it may initially appear, the obligations involved in a true case of tragic dilemma are not unfair by means of causing an unavoidable moral demerit. My account has already discussed how in cases of "faultless failure" like the wind blowing away the money to be return, the demerit which would normally be due for failing in a genuine moral obligation will be excused. Even if this were not the case, however, any moral demerits that occur as a result of failing in the individual (and mutually exclusive) obligations within such a dilemma would clearly *not* be unavoidable: it is written into the fabric of such cases that the obligated agent could have chosen to fulfil any, but not all, of them and hence could have avoided any (though not all) of the moral demerits which result from his choices.

Indeed, it is the presence of this efficacy in regard to *each* of one's obligations but not both together which contributes heavily to the sense of anguish that we classically associate with these dilemma. If I lacked efficacy in regard to some or all of my obligations here - if I could not save one or either of these men - there would certainly be an awful feeling of helplessness as I watched them die, but that would not be a moral *dilemma* in the relevant sense. Given this, it is only natural that these kinds of dilemma would pose no threat to my argument here.

At this point, one may be tempted to point out that whilst it is true that the requirement of efficacy is satisfied in regard to the individual obligations of a tragic dilemma - to save each of the dying men, say - it is certainly *not* satisfied in regard to any "conjoint" obligation we may possess to fulfil all of our moral requirements (since here we cannot do so). If such a genuine obligation existed it would therefore clearly be unfair according to my account, and my conclusion that all genuine obligations are necessarily fair in that sense would be falsified. But this would be not only to beg the question against OIC, but also to appeal to an argument that I have previously rebutted. For as discussed in detail in Chapter III, I reject the notion that our moral obligations agglomerate in the way that they ordinarily do[152] in circumstances where the resultant obligation would be impossible for its agent to fulfil. There is therefore no joint obligation of this kind which the agent involved in a tragic dilemma would lack efficacy in regard to, only the individual and mutually inconsistent obligations in regard to each of which they do possess such influence.

Despite how it may initially appear, then, all of the obligations involved in a true case of irresolvable moral dilemma are legitimately fair in the sense I am concerned with, and therefore present no problem for my arguments in this work. I shall now conclude this chapter by examining one final counter-argument to this Fairness Principle, and offering a thorough rebuttal of it: the problem of Ascription.

Section V.VIII: The Problem of Ascription

In this final section, I wish to examine an independent challenge to both the account of fairness I have defended in this chapter and the foundation for OIC which I intend that account to provide. Rob van Someren Greve has put forward an intriguing line of argument in his "Ought, Can, and Fairness"[153] (2014) which attacks not merely my particular method of how to understand the fairness of genuine moral duties, but the entire practice of categorising our genuine moral duties as fair or unfair in any sense whatsoever.

At this point, it is useful to refer back to a certain feature of a previous chapter. You may recall that the original case in favour of OIC which I put forward in Chapter II drew

[152] That is to say, if I am morally obligated to A, and morally obligated to B, I am morally obligated to (A+B). See Williams and Atkinson (1965).
[153] Van Someren Greve, Rob. 2014. 'Ought, Can, and Fairness', *Ethical Theory and Moral Practice*, 17, 5: 913–922.

heavily in both structure and motivation from David Copp's original attempt to derive PAP from OIC. Greve offers a series of criticisms that are specifically targeted at the underlying assumptions which run through Copp's argument, which concern the nature of fairness and its applicability to discussions of moral obligation - assumptions which are largely shared with my own Derivation of PAP, and that have been left broadly untested until now. In essence, Greve's critique holds that to apply the notion of fairness (in any sense) to our objective moral obligations is to commit a category mistake, since the concept of fairness can only be intelligibly *ascribed* to the expectations and requirements that moral agents hold toward each other. Any argument which is built upon this mistake should therefore be regarded as unsound, as it flows from a fundamentally erroneous conception of how fairness works. For this reason, I have grouped this set of objections into what I call the Problem of Ascription.

Even before we examine the particulars of Greve's argument, the threat posed by the Problem of Ascription should be clear. The truth of OIC in the sense which I have understood it is a central part of my Derivation of PAP, and my argument in favour of that understanding appealed in large part to what (I argued) was a legitimate sense in which moral duties can be said to be fair or unfair to their agents. If Greve is right that no such legitimate sense of fairness exists, therefore, the motivations for my entire project are suspect. If my earlier appeal to the fairness of moral obligations fails, we have substantially fewer reasons to accept OIC in the sense which I believe PAP logically follows from and the Derivation is severely weakened. Since the purpose of this entire chapter has been to properly explore and develop these fairness-based motivations for OIC and by extension PAP, I feel it is only fitting that I conclude it by taking on this problem and explaining why I believe Greve is mistaken in suggesting it is inherently wrong to think of moral obligations as being fair (or unfair) to the agents that hold them.

Before this, however, there is one final aspect of my account of the inherent fairness of moral duties which I should now clarify. As Greve rightly observes, the sense of fairness that Copp appeared to be working with in his original argument and which I have steadily categorised over the course of this chapter is not a strictly "distributive" notion. What this means is that it is possible for a moral agent, or a specific subset of agents, to be treated unfairly in the sense I have been investigating even if they have not been treated differently from anyone else (in the relevant respects). The hypothetical obligation in *Layman's Roulette* to win on exactly 13 black is an excellent illustration of this principle:

even if everyone at the table were somehow morally required to win on 13 black, and anyone that happened to be in the layman's position would also possess that precise obligation, it would still be legitimate for any or all of those agents to object that the duty was unfair to them in the sense I am interested in.[154] Greve, however, points out that such conceptions of fairness are not uncontroversial in the first instance, citing a famous anecdote from Fletcher by way of example:

"During a trial about alleged police brutality, a lawyer asked Sydney [Morgenbesser] under oath whether the police had beat him up unfairly and unjustly. He replied that the police had assaulted him unjustly, but not unfairly. The lawyer was puzzled. "How is that possible?" he queried. "Well," Sydney reportedly said, "They beat me up unjustly, but since they did the same thing to everyone else, it was not unfair."[155]

What this story suggests is that fairness might be *essentially* distributive: that a group of agents may be collectively treated wrongly or unjustly, but (assuming they each receive the same treatment) not unfairly.[156] This is a point of disagreement which will become important later, once Greve's true criticism of my model of fairness becomes clear. However, for the sake of argument he is ostensibly prepared to accept the possibility of non-distributive notions of fairness, and so I will take the existence of such notions for granted in replying to him. Having said this, I would also note that as far as my account of fairness is concerned I regard this particular objection to be a distinction without a difference. Although I have consistently utilised the language of fairness throughout this chapter and my thesis as a whole, I do not believe it would ultimately make any substantive difference to my overall argument if I were to classify the relevant concepts in terms of injustice (in the sense Greve is thinking of) rather than unfairness. I will now, therefore, press on to what I consider his more substantive objections.

The central concern that drives the Problem of Ascription is something which I myself touched on at the beginning of Chapter II: that morality is not a thing which behaves or has expectations of agents in the way that we do normally do of each other, and Greve argues that to treat it as though it does - even implicitly or through metaphor - is to think

[154] See Graham (2011) for more on this point.
[155] Fletcher, George P. 2005. 'Justice and fairness in the protection of crime victims', *Lewis & Clark Law Rev*, 9, 3: 547–557.
[156]There is an interesting comparison to draw here with my earlier discussion of the requirement of equitability. Might this condition be the only sense in which supporters of this view acknowledge that moral obligations may be fair or unfair to their agents in themselves?

of our moral obligations in the wrong way. As Arpaly puts it "this is where Anscombe would probably suspect us of being under the spell of regarding morality as a set of commands from a celestial boss."[157] The suggestion here, to put it more plainly, is that it is impossible to discern truths about the fairness of our genuine moral obligations by examining the circumstances in which our inter-personal demands and expectations of our fellow agents may be fair or unfair, because morality is clearly *not* an agent and its requirements are therefore not governed by the same principles as those we set for each other.

Though Greve is correct that it would be wrong to draw a direct parallel between the conditions under which agents' demands of each other are fair to them and those under which a genuine moral obligation may be fair to its agent, I should stress once again that this has never been the goal of my argument. At the risk of excessive repetition, what I am attempting in this chapter is to understand and to clarify what it means for our genuine, all-things considered moral duties to be fair to their agents, so that that understanding - together with the principle I have already defended, at length, that a genuine moral obligation is inherently fair to its agent in my chosen sense - may serve as a justification for my preferred sense of OIC and the Derivation of PAP beyond it. One important consequence of this inquiry is that it provides an explanation for why it would be unfair, for instance, for an agent to require me to do something I cannot do (or cannot discern, or which is not equitable, or which I otherwise do not hold efficacy in regard to). To put it another way, on my account of fairness the reason why my boss behaves unfairly when he expects me to make tea for the entire office in the next two minutes is that it is unfair, simpliciter, for *any* agent to be in a position where they are required to do something they cannot do in their current circumstances.

To his credit, Greve has considered this thought process and offers two reasons why it would be better for us to abandon this explanation. The first is that our ordinary, intuitive sense of an unfair demand - the same sense which, remember, I have spent this chapter turning into a concrete standard - is inherently tied to how we think about the fairness of the inter-personal requirements agents have of each other. It is natural and easy for us to make sense of the idea that *those* kinds of requirements may be unfair depending on the

[157]Arpaly, Nomy. 2006. *Merit, Meaning, and Human Bondage; An Essay on Free Will* (Princeton: Princeton University Press) p.107. For more on the linguistic "traps" that personification of morality in this way can lead us into error, see Kagan (1989).

surrounding circumstances, such as when an agent's expectations are impossible to fulfil or discern. Though Greve agrees that it is certainly possible to abstract from this and conceive of the idea that a moral obligation might "just" be unfair - unfair without there being an agent that is behaving unfairly by demanding it - "if we were to reverse the order of explanation, we would get things backwards conceptually."[158]

What is meant here is that all talk about the fairness of requirements as a whole, including moral requirements and my investigation into the (proposed) inherent fairness of such requirements is sourced in, and almost always stems from, our everyday thinking about the fairness of agent-requirements. Because of this, we now have cause to doubt that the best explanation of the fairness or unfairness of those inter-personal obligations is (as my account suggests) to be found in claims about the objective fairness or unfairness of moral obligations more generally, and so the veracity of the claim that such obligations *can* be fair or unfair to their agents is called into question.

Now, this by itself does not pose any significant problem for my position, and Greve admits that it is, for the moment, only a "suggestive" point. In truth, however, the goal of the above is to prepare the ground for his second, more central attack: that it is simply incoherent to ascribe the properties of fairness or unfairness to moral obligations in the manner that I, following Copp, have done throughout this thesis. As mentioned before, Greve regards this as a category mistake and his argument for this conclusion forms the true thrust of the Problem of Ascription. In his eyes, to suggest that deontic facts like an action's being morally obligatory in a set of circumstances[159] could be fair or unfair to the agent(s) involved in those circumstances is straightforwardly illogical, for these particular kinds of properties of our actions are simply not the sorts of things that can be assessed as fair or unfair. On this line of argument, fairness is a quality which is only applicable to the behaviour of a moral agent vis-à-vis other members of the moral community - "what can be fair or unfair is what we expect of others, how we treat them, how we distribute goods among them, and so on."[160]

Whilst Greve agrees in principle with one of the founding principles behind the Derivation - that it is indefensible to think that either our genuine moral obligations or

[158] Van Someren Greve (2014) p. 917.
[159] Or being praiseworthy in the circumstances, or morally wrong, and so on.
[160] Van Someren Greve, (2014) p. 918.

morality in general could be unfair to us - he argues that the defenders of the Fairness Principle have drawn the wrong conclusion from that truism. What we ought to take from the realisation that it is nonsensical to think of morality as unfair is not, as Copp and I have done, that our moral obligations must therefore be fair to us by default, but that fairness and unfairness are simply inapplicable terms when we are describing the moral or deontic status of our actions. If this assessment is correct, then the explanation of the unfairness of certain individual agent-requirements I suggested above will be false. More troublingly for my Derivation, it would also mean my efforts throughout this chapter to establish a legitimate and meaningful sense of fairness as the basis for OIC (and in the longer term, for PAP) would amount to nothing. If it should transpire that our genuine moral obligations are not after all the kinds of things which may be rightly assessed as fair or unfair to us, it cannot then be the case that the best reason to accept OIC is that we would face unfair, but nevertheless genuine moral obligations (of the kind I have consistently argued are impossible) if we do not. As Greve puts it, "if this is correct, [the Fairness Principle] must be rejected as meaningless, and therefore as incapable of supporting any conclusion whatsoever."[161]

But why ought we to accept that conclusion? Greve believes the argument form in which I have employed the Fairness Principle to motivate the premises of the Derivation to be clearly invalid, and hence we should reject the Principle as a valid foundation for OIC (or, indeed, any other moral principle). He begins his attack by pointing out that the core logic of the claim that all genuine moral obligations are necessarily fair - "it would be unfair if X, therefore ~X"[162] - is certainly not universally valid, because there is ample evidence of existing unfairness in the world. It would be unfair to pay men and woman differently for the same work, he muses, but we clearly could not deduce from this that they are always paid equally. There is no shortage of examples of this kind: it is unfair for the rich to avoid taxes which the poor cannot, but they undeniably do. It is unfair that our agent in *Unlucky Roulette* is gunned down due to mistaken identity, but he was. However, when the "X" in question concerns a moral claim, as in this case, things are more complex. For example, it would be unfair for me to drink the entire bottle of wine we bought together, and that is why I do not do so. It would be unfair for the government to levy a universal 90% rate of income tax, and so it does not.

[161] Van Someren Greve, (2014) p. 918.
[162] Put precisely, my argument in these terms would read "it would be unfair if there were genuine moral duties which we could not fulfil (or discern, or otherwise lacked efficacy regarding), therefore there are no such duties."

Still, Greve maintains that even within the moral sphere, such statements would prove far too much if we accepted them as a legitimate form of argument, and therefore we ought not to do so. He offers the following example in support of this point: imagine that Sally is a dedicated member of a team of lifeguards, but her co-workers are lazy and routinely make her do everything. One day, a child is in danger of drowning and Sally knows that she will have to be the one to rescue them, even though she has already done far more than her fair share of the work today. In this case, Greve suggests that because Sally's fellow lifeguards will not bother to help, and Sally knows this, she is genuinely morally obligated to rescue the child. This, he argues, is proof that my central claim that genuine moral duties are inherently fair to their agents is false, because he regards it as obviously unfair that Sally should have to be the one to fulfil this obligation and equally obvious that she is still genuinely obligated to rescue the drowning child regardless. Were we to stick to my foundational moral principle that we cannot possess a genuine moral obligation that is unfair to us, his argument goes, we would have to deny that Sally is under an obligation to rescue the child in these circumstances, and that is clearly mistaken.

From Greve's perspective, this point also generalises to a great many examples of "unequal" moral obligations. Far from being an outlier, he notes, it is frequently the case that our collective moral burden is unfairly distributed in this way, not because of any properties of those duties but as a result of the apathy or inaction of those who could help but choose not to. In none of these situations, however, would it be correct for us to rule (as Greve understands my argument to do) that the moral obligations involved are not genuine as a result of this kind of unfairness, and we should not accept the type of argument - that is to say, the Fairness Principle - which leads us to that false conclusion. Far from providing solid grounds on which to support OIC, then, the Problem of Ascription holds that my analysis of fairness fails to even offer a viable base for the broader view that we can sensibly ascribe the properties of fairness or unfairness to our genuine moral obligations to begin with.

Although the Problem of Ascription appears to strike at both the very heart of the account of fairness I have presented in this chapter, and in a more extended sense of my project as a whole, I am not persuaded by it. Specifically, I believe there is a major equivocation taking place in regard to Greve's appeal to fairness and unfairness at various points throughout his attack. Once we have lined up his argument against the specific sense of

fairness which I have argued is necessary to all genuine moral obligations, it will be clear that it does not undermine my Fairness Principle in any meaningful way.

To understand the important weakness in Greve's case against me, we must return to the clarification I offered at the beginning of this section. As I have explained, and as Greve himself acknowledges at the beginning of his objections, the sense of fairness which I suggest is necessarily satisfied by all genuine moral obligations is explicitly not a distributive one. As Greve's example shows, it is undoubtedly true that there are distributive notions of fairness which are well-grounded, and this is part of the reason I acknowledge that the understanding we have been discussing is not the only meaningful sense in which something may be fair or unfair to a moral agent. I do not, however, believe these distributive notions to be suitable ways of categorising the specific manner in which a moral obligation may be unfair to its agent, the details of which have been set out at length elsewhere in this work. When I say, then, that no genuine moral duty can be unfair, I must repeat for the final time that I do not mean that our moral obligations must be fair according to each and every valid interpretation of the concept.

I make this point because, Greve's argument appears to slide between making use of both distributive and non-distributive senses of fairness as it proceeds. He begins by acknowledging both the possibility of a non-distributive sense of fairness and that defenders of the Fairness Principle such as myself are not thinking of fairness in purely distributive terms. Yet he objects to the conclusion that all moral obligations must be fair to their agents, in this specific sense, by appealing to the example of Sally and the lazy lifeguards. Here things become murkier. Greve suggests Sally's genuine obligation to rescue the drowning child - and the many other cases this generalises to - is unfair to her on the basis that she has already pulled far more than her moral "weight" in comparison to her slothful fellows. To put it another way, having done so much already it is unjust that she should be morally required to do more when others have done nothing. This, surely, is an appeal to a distributive notion of fairness, which Greve has already conceded is *not* the sense in which I am suggesting all moral obligations are necessarily fair. For if Sally's compatriots had been just as diligent as she was in regard to their own moral obligations in the past, there would seem to be no unfairness - in the sense Greve seems to have in mind - in asking her to rescue this particular child, and as with the original example this same conclusion generalises to the many other examples we could imagine of this type.

Even if Greve is correct that there is a legitimate sense of fairness according to which this obligation is unfair to Sally, therefore, examples such as these pose no danger to my account of fairness or the ways in which it supports the other premises of the Derivation. All that is required for my account of fairness to function on its own merits and to perform the task of grounding my chosen sense of OIC is for our genuine moral obligations to meet the necessary and sufficient criteria for fairness that I have set out earlier in this chapter. As I have stated before, whether these obligations fail to meet such criteria for any *other* understanding of fairness is of no consequence to my argument.

It is important, as well, to recognise at this point that those necessary and sufficient conditions for a fair obligation which this chapter has explored at great length *are* satisfied both in the specific case of Sally's obligation to the drowning child and the various other cases which generalise from it, even though there is an obvious unfairness (in a different sense) in how the moral "workload" has been assigned. In the lifeguards example, Sally can undeniably discern her obligation to rescue the child and is capable of achieving it in the relevant sense, and anyone in her position would hold the same obligation which she does - indeed, the equitability of this particular obligation is plainly shown by the fact that her fellow lifeguards *are* also under it, and their refusal to act accordingly is the source of its supposed unfairness. It may also be safely assumed that Sally holds sufficient efficacy in regard to the outcome of her obligation since - based on the parameters we are given - it seems that if she does make the effort to rescue the child she will likely succeed, but is naturally not guaranteed to do so.

I therefore agree with Greve that Sally is genuinely obligated to rescue the child because of her co-workers' inaction, but I dispute that this would be an unfair obligation in the sense that I am suggesting is impossible for any genuine moral duty. Rather, as far as my account of fairness is concerned this is simply the latest in an unfortunate series of demanding, but unambiguously fair moral obligations for Sally and a straightforward case of moral wrongdoing on the part of her co-workers. It therefore does not follow that the argument form I have employed justifies too much, as Greve would have it. It simply justifies exactly the set of moral obligations which I have intended it to do all along. As a result, his argument fails to establish his desired conclusion that deontic facts like moral obligations are not items of the correct category to be assessed as fair or unfair, nor does it prove that there may be genuine moral obligations which are unfair to their agents. My

account of fairness is thus unharmed by the Problem of Ascription, as was the case with the problems of Hindsight and Helplessness, and it is now finally ready to provide the principled foundation which I have been seeking for my account of OIC, of PAP, and the premises which connect them.

By this point in my project to conclude the Principle of Alternate Possibilities from "ought implies can", both the motivations and the full reasoning in favour of my Derivation should now be fully clear to see. I have sought in this chapter to set out, in as much detail as possible, the foundational moral principles upon which the arguments of this thesis are based, and to codify the intuitive and natural sense of fairness to which we often appeal when discussing the moral requirements placed upon on agents. Having now done this, it is my honest and strongly-held assessment that our genuine moral obligations not only are inherently fair to the agents which hold them, in at least one rigorous and meaningful sense of that term, but also *must* be so for their authority and obligatory status to persist or for agents to hold any reason to abide by them. I believe this sense of fairness mandates that a genuine moral obligation is achievable, equitable, and discernible by the agent that holds it, and that they will therefore possess sufficient efficacy in regard to how their actions will affect the outcome of their obligation that the resulting changes in their moral status will be justified.

This Fairness Principle yields the conclusion that "ought implies can" - that only our achievable moral duties, as I have defined them, can possibly be genuinely obligatory - and "ought implies can", as we have conclusively seen over the course of my work in this project, yields PAP.

In Conclusion

The aim and purpose of this PhD thesis has been to defend two claims: firstly, that there is a logically valid argument which derives the Principle of Alternate Possibilities in respect to blameworthiness from the principle that "ought implies can", and secondly to show that this argument is philosophically sound. If I have succeeded in my efforts here to establish both of these claims, my work will therefore constitute a proof of both OIC and PAP in the senses that I have defended them here.

At this point, I am confident that I have accomplished the first of these objectives beyond any reasonable doubt. The Derivation that I have put forward in this work begins from the position that OIC is true (Premise 1), and from there argues that there is an inherent connection between blameworthiness and moral wrongdoing - that if an agent deserves blame for an action or inaction of theirs, then that action or inaction was necessarily morally wrong (Premise 2). Since it is morally incumbent on us not to perform an action that is morally wrong (Premise 3), it follows that, given "ought implies can", an agent may be morally blameworthy for an action only if they were able to refrain from performing it (Premise 4). The Derivation's final move is to link this ability to refrain from acting in a given way with the ability to act otherwise than they did; any agent that possesses the former ability in regard to a given action also possesses the latter (Premise 5). If we accept these five premises, as well as the unstated assumption that OIC and PAP utilise the same sense of morally relevant ability, they yield together the conclusion that an agent may only be blameworthy for an action (or inaction) if they possessed the power to act otherwise than they did. Thus, the Principle of Alternate Possibilities is logically derivable from at least a certain interpretation of "ought implies can."

With this in mind, each chapter of this thesis has been structured around the pursuit of my second aim - to show that my Derivation, presented above, is not merely valid but also sound. In Chapter I we reviewed the history of the debate over PAP in order to place the Derivation in its proper context, and show how Copp's original efforts in support of it approached the question of alternate possibilities from a new direction. Once the precise formulation of my Derivation was made clear, I assigned Chapters II and III to providing a solid foundation for its starting premise that "ought implies can." In these chapters, I

put forward the particular form of OIC that I have been interested in throughout this thesis. First, I laid out my precise formulation of the "ought", "implies", and "can" in question, before examining what I believe are the most potent reasons in favour of the principle and replying to a diverse series of arguments against it. In Chapter IV, we explored what I have called the "connecting premises" of the Derivation (Premises 2, 3, and 5), and similarly repulsed a range of challenges to them from across the philosophical spectrum.

With each premise of the Derivation having been put to the test and displayed its quality, I devoted my final chapter to a thorough examination of what I have called the Fairness Principle: my belief, first suggested in my original defence of OIC, that all genuine moral obligations must be fair to their agents. There, I clarified exactly what I believe is entailed by this requirement, explained the necessity of the qualities of achievability, discernibility, and equitability to any genuine moral obligation, and put forth an argument that those properties were collectively sufficient to constitute a fair moral obligation as I have defined it. The reason I chose to spend so much time on this Fairness Principle is simple - as we have seen, it forms the essential core of my argument in favour of OIC and provides - by extension - one of the most compelling motivations to accept the Derivation of PAP. At this stage of the proceedings, I feel secure in my argument that a genuine moral obligation inherently possesses the qualities which are required for it to be fair to its agent, and the truth of this principle is the bedrock upon which my wider project is built.

I wish now to turn my attention to an important question that may have faded into the background during our long road through the body of my project. What, precisely, is the *purpose* of the Derivation? It is an argument aimed at concluding PAP, yes, but as we covered in Chapter I there are many such arguments, and certainly no shortage of indefatigable supporters and critics of PAP to debate them. What has been gained by my focus on this specific method of arguing for the principle?

I believe there are two equally important answers to this question. The first may be considered a "tactical" reason why the Derivation matters. The reason I chose to begin

Chapter I with a comprehensive review of the history of the debate over PAP was, as I drew attention to at the time, that the classic historical debate over the veracity of the principle has steadily ground to a halt in recent years. Worse still, the focus of debate on PAP's consistency or lack thereof with determinism has meant that comparatively little attention has been paid to the principle's relationship with OIC, with many of those who reject PAP - including Frankfurt himself - supporting it without acknowledging any kind of connection between the two ideas. By highlighting, as I have done here, that those who accept OIC and the (commonly held) connecting premises of the Derivation are also committed to PAP as regards blameworthiness, I do not merely demonstrate the truth of the latter. What my work here has exposed is that the more widespread acceptance of OIC than of PAP, at least in the senses I have understood them throughout this work, is a notable instance of collective misunderstanding of the relevant concepts.

This is the first advantage of the Derivation over other strategies to defend PAP: in proving that the Derivation is sound, I have also shown that several, perhaps many of those critics who favour OIC but not PAP are mistaken about their normative commitments, and that PAP is not merely a true moral claim but ought to be regarded as a much more popular principle than it currently is.

Whilst this would be a significant achievement in its own right, my choice of the Derivation to establish PAP also carries an additional benefit. Through arguing for the necessity of alternate possibilities on the basis of their connection with "ought implies can", my argument has explored the links between our moral obligations and blameworthiness and uncovered some highly interesting truths about the interaction between OIC and PAP. By choosing to demonstrate PAP by means of the Derivation, we can now see the importance of the shared motivations which underpin both it and "ought implies can." Despite their seemingly different subject matters - OIC concerned with the scope of our moral duties, PAP with the necessary conditions for blameworthiness - my exploration of the reasons behind these principles reveals each of them, as well as the secondary premises which connect them, to be expressions of the same essential view of the function and purpose of genuine moral obligations and moral judgement.

Perhaps most noteworthy of all, my study and defence of the Derivation in this thesis has also revealed the central importance of the concept of *fairness*, so often undiscussed or under-developed, to many of the arguments and claims which are offered in defence of OIC and PAP alike. So far as I know, almost no-one has attempted to link the concepts of fairness, genuine moral obligation and blameworthiness together in the way that I have, by defending and codifying the intuitive belief that each of our moral obligations must be fair to the agents who receive them in order to be genuine, and for those agents to properly deserve blame for falling short of them. But I believe my extensive work on the subject in Chapter V to now have shown these concepts, when understood in the ways that I have, to be essentially connected in this way. What has been gained from my work on the Derivation in this thesis, then, is a deeper understanding of the role that fairness plays not just in the defence of OIC and PAP, but in how we think about our moral obligations, their role in determining our moral status and how other agents are morally entitled to treat us - a valuable addition to the debate even if my Derivation should ultimately fail in the task of proving PAP.

Though I am proud of the breadth and depth of my work in this thesis, it would be remiss of me not to acknowledge some of the topics which, whilst of obvious relevance to the questions of my research, I was not able to discuss in the course of defending my Derivation. We raised briefly in Chapter IV the question of how we ought to determine exactly *which* abilities, and abilities to do otherwise in particular, are the normatively relevant ones to use when applying the criteria set by OIC and PAP. Whilst I have offered what I believe to be the beginnings of the correct answer to this question in my discussion of Yaffe - that working to comprehend the precise nature of our genuine moral duties and of blameworthiness will yield understanding of which senses of ability are the relevant ones - I know there is a great deal more work on this subject to be done before we can be said to have reached that detailed understanding. I look forward to continuing my research into this fascinating topic in my future work.

The other major question which I have been forced to omit from this project concerns, perhaps unsurprisingly, my Fairness Principle. Whilst I discussed several of the implications of this principle in Chapter V, I am certain that I have only begun to scratch the surface of its full commitments and consequences. In particular, the claim that all

genuine moral obligations are inherently fair to their agents - even in the limited sense of "fair" that I have set out - creates some very interesting interactions with some perennial philosophical problems. One of these which I am especially eager to investigate is how this principle interacts with the classic problem of moral luck. For as we discussed at length in Chapter V, the fairness of a given obligation upon a given agent is heavily dependent on the circumstances and context in which it is held, and the phenomenon of moral luck has notoriously shown us how these are so often beyond the obligated agent's control, both in terms of those circumstances themselves and in how their actions and decisions will turn out (so-called "resultant" luck). The various roulette-themed cases we examined in Chapter V have offered only a taste of the complexities which await us here.

Finally, let us assume for a moment that my Derivation is a success and that OIC and PAP in regard to blameworthiness are true. Knowing this, where might we next direct our attention? As mentioned above, the question of what *makes* a certain kind of ability (or abilities) normatively relevant when applying moral principles like OIC and PAP remains an open line of philosophical enquiry, even if one accepted the specific sense of ability that I have argued for here as the correct one. One might also decide to see how far the broader ethical framework of our obligations and blameworthiness being constrained by fairness which I have been immersed in can be extended. Could it be the case, for instance, that the sense of fairness and the necessity of the ability to do otherwise as I have understood them here are applicable to other moral conversations, such as the nature of praiseworthiness or moral responsibility more generally, or certain areas of applied ethics? By the same token, we might also wonder whether fairness may (or must) serve the kind of guiding and constraining role I have in mind for it even when it is understood in a different sense than I have argued for. Now that we have seen that fairness can constrain our duties and, in turn, the conditions under which we may deserve blame, it is only natural to wonder whether other kinds of fairness - distributive notions, perhaps - may do the same.

I do not know the answers to these questions, nor I suspect are they likely to receive renewed attention simply because of what you have read here today. Such is the fundamental importance of the subjects of this project to the discipline of metaethics as a

whole, however, that I am sure they will continue to remain a topic of much conversation for years to come.

In writing this thesis, I owe a great intellectual debt to the work of David Copp, the original architect of the Derivation. I feel it is only fitting, therefore, that I entrust my final words on the defence of the Principle of Alternate Possibilities to him:

"I have not been able to devote as much attention to all of these issues as would be required in a fully adequate defence of PAP. I hope nevertheless to have shown at least that there is a coherent, plausible, and attractive position that accepts both PAP and the Maxim. According to this position, a person is not morally required to do something unless she can do it, and a person is not blameworthy for doing something unless she could do otherwise."[163]

Needless to say, I find myself in complete agreement.

- Sam Waters.

[163] Copp, David. 2006. 'Ought Implies Can, Blameworthiness, and the Principle of Alternate Possibilities', in *Moral Responsibility and Alternative Possibilities* (Aldershot: Ashgate Publishing Ltd) 265-301 (p. 296).

Bibliography

Alvarez, Maria. 2009. ‘Actions, thought-experiments and the ‘Principle of alternate possibilities’’, *Australasian Journal of Philosophy*, 87, 1: 61-81.

Alvarez, Maria and Clayton Littlejohn. Forthcoming. ‘When Ignorance is No Excuse’, in *Responsibility: The Epistemic Condition*, ed. by P. Robichaud and J. Wieland, (Oxford University Press).

Arpaly, Nomy. 2006. *Merit, Meaning, and Human Bondage; An Essay on Free Will*, (Princeton: Princeton University Press).

Ayer, A. J., ‘Freedom and Necessity’, in *Philosophical Essays*, (London; Palgrave Macmillan, 1954) pp. 271-284.

Bergström, Lars. 1996. ‘Reflections on consequentialism’, *Theoria*, 62, 1-2: 74-94.

Bramhall, John. 2008. ‘Selections from Bramhall, A Defense of True Liberty’, in *Hobbes and Bramhall on Liberty and Necessity*, ed. by Vere Chappell, (Cambridge; Cambridge University Press), pp. 43-69.

Capes, Justin A. 2012. ‘Blameworthiness without wrongdoing’, *Pacific Philosophical Quarterly*, 93, 3: 417-437.
—2014. ‘The Flicker of Freedom: A Reply to Stump’, *The Journal of Ethics*, 18, 4: 427-435.

Chisholm, Roderick. 1964. ‘Human Freedom and the Self’, in *Lindley Lectures* 4, (University of Kansas, Department of Philosophy).

Copp, David. 1997. ‘Defending the Principle of Alternate Possibilities: Blameworthiness and Moral Responsibility’, *Noûs*, 31, 4: 441-456.
—2006. ‘Ought Implies Can, Blameworthiness, and the Principle of Alternate Possibilities’, in *Moral Responsibility and Alternative Possibilities*, (Aldershot: Ashgate Publishing Ltd), pp. 265-301.
—2008. ‘'Ought' Implies 'Can' and the Derivation of the Principle of Alternate Possibilities’, *Analysis*, 68, 1: 67-75.

Davidson, Donald. 2001. ‘Freedom to Act’, in *Essays on Actions and Events*, 2nd edn, (Oxford; Clarendon Press), pp. 63-82.

Dennett, Daniel C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*, (The MIT Press).

Fischer, John Martin. 1994. *The Metaphysics of Free Will; An Essay on Control*, (Oxford; Blackwell Publishers Inc.).

Fletcher, George P. 2005. ‘Justice and fairness in the protection of crime victims’, *Lewis & Clark Law Rev*, 9, 3: 547–557.

Frankfurt, H.G. 1969. ‘Alternate Possibilities and Moral Responsibility’, *Journal of Philosophy*, Volume 66, 23: 829-839.

—1998. ‘What We Are Morally Responsible For’, in *The Importance of What We Care About*, (New York; Cambridge University Press), pp. 95-104.

Franklin, Christopher E. 2011. ‘The Problem of Enhanced Control’, *Australasian Journal of Philosophy*, 89, 4: 687–706.
—2015. ‘Everyone thinks that an ability to do otherwise is necessary for free will and moral responsibility’, *Philosophical Studies*, 172, 8: 2091-2107.

Ginet, Carl. 1996a. ‘In Defence of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing’, *Philosophical Perspectives*, 10, 403-417.
—1966b. ‘Might We Have No Choice?’, in *Freedom and Determinisim*, ed. by Keith Lehrer, (Random House) pp. 87-104.

Gowans, Christopher W. ed. 1987. *Moral Dilemmas*, (New York; Oxford University Press).

Graham, Peter. 2011. ‘Ought and Ability’, *Philosophical Review*, 120, 3: 337-382.

Haji, Ishtiyaque. 1998. *Moral Appraisability; Puzzles, Proposals, and Perplexities*, (New York: Oxford University Press).
—2017. ‘The Obligation Dilemma’, *The Journal of Ethics*, 21, 1: 37-61.

Hobbes, Thomas. 2008a. ‘Hobbes’ treatise of liberty and necessity’, in *Hobbes and Bramhall on Liberty and Necessity*, ed. by Vere Chappell, (Cambridge; Cambridge University Press), pp. 14-43.
—2008b. ‘Selections from Hobbes, The Questions concerning Liberty, Necessity, and Chance’, in *Hobbes and Bramhall on Liberty and Necessity*, ed. by Vere Chappell, (Cambridge; Cambridge University Press), pp. 69-91.

Honoré, A. M. 1964. ‘Can and Can’t’, *Mind*, 73, 292, 463-479.

Howard-Snyder, Frances. 1997. ‘The Rejection of Objective Consequentialism’, *Utilitas*, 9, 2: 241-248.
—2006. ‘“Cannot” Implies “Not Ought”’, *Philosophical Studies*, 130, 2: 233-246.

Hughes, Nick. 2018. ‘Guidance, Obligations and Ability: A Close Look at the Action Guidance Argument for Ought-Implies-Can’, *Utilitas*, 30, 1: 73-85.

Hume, David. 2004. *A Treatise of Human Nature*, new edn, (Mineola: Dover Publications, Inc.).
—2008. *An Enquiry concerning Human Understanding*, ed. by Peter Millican, (Oxford; Oxford University Press).

Jacquette, Dale. 1991. ‘Moral Dilemmas, Disjunctive Obligations, and Kant’s Principle that “Ought Implies Can”’, *Synthese*, 88, 1: 43-55.

Janzen, Greg. 2013. ‘Frankfurt Cases, Alternate Possibilities and Prior Signs’, *Erkenntnis*, 78, 5: 1037-1049.

Kagan, Shelly. 1989. *The Limits of Morality*, ed. by Derek Parfit, (Oxford: Clarendon Press).

Kane, Robert. 1985. *Free Will and Values*, (Albany: State of New York Press).

Kant, Immanuel. 1998. *Religion Within the Boundaries of Mere Reason and Other Writings*, ed. by Allen Wood and George di Giovanni, (Cambridge: Cambridge University Press).
—2012. *Groundwork of the Metaphysics of Morals*, revised edn., tr. by Mary Gregor, ed. by Jens Timmermann, (Cambridge; Cambridge University Press).
—2015. *Critique of Practical Reason*, revised edn., tr. by Mary Gregor, (Cambridge; Cambridge University Press).

Kittle, Simon. 2015. 'Abilities to do otherwise', *Philosophical Studies*, 122, 11: 3017-3035.

Larvor, Brendan. 2010. 'Frankfurt counter-example defused', *Analysis*, 70, 3: 506-508.

Lawford-Smith, Holly. 2013. 'Non-Ideal Accessibility', *Ethical Theory and Moral Practice*, 16, 3: 653-669.

Lemmon, E. J. 1962. 'Moral dilemmas', *Philosophical Review*, 71, 2: 139-158.

Lewis, David. 1983. 'Scorekeeping in a Language Game', in *Philosophical Papers Volume I*, (Oxford University Press), pp. 233-250.
—1987. 'The Paradoxes of Time Travel', in *Philosophical Papers Volume II*, (New York: Oxford University Press). pp. 67-81.

Lord, Errol. 2015. 'Acting for the right reasons, abilities, and obligation', in *Oxford Studies in Metaethics*, 10, ed. by Russ Shafer-Landau, (Oxford: Oxford University Press), pp. 26-52.

Nagel, Thomas. 1979. 'Moral Luck', in *Mortal Questions*, (Cambridge; Cambridge University Press) pp. 24-39.

Makin, Stephen. 1996. 'Megarian Possibilities', *Philosophical Studies*, 83, 3: 253-276.

Makinson, David. 1987. 'Fred Feldman. Doing the best we can. An essay in informal deontic logic.' Philosophical studies series in philosophy, vol. 35. D. Reidel Publishing Company, Dordrecht, Boston, Lancaster, and Tokyo, 1986, xiv + 244 pp.', *The Journal of Symbolic Logic*, 52, 4: 1050-1051.

Mason, H. E., ed. 1996. *Moral Dilemmas and Moral Theory*, (New York; Oxford University Press).

McKenna, Michael. 2012. *Conversation & Responsibility*, (New York: Oxford University Press).

Mele, Alfred R. 2003. 'Agents' Abilities', *Noûs*, 37, 3: 447–470.

Mele, Alfred R. and David Robb. 1998. 'Rescuing Frankfurt-style cases', *Philosophical Review*, 107, 1: 97-112.

Moore, G. E. 1912. *Ethics*, (London; Williams and Norgate).

Mothersill, Mary. 1996. ‘The Moral Dilemmas Debate’, *Moral Dilemmas and Moral Theory*, ed. by H. E. Mason, (New York; Oxford University Press), pp. 66-86.

Naylor, M. B. 1984. ‘Frankfurt on the principle of alternative possibilities’, *Philosophical Studies*, 46, 249-58.

Nelkin, Dana Kay. 2011. *Making Sense of Freedom and Responsibility*, (New York: Oxford University Press).

Pettit, Philip. 2001. ‘The Capacity to Have Done Otherwise’, in *Relating to Responsibility: Essays in Honour of Tony Honoré on his 80th Birthday*, ed. by Peter Cane and John Gardner, (Oxford: Hart Publishing) pp. 21-35.

Plato. 1992. *Republic*, tr. by G. M. A. Grube, ed. by C. D. C. Reeve, (Indianapolis; Hackett Publishing Company).

Robinson, Michael. 2012. ‘Modified Frankfurt-type counterexamples and flickers of freedom’, *Philosophical Studies*, 157, 2: 177-194.

Schlick, Moritz. 2008. *Problems of Ethics*, tr. by David Rynin, (New York; Nielsen Press).

Schwan, Ben. 2018. ‘What ability can do’, *Philosophical Studies*, 175, 3: 703-723.

Simkulet, W. 2015. ‘On the Signpost Principle of Alternate Possibilities: Why Contemporary Frankfurt-Style Cases are Irrelevant to the Free Will Debate’, *Filosofiska Notiser*, 2, 3: 107–120.

Sinnott-Armstrong, William. 1984. ‘’Ought’ Conversationally Implicates ‘Can’’, *The Philosophical Review*, 93, 2: 249-261.
—1985. ‘'Ought to Have' and 'Could Have'’, *Analysis*, 45, 1: 44-48.

Speak, Daniel. 2002. ‘Fanning the Flickers of Freedom’, *American Philosophical Quarterly*, 39, no. 1, 91-105.

Stern, Robert. 2004. ‘Does “Ought” Imply “Can”? And Did Kant Think It Does?’, *Utilitas*, 16, 1: 42–61.

Stump, Eleonore. 1999. ‘Alternative Possibilities and Moral Responsibility: The Flicker of Freedom’, *The Journal of Ethics*, 3, 4: 299–324.

Trigg, Roger. 1971. ‘Moral Conflict’, *Mind*, 80, 317: 41-55.

Van Inwagen, Peter. 1978. ‘Ability and Responsibility’, *Philosophical Review*, 87, 2: 201- 224.
—2008. ‘How to Think about the Problem of Free Will’, *The Journal of Ethics*, 12 3-4: 327-341.

Van Someren Greve, Rob. 2014. ‘Ought, Can, and Fairness’, *Ethical Theory and Moral Practice*, 17, 5: 913–922.

Vihvelin, Kadri. 2000a. ‘Freedom, foreknowledge, and the principle of alternate possibilities’, *Canadian Journal of Philosophy*, 30, 1: 1-23.
—2000b. ‘Libertarian Compatibilism’, *Noûs*, 34, 14: 139-166.
—2004. ‘Free will demystified: a dispositional account’, *Philosophical Topics*, 32, 1: 427-450.
—2013. *Causes, laws and free will: Why determinism doesn’t matter*, (New York: Oxford University Press).

Vranas, Peter B.M. 2007. ‘I ought, therefore I can’, *Philosophical Studies*, 136, 2: 167-216.

Whittle, Ann. 2016. ‘Ceteris Paribus, I Could Have Done Otherwise’, *Philosophy and Phenomenological Research*, 92, 1: 73-85.

Williams, Bernard. 1973. *Problems of the Self*, (Cambridge; Cambridge University Press).
—1981. *Moral Luck; Philosophical Papers 1973-1980*, (Cambridge; Cambridge University Press).

Williams, Bernard and W. F. Atkinson. 1965. ‘Symposium: Ethical Consistency’, *Proceedings of the Aristotelian Society, Supplementary Volumes*, 39: 103-138.

Widerker, David. 1991. ‘Frankfurt on ‘Ought Implies Can’ and Alternative Possibilities’, *Analysis*, 51: 222-224.
—1995. ‘Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities’, *The Philosophical Review*, 104, 2: 247-261.

Wyma, Keith D. 1997. ‘Moral Responsibility and the Leeway for Action’, *American Philosophical Quarterly*, 34, 1: 57-70.

Yaffe, Gideon. 1999. ‘'Ought' implies 'can' and the principle of alternate possibilities’, *Analysis*, 59, 3: 218–22.

Young, Garry. 2016. ‘The Principle of Alternate Possibilities as Sufficient but not Necessary for Moral Responsibility: A way to Avoid the Frankfurt Counter-Example’, *Philosophica*, 44, 3: 961-969.

Zimmerman, David. 1994. ‘Acts, Omissions and "Semi-Compatibilism"’, *Philosophical Studies*, 73, 2-3: 209-223.

Zimmerman, M. J. 1996. *The Concept of Moral Obligation*, ed. by Ernest Sosa, (Cambridge: Cambridge University Press).
—1997. ‘A Plea for Accuses’, *American Philosophical Quarterly*, 34, 2: 229-243.