

Liability and Culpability



Matthew Oliver

Merton College

University of Oxford

Submitted for the degree of,

Doctor of Philosophy

Trinity 2018

Liability and Culpability*

Abstract:

A person's choices and actions can make them liable to defensive harms. Yet the nature of moral liability, its limits and justifications, are the subject of considerable debate. This thesis is a contribution to those debates. Its starting point is the idea that we can better understand both liability and culpability by examining the connection between the two.

I offer an account of moral liability according to which a liability is a reason to give less weight to the interests of the liable person. To defend this account I discuss proportionality, necessity, and hard cases of aggregation and multiple justifications. I justify liabilities by appealing to the significance of our choices, the costs attached to those choices, and the reasons we have for making them. Culpable choices, such as the decision to attack another, can make a person liable to defensive harms. Our enforceable duties are closely related to liabilities, yet do not themselves directly explain or justify our liabilities. I discuss the relationship between moral responsibility, causal responsibility, and liability. I argue that causal responsibility for a threat of harm is not a necessary condition of liability to be harmed in the prevention of that threat.

My account of the justification for liability is directly linked to my account of the nature of culpability. I argue that a person is culpable to the extent that they manifest insufficient sensitivity to action-guiding moral reasons in the exercise of their agency. I elaborate this account to show how it fits in between volitionist and attributivist theories of culpability. I discuss the ways in which we can be culpable for past actions and for our possible future actions, the problem of moral luck, of moral ignorance, the relationship between culpability and permissibility, and excuses that do not deny culpability.

* I thank the Clarendon Fund, Merton College, and my inspiring supervisors Jeff McMahan and John Gardner. I also thank Victor Tadros, Tom Sinclair, Patrick Tomlin, Cecile Fabre, Ralf Bader, and Ketan Ramakrishnan. The many mistakes in this thesis are likely a result of my failure to follow the advice of these brilliant philosophers. I could not have completed this thesis without the support and encouragement of Yosiane White.

Contents

Introduction	4
Chapter 1: What are Moral Liabilities?	8
Chapter 2: Choice and the Justification of Liabilities	43
Chapter 3: Duties and Liabilities	76
Chapter 4: Culpability, Causation, and Liability	106
Chapter 5: Culpability in Theory	144
Chapter 6: Culpability in Action	183
Conclusion	218
Bibliography	220

Introduction

Other things being equal, it is wrong to kill five people in order to save just one. This is about as self-evident as such sweeping statements get. It reflects two widely held assumptions about what we ought to do. The first is that we ought to minimize the amount of harm that occurs. One death, while bad, is not as bad as five deaths. The second is that killing a person is much harder to justify than allowing a person to be killed.

As we know, other things are not always equal. Perhaps the five would die anyway in a few seconds, or perhaps killing the five is also a necessary side-effect of saving a further million lives. Either of these possibilities might justify killing the five, without making an exception to our basic assumptions. Yet there are also exceptions, based on facts far less palpable than the insignificance of a few seconds or the significance of a million lives.

Here is an example. Suppose that one innocent victim is attacked by five highly culpable would-be murderers. We can save the victim from being killed, but only by killing the five attackers. In this case, it seems obvious that we should kill the five to save the one. Why is this obvious? In recent years, these cases have received considerable attention. This literature addresses what it is, precisely, about the five attackers that makes them liable to be killed. In this thesis, I argue that it can be permissible to kill the five because they chose to attack the one. In other words, I argue that choice and culpability are a key basis for liability to defensive harming.

More broadly, this thesis explores the ways in which those things that make us moral agents—our capacity to make morally significant choices, to bear duties, to respond to moral reasons—alter the way we can be permissibly treated by others. These capacities may give us a special moral status, and entitle us to special protections, but they also make us uniquely vulnerable.

As moral agents, we risk becoming culpable, and we also risk becoming liable to be harmed.

Culpability and liability are, in many ways, the price we pay for responsibility.

This thesis is divided into six chapters. In Chapter One I introduce the concept of moral liability and two respects in which I believe our current understanding of liability could be improved. I argue that liability should not be defined as the absence or forfeiture of rights. We should also distinguish between being liable to some harm and having a liability. A liability is a reason to give less weight to defensive harms applied to the liable agent when deciding what to do. I show that this approach provides us with a clear and parsimonious account of the constraints of proportionality and necessity, and helps to clarify hard cases involving multiple aggressors and multiple justifications.

Chapter Two considers the role of choice in the justification of liabilities. I begin with T. M. Scanlon's argument for the significance of choice as a justification for self-defense. I show how this argument can be modified and elaborated to account for the constraints of necessity and proportionality. The conclusion of this argument is that culpable choices, such as the choice to attack another agent, can give the culpable person a liability. This argument also suggests, contrary to widespread assumptions, that causal responsibility for a threat is not essential for liability. I go on to consider the plausible suggestion that by choosing to impose risks on others we can make ourselves liable. I argue that this is only plausible once we have separated cases of risk imposition from cases of risk assumption. To make this distinction, we must return to the significance of choice argument.

In Chapter Three I discuss another possible justification for liabilities: Victor Tadros's suggestion that our enforceable duties explain and justify our liabilities. While I agree that our duties and liabilities are closely related, I argue that this connection is akin to what scientists call correlation but not causation. I suggest that our duties and liabilities often correlate in stringency because they are justified by the same underlying moral considerations. However, there are good reasons to

suppose that our duties do not themselves generate our liabilities. This correlation between duties and liabilities leaves open the possibility that thinking about our duties can give us an insight into our liabilities. I consider this possibility and argue that our duties give us further reason to suppose that causal responsibility for a threat of harm is not a necessary condition of liability.

Chapter Four continues this discussion of the relationship between culpability, causation, and liability. I offer a series of arguments against what I call the Causal Connection Requirement. This is the widespread assumption that a person must be causally responsible for some threat in order to be liable to be harmed in the prevention of that threat. I show how this assumption is undermined by our intuitions about proportionality, necessity, elements of the tort law, and the possibility of moral responsibility without causal responsibility. I draw attention to the fact that most cases of self-defense require preemptive harming, and that such cases can be especially problematic for the Causal Connection Requirement.

To conclude this chapter, I consider the suggestion that the Causal Connection Requirement, while occasionally unintuitive, should be retained because without it there is no way to restrict the scope of liability. I reply to this objection by arguing that culpability itself can restrict the scope of liability, at least when culpability is the justification for the liability. We can see this by looking more closely at the nature of culpability, and the way it changes over time.

In the remainder of the thesis, Chapters Five and Six, I develop an account of the kind of culpability that justifies liability. This culpability, I argue, is a measure of a person's sensitivity to action-guiding moral reasons in the exercise of their agency. A person is culpable when they form a subjectively wrongful intention, by which I mean an intention that manifests insufficient sensitivity to action-guiding moral reasons. In Chapter Five I lay out the basic structure of this account and discuss the relative importance of actions and intentions, the relationship between culpability and

permissibility, the problem of moral luck, and an interesting set of excuses that do not deny culpability.

In Chapter Six, I discuss what it means for an intention to manifest insufficient sensitivity to action-guiding moral reasons. I argue that this depends both on the content of the intention and on the way in which the intention was formed. In particular, it depends on how wrongful the intended action would be, from the belief and evidence-relative perspectives, and on whether the person would still have formed the intention if they cared enough about the relevant morally significant considerations. I show how this account offers a promising middle ground between volitionist accounts of culpability, which focus on our choices, and attributivist accounts of culpability, which also focus on non-voluntary states such as our desires and moral beliefs. I conclude by showing how my account provides a good description of our culpability for actions we have not yet performed, and how it can address hard problems raised by cases of moral ignorance.

The thesis, and each chapter, ends with a brief conclusion.

Chapter One:

What are Moral Liabilities?

Introduction

In recent years the concept of moral liability has become widely used. It is central to the way many now understand the permissibility of defensive and compensatory harming. Interesting work has been done on the necessary and sufficient conditions of liability, and on the constraints thought to restrict it.¹ Most of this work has been undertaken in the tradition of applied ethics, with an interest in the nuances of practical cases. Thus, while the applications of the idea of moral liability are widespread and increasingly sophisticated, our theory of the nature of moral liability has remained somewhat primitive. In this chapter I identify two ways in which our current understanding of moral liability might be improved.

First, it is often taken for granted that the idea of moral liability is inseparable from the forfeiture of rights. For many authors, this is simply what moral liability is. When an aggressor culpably attacks a victim, we suppose that the aggressor forfeits certain of her rights.² This is the basic structure of most liability based justifications for defensive harm. While liabilities may mark an exception to moral protections such as rights, this is not all that liabilities do. I argue that we should not define liabilities simply in terms of the forfeiture of rights.

Second, I argue that we must distinguish between two distinct roles that claims about liability can play in our deliberations. Too often, liability is simply a conclusion of deliberations about what it

¹ See, for example: Jeff McMahan, "The Basis of Moral Liability to Defensive Killing," *Philosophical Issues* 15, 1 (2005): 386–405. Helen Frowe, *Defensive Killing: An Essay on War and Self-Defence* (OUP, 2014), Jonathan Quong, *The Morality of Defensive Force* (Forthcoming), and essays in Christian Coons and Michael Weber eds. *The Ethics of Self-Defense* (OUP, 2016).

² McMahan, "The Basis of Liability," 386. This approach begins with Judith Jarvis Thomson. See *The Realm of Rights* (Harvard University Press, 1990) and "Self-Defense" *Philosophy and Public Affairs* 20, 4 (1991): 283-310.

is permissible to do. Used in this way, claims about liability merely summarize, but cannot explain or justify, our conclusions. The concept of moral liability is more useful when it is treated as an input into deliberations about what we should do.

Therefore, rather than treating liability as the mere absence or forfeiture of rights, we should treat it as the presence of a distinctive *pro tanto* moral consideration that counts against the moral significance of defensive harms applied to a person. We can call this distinctive moral consideration a liability. We should distinguish between the liabilities a person has, those considerations that make him morally susceptible to bear costs in general, and how much harm he is liable to bear in a given situation. We can better understand cases of self-defence, and constraints such as proportionality and necessity, by developing a theory of liabilities.

In the first half of this chapter I elaborate on these two ways in which our understanding of liability might be improved. In section one, I argue that it is a mistake to simply define liability as rights forfeiture. In section two, I argue that we should distinguish between the judgment that a person is liable to some harm, and the claim that a person has a liability that might make him liable in some circumstances. I go on, in the second half of the chapter, to sketch a theory of moral liabilities and to show the advantages of this approach in explaining the constraints of proportionality and necessity. This approach is particularly useful for clarifying otherwise complex cases of aggregation or multiple justifications.

1. Beyond the Forfeiture Definition of Liability

Moral liability is the dominant concept used to explain the ethics of self-defense and related forms of permissible defensive harming. The recent interest in cases of permissible defensive harming arose, and has been mainly pursued, within a broadly deontic and rights based tradition. It is understandable, therefore, that liability has consistently been defined in terms of rights. Moral

liability could be said to have developed as a tool for explaining cases of defensive harming within a framework of rights.

However, I believe that despite this history, and even because of it, we should not *define* moral liability as the absence or forfeiture of rights. Those who use the concept of liability alongside the concept of rights will need some account of the relationship between these two. Which account works best will depend as much on liability as on the theory of rights that we prefer. This is one way of understanding the debate between specificationist and forfeiture theories of liability. On the former, rights are not the sort of things that come and go, but can be highly specified to account for the vagaries of proportionality and necessity. On the latter, rights are the sort of things that come and go, and do not need to be highly specified. The conflict here is as much about the nature of rights as it is about liability.

We should not confuse an account of the relationship between rights and liability with an account of what liability is. I believe that we should be open to the possibility of other relationships between liability and rights, including no relationship at all. This would make liability based justifications accessible to those who do not wish to make rights central to their treatment of cases of permissible defensive harming.

It would also allow us to see moral liability, for the first time, not as an absence or lack of something but as a substantive moral consideration. This is closer to the way the concept of liability is used in other contexts. In finance, for example, a liability is not merely the absence of a protection against losses, it is the presence of a particular susceptibility to suffer losses. As I argue below, we can make progress in our understanding of liability by treating it as a type of moral reason: a substantive moral consideration in its own right.

Simply defining liability as the forfeiture of rights does nothing, on its own, to justify our claims about liability. This point has been made elsewhere, in an excellent paper by Massimo Renzo.³

The point can be made using a simple case:

Self-Defense: A person, Villain, culpably attacks another person, Victim. Victim can defend herself but only by killing Villain. If Victim does nothing she will be killed by Villain.

Most, I will assume, have the intuition that it is permissible for Victim to kill Villain in self-defense.

Seen in a certain light, this conclusion is surprising because in most other contexts it is impermissible for us to kill others merely for our own benefit. This is what makes cases of self-defense interesting.

We want to know more about why it is permissible to kill the Villain in this case.

Perhaps the most popular answer to this question is that it is permissible to kill Villain because she has made herself liable to be killed to save Victim. However, as Renzo points out, this is clearly more of a placeholder than an explanation.⁴ This kind of placeholder can be useful, because it gives us a word to refer to the moral phenomenon we are interested in. Instead of asking “what is it that separates Villain (and other similar people) from innocent people” we can ask “what makes Villain liable in this case” or “what are the necessary and sufficient conditions for moral liability.” Despite being useful, this kind of placeholder is not the same as a justification.

For McMahan, the claim that Villain is liable to be killed is another way of saying that Villain does not have a right not to be killed and Victim does not have a duty to refrain from killing Villain.⁵ This is a more precise way of describing our intuitive conclusion about this case, this time using Hofeldian claim rights.⁶ Such a description is essential if we want to talk about the relationship between liability and Hofeldian claims. Thus, my point is not that this description of liability is false.

³ Massimo Renzo, “Rights Forfeiture and Liability to Harm,” *The Journal of Political Philosophy* 25, 3 (2017) 324-342.

⁴ Renzo, “Rights Forfeiture and Liability to Harm,” 335.

⁵ In conversation.

⁶ For a classic account of such rights see Thomson, *The Realm of Rights*, part one.

My point is just that this is not an account of what liability is. Nor is it a description that has much explanatory power. We are no closer to knowing why Villain has lost this right, why Victim lacks this duty, or why Victim's defense must be proportionate and necessary. Ultimately, these are the further facts that we need to investigate, in order to understand liability.

I concede that liability often involves an exception to some of the moral protections that people are usually thought to enjoy. A liable person may be deliberately harmed, either as a side-effect or as a means, despite the fact that harming others in this way is usually impermissible. This is what would happen in *Self-Defense* if Villain was killed to save Victim. To this extent, if these usual protections are described as rights, liability may involve an exception to these rights.

Yet, liability can mark an exception to our usual rights, on some theories, without being defined as the absence or forfeiture of rights. There are other possible relationships between rights and liability.

Moreover, rights may not be the only moral consideration to which liabilities mark an exception. Consider,

Self-Defense II: Two villains attack an innocent victim. The victim will be killed unless both villains are killed in his defense.

In this case, it is permissible for the victim to kill both villains in self-defense. The villains, we could say, are liable to be killed. This marks an exception to the usual protections people enjoy against deliberate harming. Yet it also marks an exception to the common moral assumption that we should minimize the amount of harm that occurs overall. For many theorists, rights are an exception to this common assumption; liability seems to be another. Since liability does more than make an exception to our usual rights, it is a mistake to define liability as the absence or forfeiture of rights.

There may be some theories that can explain this background assumption in terms of rights. Perhaps the villains, in this case, each have a right not to be killed in a way that brings about more harm overall. I do not mean to deny this possibility. However, not everyone who thinks and writes

about liability must understand rights in this way. Many may understand rights as an exception to the background assumption that we should minimize the amount of harm that occurs overall. For these people, liability is more than an exception to our usual rights.

Indeed, it is possible to use the concept of moral liability without using the concept of rights. One could claim that it is perfectly permissible to deliberately harm others, and that all that matters is minimizing the amount of harm that occurs overall. This would be a kind of simple consequentialism. Such a theorist might then make an exception in cases like *Self-Defense* and *Self-Defense II*. They could claim that we should always try to minimize the amount of harm that occurs overall, except in the case of those who culpably attempt to harm others. They might claim that these people are liable to be harmed, so that harms applied to them do not count when deciding which possible outcome is best. Such a view might not be correct, but it would not be conceptually mistaken. It would not be a mistaken application of the concept of liability. If liability can be used without a background theory of rights, then it is surely a mistake to define liability in terms of rights.

What then is liability, if it is not just the absence or forfeiture of rights? I turn to this question in section three, but I can give a part of my answer now. Recall the role that Renzo identifies for liability; it is a placeholder for whatever it is about people like Villain that make them vulnerable to harms in a way that Victim and innocent bystanders are not. There is something special about Villain in this case that gives us, and Victim, a reason to harm her rather than a bystander, and rather than allowing Victim to be harmed. This, I take it, is the core of what liability is. It is the moral reason we have, given what Villain has done, to prefer defensive harms to Villain, over harms to Victim or somebody else.

Notice two crucial aspects of this basic account of liability. First, it describes liability as a substantive moral consideration, not just as the absence of something else. Liability is a way of describing a certain type of moral reason. We can describe the consequences this reason has for our

conclusions about rights. This would describe a consequence or implication of liability, but is would not capture all of what liability is. We can discuss liability, and debate its deeper justification and necessary conditions, without making any reference to rights.

Second, by describing liability as a type of moral reason, I am turning liability from an output of our deliberations about what to do, into an input into our deliberations about what to do. The claim that Villain has forfeited her right not to be killed lacks explanatory power, because it is a conclusion we reach by considering what it would be necessary, proportionate, and permissible to do in this case. I believe that liability can do more than this. I think liability can explain why it would be necessary, proportionate, and permissible to kill Villain. However, distinguishing these two roles for liability, as an input and output of our deliberations, can be confusing. Making it clear requires a bit more terminology.

2. From 'Being Liable' to 'Having a Liability'

The way liability is currently discussed can be confusing because liability can be used to do several things at once. First, it is sometimes used to specify how much defensive harm can be permissibly inflicted on a person in a given situation. Used in this way, judgments of liability are an output of deliberations about what we ought to do. Second, liability can be used to explain the difference between people like Villain, who are morally susceptible to bear harms, and those innocent people who are not. Used in this way, liability becomes an input into deliberations about what we ought to do. Things get confusing, because these two roles place different demands on the way we discuss liability.

When liability is used to fulfil the first role, then a person's liability must be relative to a given situation. This is how the term liability is used in the influential work of Jeff McMahan, who considers proportionality to be internal to liability. In other words, to know how much harm an

aggressor is liable to suffer, you must first know how much harm it is proportionate to inflict. While this is a reasonable application of the concept of liability, it creates problems when the same concept is also tasked with the second role: explaining the difference between people who are morally susceptible to bear harms and those who are not.

Whatever concept explains the difference between people who are morally susceptible to bear harms (such as culpable or morally responsible people) and those who are not, must be an *input* into deliberations about proportionality. It is proportionate to inflict more harm, relative to the same harm prevented, on a culpable person when compared to an innocent person. If we asked McMahan why this is, he might be tempted to reply that it is because culpable people are liable while innocent people are not. This seems like the right reply, and this explanatory power is part of the appeal of McMahan's general approach.

Yet, as McMahan realizes, this reply is unavailable. If proportionality is internal to liability, then liability cannot also be an input into deliberations about proportionality. This leaves McMahan relying on phrases like "potentially liable" to explain the difference between culpable and innocent people.⁷ There is nothing incorrect about this, but it is unsatisfying and can be slightly confusing.

To remedy this, we can borrow a distinction from the way liability is understood in other contexts. In the legal or financial context, for example, we distinguish the general question of what liabilities a person or corporation has, from the specific question of how much that person or corporation is liable to pay in a given circumstance. Suppose I trade in stock futures and have agreed to sell some number of shares in a company for a set price at some time in the future. If the price of these shares rises unexpectedly, I will suffer a loss. We can describe this situation by concluding that my agreement to sell these shares is a liability. How much I will actually be liable to pay depends on whatever the price ends up being when I am forced to sell. In this way, a description of my liabilities

⁷ Jeff McMahan, "Proportionate Defense," *Journal of Transnational Law and Policy*, 23 (2013) 7.

is a description of my susceptibilities to suffer losses in general, while a description of how liable I am is relative to a specific situation.

This distinction is crucially important, but in the moral context it is often overlooked. The question ‘is Jack liable?’ can be ambiguous between the question ‘what moral/legal/financial liabilities does Jack have at present,’ and the question ‘how much of a cost is Jack morally/legally/financially liable to pay in a given circumstance?’ We should distinguish between the liabilities a person has, those considerations that make him morally susceptible to bear costs in general, and how much harm he is liable to bear in a given situation. Making this distinction helps clarify the way we discuss liability. The two are linguistically distinct as a person’s *liability* is a noun, while how *liable* a person is in a given circumstance is an adjective.

A person’s liabilities are crucial inputs into deliberations about how much harm that person is liable to suffer in a given situation. To begin deciding what harms a person might be liable to suffer, we first need to know about that person’s liabilities. In the remainder of this chapter I will argue that we can make progress by taking liabilities seriously. This means treating liabilities as independent moral considerations and not defining them as the forfeiture of rights.

3. What are Moral Liabilities?

I have argued that we should treat liabilities as independent moral considerations that are inputs into our deliberations about what to do. By doing this we can remain open to different possible justifications of liabilities, and we can avoid confusion between the different roles the concept of liability can play in our deliberations about what to do. This approach is warranted because it allows the concept of liability to better fulfil its role as a ‘placeholder.’ It is further supported, as I will argue in part four, by the easy way in which this approach can explain the classic constraints of proportionality and necessity.

Yet, it might be wondered, what could liability mean if it is not defined in terms of rights. I contend that the core concept of moral liability can be preserved, and more clearly revealed, when liability is made to stand on its own. As we have seen, moral liabilities make some people, like Villain, particularly morally susceptible to bear costs. It might help to think of liabilities as being in some sense the opposite of rights. While rights are *pro tanto* considerations, or reasons, that insulate us from the imposition of harms, liabilities are *pro tanto* reasons that make us particularly vulnerable to the imposition of harms. Yet, since I do not wish to defend a particular theory of rights, I consider relationship between rights and liabilities to be an open moral question.

More precisely, we could observe that a liability is a reason that reduces the weight that we should give to defensive harms applied to the liable person when we are deciding what to do. In this way, a liability is not a directly practical reason; it is not a reason for or against doing something. Certainly, it is not simply a reason to impose harms on a culpable aggressor. Rather, a Liability is an indirect or theoretical reason. It is a reason to discount the interests of the culpable aggressor when deciding which possible distribution of defensive harms we should bring about.

This approach follows a long tradition in Anglo-American and French law, according to which self-defence is a justification, grounded in the diminished importance of the interests of the culpable aggressor.⁸ In his 1978 *Rethinking Criminal Law*, George Fletcher describes this tradition writing,

[T]he Aggressor is entitled to lesser consideration in the balancing process. His interests are discounted, as it were, by the degree of his culpability. The extent to which his life is discounted determines whether the defender may use deadly force to defend against rape, serious bodily harm, loss of irreplaceable property and even, possibly, the loss of less important interests. The underlying premise is that if someone endangers the interests of another, his interests are less worthy

⁸ George Fletcher, *Rethinking Criminal Law* (OUP, 2000) 859-60. This is a reprint of the original Little, Brown edition of 1978.

of protection ... The entire approach demands sensitivity to conflicting values and to the moral significance of culpability.⁹

This, Fletcher writes, is how the Anglo-American and French legal systems understand self-defence when it functions as a justification, rather than as an excuse. It overlaps perfectly with how I have come to believe the nature of moral liability, and the moral justification for self-defence, should be understood.¹⁰

A liability is the reason we have to discount the interests of the aggressor. From this perspective, it seems clear that some liabilities are stronger than others. Highly culpable people, for example, are more morally susceptible to bear costs than minorly culpable people. To capture this, we can assert that defensive harms applied to highly culpable people are much less significant than defensive harms applied to less culpable people.

What does it mean to say that we give a consideration, such as a harm, less weight? There are many ways in which this might be understood, depending on how we understand moral deliberation. I will take a simple, and I hope ecumenical, approach. I assume that the magnitude of a harm counts against imposing that harm, even if imposing the harm helps us achieve some good. For example, we have more reason not to break an arm to save a life than we have not to break a finger to save a life.

Now suppose that the arm belongs to a culpable aggressor who threatens the life we can save, and that the finger belongs to an innocent bystander. This changes the weight we should give to these to harms. In such a case, let us assume, we should break the arm and not the finger. One way to describe this is to say that we are treating the broken arm as if it is less morally significant

⁹ Fletcher, *Rethinking Criminal Law*, 858-9. Fletcher describes this justification as a “variation of lesser evils.” In the contemporary literature ‘lesser evil’ justifications apply when the agent who can be defensively harmed does not have a liability. However the same approach, of weighing the moral significance of one set of harms against another, applies in each case.

¹⁰ I reached my conclusions independently, though no doubt influenced by the Anglo-American legal tradition which Fletcher describes and under which I have always lived.

than a broken finger. In my view, this is what liabilities do. They describe one reason we can have to count certain harms as less significant than they would otherwise be, for the purpose of making certain moral decisions.

The stronger the liability, the more we can discount those defensive harms applied to the liable person. In mathematical terms, if a broken arm is a harm of 10 and a broken finger is a harm of 1, then the liability in this case reduces the significance of defensive harms by at least ten times. I am not suggesting that these numbers are correct or useful for solving real cases. Indeed, I doubt whether we can ever expect mathematical precision in ethics. However, as I will show in part five, it is sometimes helpful to use mathematical language for its clarity, even if we do not expect its precision.

It might appear that liability judgments are inherently comparative, so we cannot know the force of a person's liability until we know the context. Consider,

Two Villains: Two people are each culpable responsible for creating a substantial threat of harm to Victim. Both people are culpable for this, and we can assume it would be proportionate to impose the defensive harm necessary to prevent the threat on either. However, one is far more culpable than the other.

If the defensive harm needed to prevent the threat cannot be divided, then it ought to befall the more culpable person.¹¹ Given the context, the less culpable person is not liable to any harm at all. Yet if the more culpable person had not been present, the less culpable person would have been liable to suffer the defensive harm. This sort of case could lead one to believe that the strength of a person's liability is relative to the situation.

This reasoning confuses the distinction between liabilities and being liable. It is true that how much harm a person ends up being liable to suffer will depend on the context, and in particular on the range of defensive options available. When harming the more culpable person is an option,

¹¹ McMahan makes this point about a similar case in "The Basis of Moral Liability," 394.

the less culpable person will not be liable to be harmed. This does not mean that the less culpable person's liability has become weaker or disappeared. It is simply because, in this specific context, the less culpable person's liability is not powerful enough to make that person liable to harm.

An analogy with duties makes this clearer. Suppose that several people are drowning, and you can either rescue a lone person, or a pair of different people. Each person has a claim to be rescued, and you have a *pro tanto* duty to rescue each. However, given the circumstances, you ought to rescue the pair, and therefore cannot rescue the one. This does not mean that your duty to rescue the lone person has weakened, or has disappeared. It simply means that, in this specific context, your duty to rescue the lone person has been outweighed.

In the rescue case, as in the liability case, we can see the importance of separating the inputs from the outputs of deliberation. Because these are both forced choice cases, some of the inputs will not be discernible in the outputs. In the rescue case, for example, the way you ought to behave is identical to the way you ought to behave if the lone person did not exist. It is a mistake, however, to suppose from this that your duty to rescue the lone person has no force in this case. We need to posit this duty as an input in order to make sense of your deliberations.

The same is true with liabilities. Both culpable people in *Two Villains* have liabilities, and the strength of these liabilities is determined, perhaps amongst other things, by their culpability. Depending on the circumstances, some people with liabilities will not be liable to be harmed. Yet, as I will argue below, we must still posit these liabilities in order to make sense of our deliberations about what to do.

Liabilities are complicated and important moral considerations about which we still know very little. The situation is akin to having no theory of rights in general, despite having detailed views of who has an all-things-considered right to what in specific situations. If we take liabilities seriously, then we see that there is much we have yet to discover.

4. Proportionality and Necessity

The role for liabilities I have proposed is broadly supported by the classic constraints of proportionality and necessity. By focusing on liabilities we can offer a clear and parsimonious account of these constraints. Understood as independent considerations, or when defined in terms of rights, the complexity of proportionality and necessity can be daunting.¹² Yet we can simplify these constraints, and explain their underlying logic, by focusing on the role of liabilities.

Any defensive harming can be understood as a trade of one set of harms for another. If we allow a villain to harm a victim, a certain distribution of harms will result. We can bring about a different distribution of harms if we intervene, by harming the villain to save the victim. There are many moral considerations that can make one distribution of harms better than another. Liabilities are one such consideration. They are considerations that make defensive harms less significant than they would otherwise be. If some of the people we could inflict defensive harms on have liabilities, then defensive harms imposed on those people count for less than harms of comparable magnitude imposed on innocent people.

Once liabilities are taken into account, we can understand proportionality as the requirement that we only accept good trades. Consider the most basic sort of proportionality judgment, that of narrow proportionality.¹³ I will follow McMahan in assuming that proportionality is a matter of the relation between three variables: the harm inflicted on the ‘potentially liable’ person, the harm to others thereby prevented, and the degree of the ‘potentially liable’ person’s culpability.¹⁴ The greater

¹² For a discussion of the three kinds of proportionality see McMahan, “Proportionate Defense” and McMahan “Liability, Proportionality, and the Number of Aggressors,” in *The Ethics of War*, ed. by Saba Bazargan and Samuel Rickless, (OUP, 2017) 3-28. For an overview of necessity see Seth Lazar, “Necessity in Self-Defense and War,” *Philosophy and Public Affairs*, 40 (2012) 3-44.

¹³ McMahan, “Proportionate Defense,” 7.

¹⁴ For a rival approach that focuses on the strength of the right that is threatened, rather than on the degree of the threatener’s culpability, see Jonathan Quong *The Morality of Defensive Force* (Forthcoming).

the person's culpability, the more harm that can proportionally be inflicted relative to the same harm prevented.

This result is explained by liabilities. The stronger the liability, the more harm that can permissibly be inflicted relative to the same harm prevented. This is because the extent to which the significance of the harm inflicted is discounted depends on the strength of the liability. However, because defensive harms inflicted on the person with the liability are merely reduced in significance, not ignored, there will always be some ratio of harm inflicted to harm prevented beyond which the defensive harm is morally worse than the harm it could prevent, even after it has been discounted. Defensive actions are disproportionate when the harms they impose are still morally worse than the harms they prevent, even after all the relevant liabilities have been taken into account.

The same is true of wide proportionality, although different considerations will be relevant. Wide proportionality measures whether the harm a defensive action imposes on innocent bystanders is proportionate to the harm it prevents.¹⁵ Wide proportionality asks whether the harms a defensive action would inflict on bystanders is morally preferable to the harm that would otherwise occur, even after any relevant rights of those bystanders have been taken into account.

Having different measures of proportionality may be practically useful, as a way of structuring decision making, but it is unnecessary in theory. All we need is the fact that liabilities make some harms less morally significant than others, and the requirement that we never bring about a set of harms that is worse than that which would otherwise occur. Together these considerations generate all the intuitions that are traditionally explained by positing several separate proportionality constraints. This simplification also means that it is much easier to understand the relationship between proportionality and the constraint of necessity, which is also thought to restrict the permissibility of defensive harming.

¹⁵ McMahan, "Proportionate Defense," 7.

The constraint of necessity picks up where proportionality leaves off. If there is more than one trade available, then the requirement that we only make good trades is not sufficient. Some good trades are much better than others. The constraint of necessity is the requirement that we make the best trade available.¹⁶ It means that we must choose the option that brings about the best distribution of harms available, once all the relevant rights and liabilities have been taken into account. Consider:

Deadly Pincher: A villain has attached a poisoned prick to her thumb and is now culpably attempting to kill an innocent victim. If the villain succeeds, the victim will suffer a hard pinch followed by death. You can intervene to save the victim by either, a) removing the poisoned prick, in which case the victim will still suffer a strong pinch but will not die, or b) killing the villain, in which case the victim will be unharmed.¹⁷

In this case, it seems to me that option (a) is correct. Under the circumstances, (b) would be deeply wrong. This is despite the fact that option (b) is proportionate, since killing the villain is better than allowing the victim to die, and is necessary to save the victim from the *entirety* of the threatened harm.¹⁸

In *Deadly Pincher* we can see how similar proportionality and necessity are, and how both constraints are captured by focusing on liabilities. The reason (a) and (b) are both proportionate is that both bring about harms that are better than the harms they prevent, once the villain's liability has been taken into account. The reason (a) satisfies necessity and (b) does not, is that (b) brings about a distribution of harms worse than (a).¹⁹ Harms applied to the villain can be heavily discounted because of his liability. However, the harm of death is so much greater than the harm of a strong pinch that it is better for the innocent victim to suffer a strong pinch than it is for the

¹⁶ The understanding of necessity I am appealing to here is a simplified version of that defended by Lazar in "Necessity in Self-Defense and War."

¹⁷ This is a variation on a case used by McMahan.

¹⁸ Lazar makes a similar point about a different case in "Necessity in Self-Defense and War," 13.

¹⁹ This is consistent with the approach to necessity advocated by Lazar in "Necessity in Self-Defense and War."

culpable aggressor to be killed. The villain's liability merely discounts, but does not exclude, the significance of defensive harms. It is for this reason that (a) brings about a better result than (b).

In some ways, the constraint of necessity is simply a further application of the constraint of proportionality. We can tell that (a) is better than (b) by isolating the difference between the two, and asking whether that difference would be proportionate. Suppose that (b) consisted of (b1) removing the poisoned prick to save the victim's life, followed by (b2) killing the villain to prevent the hard pinch. We can see that (b1) is identical to (a), so that the difference between (a) and (b) is the addition of (b2). Taken on its own, (b2) is disproportionate. Thus, we can conclude that choosing (b) over (a) would violate necessity, because it would bring about a result that is much worse than another available option. A defensive action violates necessity if the difference between that action and any other available option is itself disproportionate.²⁰

Ultimately, narrow proportionality, wide proportionality, and necessity, are all just different ways of measuring the same thing. All these considerations assess whether a defensive action brings about the best distribution of harms available, once all the relevant liabilities have been taken into account.

These constraints only produce different results because they exclude various things from consideration. For example, judgments of narrow proportionality exclude from consideration harms applied to people other than a specific liable person. Wide proportionality excludes harms applied to people other than bystanders. All types of proportionality exclude consideration of all other options besides doing nothing. While these constraints limit the options and harms they consider in various way, the question they all ask is the same: does this defensive option bring about the morally best distribution of harms available?

²⁰ McMahan recognizes this relationship in "The Limits of Self-Defense," in Christian Coons and Michael Weber, eds. *The Ethics of Self-Defense* (OUP, 2016).

In this way, by treating liabilities as independent moral considerations, rather than defining them as the absence or forfeiture of rights, we can offer a clear and straightforward account of liabilities, proportionality, necessity, and the relationship between the two. On this approach we can use liabilities to explain proportionality and necessity. If we can then agree on a justification for liabilities, then we will also have explained proportionality and necessity. This shows how useful liabilities can be as a placeholder, when they are not defined in terms of rights. We can remain open to the possibility that a theory of rights can provide the best justification for liabilities, without including rights in our definition of moral liability.

4.1 Can Liabilities Explain Compensation?

The concept of liability has been used to distinguish between cases where a person is harmed in a way that requires compensation, and other cases where a person is harmed in a way that does not require compensation. Suppose we believe that:

- a) If a person suffers a harm to which they are liable, then there is no *prima facie* case for compensating that person for that harm.

This claim may be defended in terms of rights forfeiture, if we further believe that:

- b) If a person suffers a harm that they have forfeited their right not to suffer, then there is no *prima facie* case for compensating that person for that harm.

This claim, or something like it, lies behind much current thinking about liability. It might be objected that if liability is no longer defined as the forfeiture of rights, then (b) can no longer be used to defend (a).

We can meet this objection by showing that liabilities work equally well as part of an account of compensation. Like a case of defensive harming, any case of compensation can be treated as a trade of one set of harms for another. We take something from the person who pays compensation, and give it to the person who ought to receive compensation. As in cases of defensive harming,

liabilities are a consideration that matter when determining whether a given amount of compensation makes a good trade.

Suppose a villain culpably harms a victim. We can suppose that this culpable attack gives the villain a liability. Thus, when considering a trade of harms between the villain and the victim, we should discount those harms applied to the villain to protect the victim from the threatened harm. As we have seen, this discounting can help explain self-defense. The same explanation can also work for compensation. If we force the villain to compensate the victim, then we are harming the villain to protect against, or in this case remedy, harm suffered by the victim. This is permissible because the harm that would be suffered by the villain, if he were forced to pay compensation, is less morally bad than the harm that the victim would suffer if she were not compensated.

By contrast, if the villain is permissibly harmed by the victim in self-defense, then the villain lacks a case for compensation from the victim. Assuming the defensive action was necessary and proportionate, then the defensive action made a good moral trade. There is thus no *prima facie* case for reversing this trade through compensation. In this way, liabilities can explain compensation, in the same way they can explain self-defense, without assuming a connection between liability and rights.

This not meant to be an account of compensation. My point here is simply to show that liabilities can plausibly be included in an account of compensation. For such an account to be compelling there are a host of details that must be further explored. We need to know if liabilities include a threshold magnitude of harm beyond which they do not apply. We might also ask how liabilities apply over time, whether, and in what way, causation matters for liabilities, and how liabilities interact with other deontic considerations like the distinction between doing and allowing. Most of all, we want to know how strong liabilities are, and what further moral theory best justifies liabilities.

5. Proportionality in Aggregation

A liability reduces, but does not itself eliminate, the moral significance of defensive harms applied to the liable person. This kind of moral weighting already underlies our thinking about proportionality and necessity. Focusing on liabilities, rather than on rights forfeiture, brings this to the surface. It simplifies and clarifies our treatment of these cases. This clarification can help us understand more complex cases, such as those involving aggregation and multiple justifications.

One consequence of the approach I have been defending is that there is always some *pro tanto* moral reason that counts against imposing harms on even highly culpable people. This is because liabilities merely reduce the significance of harms applied to liable people. This is not a by-product or unexpected consequence of my account. Rather, it is central to the way I think liabilities work. Nor should it be, in my view, a controversial claim. The nature of pain, suffering, and other harms is such that we always have a moral reason to avoid inflicting it. This reason can, of course, be outweighed. Moreover, as liabilities show, there is hardly ever a clear correlation between the magnitude of harms and the strength of the moral reasons we have to avoid them.

There are, however, some who may consider this feature of my view to be disqualifying since it implies that there is some magnitude of harm applied to some number of highly culpable people that is morally worse than even the harm of death applied to an innocent person. Here is the sort of case about which we might disagree:

Many Murderers: A victim is attacked by a large number of culpable threateners. There are only two possible outcomes, a) kill all the would be murderers, or b) allow the victim to be killed.

This is, in my view, a difficult case. In most realistic cases (a) will be preferable since the harms applied to highly culpable people can be heavily discounted. Yet I suspect that it remains theoretically possible that there will be some number of murderers so great that we are required to

choose (b). For example, it would not be proportionate to kill thousands of culpable soldiers on the unjust side of a war, simply to save one innocent life. Some may disagree. They might hold that regardless of the number of culpable aggressors we should always choose (a).²¹ Others prefer a more nuanced answer. They might claim that whether the number of threateners matters depends on just how culpable they are.²²

To get to the bottom of the disagreement over *Many Murderers* there are a series of distinct questions that need to be distinguished. First, we should set aside the question of aggregation, and first consider whether there is some magnitude of harm so great that it cannot be proportionally inflicted on a single highly culpable person to prevent the harm of death befalling an innocent person. To me, at least, the answer to this is plainly yes. This follows from widely held intuitions about proportionality.

I am assuming here that it is wrong to kill even a highly culpable person to prevent a strong pinch. If this is the case, then it should similarly be wrong to inflict some great harm on a highly culpable person to prevent an innocent death whenever that great harm is as much worse than the harm of death as death is worse than a strong pinch. Clearly, this will mean that the great harm must be very very bad. Perhaps in practice this will be impossible. At the level of theory, however, it seems that there must be some level of harm beyond which it is better to allow the innocent person to die.

The *pro tanto* badness of defensive harms applied to culpable people is implicit in the constraint of proportionality. This constraint, after all, functions to limit the permissibility of applying excessively large harms to even highly culpable people. It can make it impermissible to impose a disproportionate defensive harm on a highly culpable person, even if this is the only way to

²¹ For example, Francis Kamm, *Ethics for Enemies: Terror, Torture, and War* (OUP, 2011).

²² This view is suggested by McMahan in "Liability, Proportionality, and the Number of Aggressors," in Saba Bazargan and Samuel Rickless, eds., *The Ethics of War* (OUP, 2017), and David Rodin, "Justifying Harm," *Ethics* 122 (2011) 74-119.

prevent an unjust harm from befalling an entirely innocent person. I see no clear way to explain this constraint without positing that harming culpable people is something to be avoided.

It might be argued that a sufficiently large harm to an innocent person, such as the harm of death, cannot be outweighed by any magnitude of harm to a culpable person. Such a view would not be contradictory. It would, however, be a repudiation of the constraint of proportionality in high stakes cases. I cannot disprove such a view, although I consider it to be somewhat arbitrary and, in practice, excessively cruel.

If this is agreed, what else could drive the view that the number of murderers in *Many Murderers* is irrelevant? The answer seems to be the suggestion, first made by Francis Kamm, that proportionality does not aggregate harms, but rather proceeds by “a series of pair-wise comparisons.”²³ Thus, since it would be permissible to kill any one of the many murderers to save the life of the victim, it is permissible to kill each of the many murderers no matter how many there are. This is, in my view, a mistake. In the next section I argue, following McMahan, that narrow proportionality cannot tell us how to solve *Many Murderers*. I go on to argue that not aggregating the claims of the threateners in *Many Murderers* may be plausible, but not for the reasons that are usually proposed.

5.1 Aggregation and Narrow Proportionality

As I showed in parts three and four, focusing on liabilities allows us to see clearly the underlying structure of proportionality and necessity. These are not independent constraints on liability, rather they are different ways of measuring the same thing. They are different ways of measuring whether a given defensive action makes a good trade, once all the relevant rights and liabilities have been taken into account. These measures differ because they allow us to focus on

²³ Kamm, *Ethics for Enemies*, 133-4. Quoted in McMahan, “Liability Proportionality and the Number of Aggressors.”

different ways in which one trade might be better than another, by excluding some factors from consideration. For example, narrow proportionality asks if a defensive action makes a good trade compared with doing nothing, when all harms other than those applied to the victim and a single liable person are set aside.

It is true, in this sense, that narrow proportionality involves a series of pair-wise comparisons. However, this does not show that the underlying justification for defensive harming relies on a series of pair-wise comparisons. Narrow proportionality is limited to pair-wise comparisons because it is a pair-wise measure. This is a feature of the way we analyse these cases, not of the underlying justification.

If we look at the underlying justification, we can see that more than pair-wise comparisons are involved. For example, in cases of necessity it matters not only whether a defensive harm is narrowly proportionate, but also whether it is the best narrowly proportionate option available. The same underlying principles are at work in cases of necessity as in cases of proportionality. We discount defensive harms applied to people with liabilities, and then decide which possible distributions of harms we should bring about. The fact that some defensive action is narrowly proportionate is not sufficient to show that it is permissible. Even if we assume that killing each of the many murders is narrowly proportionate, this would not establish that killing them is permissible.

Yet, we have good reason to doubt even this assumption. If we are making pair-wise judgments of narrow proportionality, we must compare the harm that could be imposed, the harm it can prevent, and the strength of the would-be murderer's liability. Consider the first would-be murderer. How much harm can we prevent by killing him? Not much. Killing him might buy the victim a few seconds of life before the next murderer arrives, and perhaps not even that. Thus, if we are making strictly pair-wise comparisons we should conclude that it is narrowly disproportionate to

kill each of the many murderers, except perhaps the last. Killing no other one of them would, on its own, do enough good to be justifiable.

At best, the narrow proportionality of killing each murderer is indeterminate until we know whether the rest of them will be killed.²⁴ If we know that the rest of them will be killed, then killing one will be narrowly proportionate. However, as McMahan correctly points out, this does not mean that killing all of them can make it permissible to kill each.²⁵ We must consider whether killing all the murders is proportionate to saving the life of the Victim. This is what McMahan calls proportionality in the aggregate.²⁶

Asking whether we can justify killing each of the murderers is not the right question. We cannot know this until we know whether we can justify killing all the murderers. If we ask whether we can justify killing all the murderers, then it is relevant to consider whether all the murderers together can object. Once again, we are forced to move beyond considerations of narrow proportionality, and consider questions of aggregation.

5.2 Aggregation and Liabilities

To understand cases like *Many Murderers* we must consider whether all the harms to the murderers, taken together, outweigh the good of saving one life, once all the relevant liabilities have been taken into account. However, this does not mean that we must simply add all the harms applied to the murders together. Aggregating harms is a difficult and complicated business. I believe that the same considerations that apply to aggregating harms in other contexts, apply in cases involving liabilities. We should consider two different respects in which aggregating harms can require more than simple addition.

²⁴ McMahan, "Liability, Proportionality, and the Number of Aggressors," section two.

²⁵ McMahan, "Liability, Proportionality, and the Number of Aggressors," section two.

²⁶ McMahan, "Liability, Proportionality, and the Number of Aggressors," section two.

5.2.1 The Distribution of Harms

The first is that the distribution of a given quantity of harm can affect its desirability.

Suppose that two broken wrists are inevitable. We can either break both of one person's wrists at once, or we can break one wrist of one person and one of another. These options are not morally equivalent. Other things being equal, it is better to spread the harm between the two people. The same is true in cases involving liabilities. Other things being equal, it is better to divide a harm between two culpable aggressors, rather than impose it all on one aggressor.

This has implications for cases such as *Many Murderers*. For simplicity we can use numbers to represent the amount of harm involved in different defensive actions. Suppose that death is a harm of 100, and that we can discount defensive harms applied to a culpable aggressor by a factor of ten. These numbers are arbitrary, and are simply meant to illustrate a general point. Now suppose that a culpable aggressor threatens to kill a victim. The aggressor's liability means that it is proportionate to impose a harm of up to 1000, if this is the only way to save the victim's life. If this is correct, does it also mean that it is only proportionate to kill up to ten culpable aggressors to save the victim's life?

The answer to this question is probably not. If we had a choice between saving the victim by imposing a harm of 100 on one aggressor, or a harm of 10 on ten aggressors, we should clearly prefer the latter option. As in other contexts, spreading the harm out makes it much more morally attractive. Since it is more morally attractive, it ought to be easier to justify. Thus, if we have a choice between saving the victim by imposing a harm of 100 on one aggressor, or a harm of 10 on eleven aggressors, it may still be better to choose the latter option. Spreading out the harm makes it possible to justify a greater total harm.

Returning to our earlier question, if it is proportionate to impose a harm of 1000 on one aggressor to save the life of a victim, it may be proportionate in the aggregate to impose a total of

more than 1000 if this is spread out between many aggressors. For example, it might be proportionate to impose a harm of 600 on each of two aggressors. The more the harm is divided, the stronger this effect will become. Thus, if it is proportionate to impose ten times the harm of death on one aggressor to save a single victim, it ought to be proportionate in the aggregate to impose the harm of death on more than ten culpable aggressors. This can go some way towards explaining our intuitions in cases like *Many Murderers*.

5.2.2 The ‘Relevance’ of Harms

There is a second respect in which aggregating harms can require more than simple addition. This is the intuitive suggestion that some harms can be too small to be counted when much larger harms are at stake. Suppose we have a choice between saving one life, or preventing some number of temporary headaches. Many people have the intuition that we should save the life no matter how many headaches can be prevented. Scanlon explains this intuition by arguing that the harm of a headache is simply too small to be ‘relevant’ when compared to saving a life.²⁷ How relevance works in more complicated cases has proved to be very difficult.²⁸ Despite these difficulties, let us assume that some version of this relevance view, or something like it, is correct. There may be some threshold below which a harm is too small for any number of those harms to outweigh the harm of death.

This has important implications for cases of self-defence, once we consider the role of liabilities. I have suggested that liabilities reduce the significance of defensive harms applied to liable people. For example, if saving a victim from a culpable aggressor requires breaking that aggressor’s arm we should give that broken arm less weight than we otherwise would when deciding what to do.

²⁷ T. M. Scanlon, *What We Owe to Each Other* (Harvard, 1998) 239-40. Quoted in Patrick Tomlin “On Limited Aggregation,” *Philosophy and Public Affairs* 45, 3 (2017) 232-260.

²⁸ See Patrick Tomlin “On Limited Aggregation.”

We might give it the consideration that we would normally give to a broken finger. If saving the victim requires us to kill an aggressor, we should consider killing the aggressor to be morally equivalent to much less than the harm of death.

I propose that this moral weighting should be taken into account when deciding whether harms applied to some number of liable people are ‘relevant’ in Scanlon’s sense. After all, the relevance view is generally elaborated not in terms of magnitudes of harm, but in terms of the strength of claims. The very core of a liability is that it reduces the strength of the claim that liable people can make against certain kinds of defensive harms.

Suppose for the sake of argument that, because of his liability, killing a highly culpable attacker is morally equivalent to allowing an innocent person to lose an arm. Other things being equal, this might tell us that killing a highly culpable attacker is permissible if this is the only way to stop him from cutting off a victim’s arm. This judgement also has implications for cases involving aggregation.

We must ask whether the harm of losing an arm is substantial enough to be relevant when compared to the harm of death. Setting aside cases of liability, this is a reasonable question. If we can rescue one person from death, or a very large number of people from losing an arm, it is unclear, to me at least, whether we need to know how many arms we can save. I am tempted by the thought that saving an arm, while morally significant, is just not significant enough to be relevant when compared with saving a life.

We need not reach a clear judgement in this case. My point is simply that, given the assumptions I made earlier, whatever we say about aggregating lost arms in this case, we should also say about killing culpable attackers in cases like *Many Murderers*. Indeed, I think that *Many Murderers* is a difficult case precisely because it is unclear whether the morally weighted harm of killing a culpable aggressor is significant enough to be relevant when compared to allowing the victim to die. This case

is difficult because it is always difficult to compare a large number of less significant harms to a few highly significant harms.

Recently, McMahan and David Rodin have both suggested a seemingly contradictory position on the question of proportionality in the aggregate.²⁹ They both propose that defensive harms applied to culpable threateners should not be aggregated, but that defensive harms applied to non-culpable but morally responsible threateners should be aggregated. It is not immediately clear how McMahan can maintain this position, since he supposes that moral responsibility, and not culpability, is the basis of liability.³⁰

However, once we think in terms of liabilities, we can find a justification for this proposal. It is plausible to suppose that highly culpable threateners have much stronger liabilities than merely responsible threateners. Thus, the harm of death applied to a culpable threatener is morally equivalent to a much smaller harm than the harm of death applied to a merely responsible threatener. McMahan and Rodin appear to suppose that the threshold for relevance falls somewhere between these two points. This is a reasonable claim to make and it can be explained by liabilities.

Once liabilities are taken into account, proportionality in the aggregate should be understood as simply another application of the principles we accept in other cases of aggregating relatively insignificant harms and comparing them with more significant harms. These are very difficult cases, so it is not surprising that they are difficult in this context as well. While there is no consensus view of how a 'relevance' threshold could be made to work, and many obstacles for such a view to overcome, there is considerable convergence around the intuition that some such threshold of 'relevance' is need. As Larry Temkin writes, "virtually all agree that, other things equal, it would be worse if fifty people suffered from AIDS, quadriplegia, severe psychosis, or being deaf, dumb, and

²⁹ McMahan, "Liability, Proportionality, and the Number of Aggressors" and Rodin, "Justifying Harm," *Ethics* 122 (2011) 74-119.

³⁰ In conversation, McMahan suggests that he is uncommitted on the subject of aggregation.

blind, than if virtually any number of people suffered from a minor nosebleed, a slight cold, a sprained finger, or a short mild headache.”³¹

I will not attempt to justify or elaborate a ‘relevance’ view here. My claim is just that such a view is intuitively plausible and, if true, ought to take liabilities into account. One possible nuance is that claims of different strengths can be more or less relevant to each other. Claims that are simply too weak to be relevant might be ruled out completely, and not aggregated, while claims that are only partially relevant might be aggregated in a substantially limited way. Such a multi-level view of relevance would add a further nuance to our treatment of proportionality in aggregation cases.

Once again, the key advantage of liabilities is that they allow us to explain our diverse and complex intuitions about self-defence using a very few principles. In fact, we can explain our complex intuitions about cases like *Many Murderers* without positing any additional principles at all. Once liabilities have been taken into account, these cases can be explained as a further application of principles for aggregating competing claims in other contexts. This clear and parsimonious approach to liability is particularly valuable when we turn to the much more complex cases of multiple justifications.

6. Multiple Justifications

The key advantage of a liabilities based approach to proportionality and necessity is the clarity and parsimony that it achieves. This is valuable in its own right, but it is also valuable because it helps us when we turn to complex cases of multiple justifications. In some cases we may have both narrowly and widely proportionate defensive options. As they stand, the forfeiture based views of liability can give us no clear guidance in such a case. Consider,

³¹ Larry Temkin, *Rethinking the Good* (OUP, 2012) p. 33. This is quoted by Tomlin in “On Limited Aggregation.”

Multiple Justifications: Villain has culpably attacked Victim. You can defend Victim by either imposing a large harm on Villain or imposing a much smaller harm on an innocent bystander. Either option would be permissible if it were your only choice.

We can assume in this case that harming the Villain would be narrowly proportionate and that harming the bystander would be widely proportionate. What should you do?

One tempting proposal might be to adopt a hierarchy of justifications. We might suppose that we should always prefer a liability based justification, when one is available, and consider a lesser evil justification to be a last resort. The problem with this view is that it seems to violate something like the constraint of necessity. Suppose the harm that must be imposed on the bystander is negligible, equivalent to a few minutes' delay and nothing more. It would be immoral to kill the Villain rather than impose this delay on the bystander. Moreover, it is immoral in the same way that unnecessary defensive actions are immoral. Although it would be permissible to kill the Villain if there were no better option, it is immoral to kill him since a much better option is available.

In fact, it might be possible to treat *Multiple Justifications* as simply another case of necessity. If the constraint of necessity simply tells us to choose the best defensive option available, then multiple justification cases are just another kind of necessity case. It might appear that trading harms between the Villain and the Bystander can be made in the same way that the constraint of necessity makes trades between harms to the Villain and harms to the Victim. After all, as in standard cases of necessity, we are in effect asking whether an innocent person can be reasonably expected to sacrifice themselves to save a culpable aggressor from a harm they would otherwise be liable to. Yet despite its initial plausibility, this approach should be resisted, at least if we want our judgments to be consistent.

I have described a liability as a consideration that alters the moral weight we should give to defensive harms applied to the liable person when deciding what to do. In the previous section, I

even used some arbitrary numbers to show how this might, in theory, work. This numerical approach can also help us think clearly about multiple justification cases. We must start by considering what our benchmark level of harm will be. Suppose we stipulate, for the sake of discussion, that harms applied to highly culpable aggressors are discounted by a factor of ten. We need to know what this discount is relative to. I think it is most useful to choose the harms that would otherwise befall the victim as our benchmark.

Now we must consider how harms imposed on the bystander compare to harms we allow to befall the victim. To keep things simple, let us assume that both Villain and Bystander will be harmed merely as a side-effect of saving the victim. Even so, the Bystander ought to have some protection against being harmed as a side-effect of saving the victim. The Bystander can reasonably expect that she will only be harmed on the basis of a lesser evil justification when the harm that will otherwise befall the Victim is significantly greater than the harm that must be imposed on Bystander to prevent it.

Once again, and purely for the sake of discussion, let us stipulate that harms imposed on the Bystander should be treated as more morally significant by a factor of ten. This is another way of saying that, other things being equal, we can harm the bystander to save the victim on the basis of a lesser evil justification in cases where the victim would otherwise suffer a harm that is at least ten times as great. So, for example, if four broken limbs is ten times as harmful as a broken wrist, we can permissibly break the bystander's wrist as a side-effect of saving the victim from four broken limbs. We could not, however, break the bystander's wrist as a side-effect of saving the victim from one broken limb. Whether these are the right numbers or mutilations does not matter for my purposes. Different numbers and examples could be substituted if this makes the argument more plausible.

The liabilities based account tells us that we can discount defensive harms applied to the liable person. The lesser evil justification tells us, roughly, that a widely proportionate harm must be a substantially lesser evil. Both these principles appeal, at least implicitly, to the idea that defensive actions are permissible when they make good trades, once all the relevant liabilities and protections have been taken into account. What we ought to do, when combining these principles, is make the best trade available.

We can see how this works by considering a case where allowing the victims to be harmed can be ruled out. Consider,

Multiple Justifications II: Villain has culpably launched trolley, carrying nuclear waste, that threatens to destroy the entire town. We can save the town either by imposing a very great harm on the Villain or by imposing a much smaller harm on an innocent bystander.

Clearly, we cannot allow the town to be destroyed. To make sense of this case we must have some idea of how we can trade harms between aggressors with liabilities, and innocent bystanders who will not otherwise be harmed. The account I have been developing does not tell us exactly how much harm we can permissibly apply to each. However, it does tell us how we should think about this case, and how we can keep our judgements consistent. We can now see why we cannot trade harms between the villain and the bystander in the same way that necessity trades harms between the villain and the victim.

This is because we earlier assumed that harms to the bystander should be treated as ten times as significant as harms to the victim, and that harms to the villain should be treated as ten times less significant than harms to the victim. Thus, we cannot consistently accept that exchanging harms between the bystander and the villain is akin to trading harms between the victim and the bystander. To be consistent, we must treat harms to the bystander as a hundred times as significant as harms to the villain.

Justifying harming the bystander rather than the villain, is much harder than justifying harming the bystander rather than the victim. This is because the bystander has special protections against being harmed that apply in addition to the villain's liability. We must conclude that in *Multiple Justification II* we should be extremely averse to harming the bystander rather than the villain. In effect we end up with something like the principle with which we started. In practice we should almost always prefer to use a liability based justification rather than a lesser evil justification.

We have been assuming that the bystander's special protection depends on the fact that she, unlike the victim, will not be harmed if we do nothing. This raises the possibility that in some cases it may be unclear who is the victim and who is a bystander. Consider,

Multiple Justifications III: The same as *Multiple Justification II* except that diverting the trolley away from the town requires us to first pull a lever to divert it on to a track where it threatens to harm the bystander. We can then pull another lever that will divert it, before it hits the bystander, onto a track where it will harm the villain who launched it instead.

The status of the bystander is somewhat unclear in this case. If we could just pull one lever to divert the trolley towards the villain, then the bystander would be clearly a bystander. Yet if we must first divert the trolley towards the bystander, it might appear that the bystander then becomes a victim and has a different status in our deliberations.

My inclination is that the difference between pulling one levers and two levers should not in itself be significant in this case. What matters is the intention with which we act. If we pull the first lever in order to save the town by sacrificing the Villain, then the bystander remains a bystander although she temporarily occupies the causal place of a victim. On the other hand, if a gust of wind blew the first lever, and we then had to decide whether to pull the second, then the bystander would genuinely become a victim and ought to be treated as such. This case raises difficult questions that apply elsewhere in deontic ethics, but I will not pursue the issue further here.

As in preceding sections I have avoided making controversial claims about what we ought to do in concrete cases. For the sake of argument, I have assumed that certain actions are permissible or impermissible, but I have not tried to argue for or against permissibility in any specific cases. I have specifically tried to avoid such claims. My focus is not on what it is permissible to do in cases of self-defense: whether it is proportionate to kill 10 villains or 100 to save an innocent victim. Rather, I have been arguing for a certain way of understanding these cases, a way of connecting our intuitions into a parsimonious and coherent theory.

7. Conclusion

Above, I stated that there were two respects in which I believe our current understanding of moral liability can be improved. First, I argued that we can make progress by separating the idea of moral liability from the idea of rights forfeiture. We should be open, I suggested, to other relationships between rights and liability. Second, I argued that we must distinguish between the claim that a person is liable to some harm in some circumstance, and the claim that a person has a liability which might make him liable to some harm in some circumstances. To understand liability, I claimed, we need to develop a theory of liabilities.

To defend these two points, I had to show that we can make sense of liabilities without defining them in terms of rights. To do this I considered what would be left of the concept of moral liability if it is not defined in terms of rights. Liabilities, I suggest, can be understood as moral considerations that decreases the significance of defensive harms applied to the liable person. This is a promising approach because it offers a clear and simple way of understanding the otherwise complex constraints of necessity and proportionality. To further this approach, we must investigate how liabilities, and with them necessity and proportionality, can be justified. In the next chapter I

consider the possibility that choices, such as the decision to culpably attempt to harm another, can be used to justify liabilities.

Chapter Two:

Choice and the Justification of Liabilities

Introduction

At Brasenose College, in May 1986, T. M. Scanlon delivered the Tanner Lectures on Human Values.³² His theme was the significance of choice. In these lectures, Scanlon laid out two influential theories. The first was a theory about what it takes to be blameworthy, and what we do when we blame other people. Scanlon argued that blame is a response to the bad quality of a person's will. The second was a theory about the justification for liabilities. Scanlon argued that we can justify self-defense, and other liability based practices, by comparing the costs the attacker would have to suffer if he refrained from attacking, to the costs the victim would have to bear if she allowed herself to be attacked.

At the very end of these lectures Scanlon conceded that “both analyses are required to account for the significance of choice in morality, and both are required to explain its force in the law.”³³ The fact that attackers choose to create threats to others, when they could easily refrain from doing so, provides a good justification for liability. Yet, this justification side-steps the central role that blameworthiness, or culpability, seems to play in our intuitions about these cases. The dilemma Scanlon faced, which is shared by all those who have taken up his approach, was to show how these two analyses can be combined.

In this chapter, I take up the significance of choice as a justification for self-defense and liability. I leave open the possibility that there are other available justifications that do not involve choice.³⁴ However, I am interested in the role of choice. To understand the significance of choice, I

³² T. M. Scanlon, “The Significance of Choice,” *The Tanner Lectures on Human Values* 7 (1986) 149-216.

³³ Scanlon, “The Significance of Choice,” 211.

³⁴ See, for example: Helen Frowe, *Defensive Killing: An Essay on War and Self-Defence* (OUP, 2014) Chapter Two.

argue, we must look beyond costs, and consider the reasons that each person has for acting. It is the person's response to reasons that gives her choices their moral significance.

In the first section of the chapter I argue that we can use the attacker's choices to justify self-defense once we consider the morally weighted costs of avoiding the attack. These costs are morally weighted to account for the moral reasons that the person has for or against attacking. With this adjustment, the significance of choice argument can explain the permissibility of self-defense against culpable attackers, as well as the constraints of necessity and proportionality that limit it. A person's sensitivity to action guiding moral reasons is directly related to her culpability. In this way, by attending to moral reasons for action, we can unify the two parts of Scanlon's discussion of the significance of choice.

In section two, I consider some of the implications of this justification for self-defense. I point out that both causal and moral responsibility are not necessary elements. It may be the culpability, and not the consequences, that justifies liability in some cases. In fact, this way of justifying self-defense suggests that it would be unfair for flukes of causation to make a moral difference. This does not mean that causal and moral responsibility are irrelevant. However, it does suggest that causal responsibility may be far less central to liability than is often supposed.

In section three, I consider a second possible justification for liability, one that depends on the risks that we impose on others. It is tempting to suppose that we ought to internalize these risks whenever possible. This might provide a separate justification for liability, one that does not depend on culpability or the significance of choice. If we ought to internalize these risks whenever possible, then we cannot reasonably object to liability.

I argue that to make sense of this justification, we must consider both those risks that we impose and those risks that we assume. Some cases will be more a matter of risk imposition, and others will be more a matter of risk assumption. To make this distinction, we must consider the

costs that people face and the reasons they have for acting. Thus, despite first appearances, this justification relies on the same sort of significance of choice analysis. This raises the possibility of a unified theory of the justification of liabilities, one that depends on choices and reasons for action, rather than on responsibility.

1. The Significance of Choice

Many justifications for self-defense, against culpable aggressors, begin by recognizing that the aggressor is responsible for bringing about a situation in which he or the victim must be harmed.³⁵ The aggressor could have chosen not to create this situation, at relatively little cost to himself.³⁶ I argue, building on the work of Scanlon and Victor Tadros, that this fact, and not the responsibility itself, can justify self-defense, and liability more broadly.³⁷ Once this argument has been further developed we will see that culpability, without moral or causal responsibility, can justify liability to defensive harming.

This justification begins with the premise that, other things being equal, each person ought to have the best possible opportunity to avoid being harmed. For this reason, it is hard to justify imposing harms on a person who does not have a good opportunity to avoid those harms. It is comparatively easy to justify imposing harms on a person who does have a good opportunity to avoid those harms.

There is a subtle shift between this starting premise and the common sense assumption that we ought to minimize the amount of overall harm. It is plausible that the best way to minimize the

³⁵ See McMahan "The Basis of Moral Liability."

³⁶ Helen Frowe argues that this opportunity is part of what makes them responsible. See Frowe *Defensive Killing*, Chapter Three.

³⁷ T. M. Scanlon, *What We Owe to Each Other* (Harvard University Press, 1998) Chapter 6; T. M. Scanlon, "Thomson on Self-Defence" in A. Byrne, R. Stalnaker, and R. Wedgwood eds. *Fact and Value: Essays on Ethics and Metaphysics for Judith Jarvis Thomson* (MIT Press, 2001); and Victor Tadros, *The Ends of Harm* (OUP, 2011) Chapter 8.

overall amount of harm, in the long run, is to give each person the best possible opportunity to avoid being harmed. Yet we also have independent reasons for caring about the opportunities we each have to avoid harm, in addition to caring about the amount of harm overall. In doing so we treat each other as responsible agents, rather than simply as potential sufferers of harm. It may be good for us to have our choices matter in this way.³⁸ It may also just be the case that our choices are morally significant, given the kind of creatures that we are. It is intuitive, even to those who are not contractualists, that a person can reasonably object to suffering a harm he could not easily have avoided. It is less reasonable to object to a harm that one could easily have avoided.

These ideas can be used to justify self-defense. Consider a simple case:

Self-Defense: One person, Villain, culpably attacks another person, Victim. Victim can defend herself but only by killing Villain. If Victim does nothing, she will be killed by Villain.³⁹

We want to know whether, and why, it is permissible for Victim to harm Villain in self-defense. I contend, following Scanlon and Tadros, that it is permissible because a moral rule allowing Victim to harm Villain in self-defense gives each of them the best opportunity to avoid being harmed.

In order for Villain not to be harmed in this case, Victim would have to decide not to defend herself. This would be very costly for Victim because she would have to resist her instinctive desire to defend herself, and she would have to suffer whatever harm Villain inflicts. A moral rule that forbids Victim from defending herself would force Victim to bear a substantial cost. Victim cannot easily avoid this cost because, we can assume, she cannot easily prevent people from attacking her. We can improve Victim's opportunity to avoid being harmed, by permitting her to harm Villain in self-defense.

³⁸ Scanlon has developed his own account of the value that choice can have. See T. M. Scanlon, "Responsibility and the Value of Choice," *Think* (Spring, 2013). I assume that choice is valuable, whether for the reasons Scanlon suggests or for other reasons.

³⁹ The *locus classicus* for this style of case is Judith Thomson, "Self-Defense," *Philosophy and Public Affairs* (1991) 283-310.

She could not be permitted, on this basis, to harm *anyone* in self-defense. An innocent bystander would have no better opportunity to avoid being harmed in defense of Victim than Victim had to avoid being harmed by Villain. The same is true of an entirely innocent threatener, one who poses a threat of harm to others without having any opportunity to avoid posing this threat. If Victim were permitted to defend herself against an innocent threatener, that innocent threatener would have no better opportunity to avoid harm from Victim than Victim had to avoid harm from the threatener. Other things being equal, the significance of choice argument will not help us choose between them.⁴⁰

In the case of a culpable aggressor, things are very different. A rule that allows Victim to defend herself by harming Villain, would not force Villain to suffer a harm that she could not easily avoid. All Villain must do to avoid being harmed in self-defense, according to this rule, is to avoid culpably attacking others. This should be relatively easy for Villain to do. It might force Villain to repress her natural aggression, or forgo some benefit she can achieve by harming Victim, but these costs are not as morally significant as the cost to Victim of allowing herself to be attacked. Innocent bystanders, and even innocent threateners, can reasonably object to a moral rule that allows them to be harmed in defense of victim. They can object because such a rule would not give them a good opportunity to avoid being harmed. Crucially, this objection is not available to culpable threateners such as Villain.

This is because Villain is morally obligated not to attack Victim. If this attack is made more costly for Villain, she cannot fairly complain. She has only lost an opportunity to do something that she was already not permitted to do. Villain can avoid being harm, according to such a rule, simply

⁴⁰ This is not to say that there is no other way of justifying self-defense against innocent threats. See Helen Frowe, *Defensive Killing*, Chapter 2.

by doing what she is already morally obligated to do. This is the second crucial step in the significance of costs argument. Scanlon summarizes this argument, writing,

The fact that V could avoid an undesirable consequence of another person's act by doing X may undermine V's objection to a principle permitting the other's act. But whether it does so depends on the costs to V of doing X. If X involves very great sacrifice for V, then the availability of this option may not diminish the force of V's objection. This is where fault becomes relevant. If X is something that V is morally required to do, or, to say the same thing, if V would be acting wrongly by failing to do X, then V could not appeal to the sacrifice involved in doing X in order to argue that the availability of this option does not diminish his objections to the principle in question.⁴¹

If we substitute V for the Villain, and X for not culpably attacking others, this argument can be used to justify liabilities for culpable aggressors. For Villain to avoid acquiring a liability, all she has to do is that which she is already morally obligated to do. In contrast, if Victim was not allowed to morally discount defensive harms applied to Villain, he would have to bear a substantial cost. Of course, this argument does not tell us exactly how much Victim can discount Villain's interests, or exactly what Villain must do to acquire a liability, but perhaps we should not expect this kind of precision from a theoretical justification.

It is important to notice the role that choice plays in this argument. It is plausible to suppose that if someone brings a risk upon themselves, despite having a good opportunity to avoid it, then that person has a reduced claim to be rescued. For this reason, it might be thought that by choosing to culpably attack Victim, Villain is willingly choosing to make herself liable. If this was our justification for liability, then it would be circular. It would use the fact that Victim is permitted to defend herself to explain why Victim is permitted to defend herself.

The significance of choice justification for liability operates at a higher level of abstraction. It justifies liability as a moral rule on the grounds that it gives each person the best opportunity to avoid harm, and on the grounds that culpable aggressors, such as Villain, cannot reasonably object

⁴¹ Scanlon, "Thomson on Self-Defense," 205. This is quoted by Tadros in *The Ends of Harm*, 172.

to it. Ultimately, these are the same sorts of reasons that could justify the assumption that people who willingly bring risks upon themselves have a reduced claim to be rescued. In both cases, the quality of the opportunity a person has to avoid suffering from a moral rule undermines that person's capacity to reasonably object to that rule.

This justification, building on the work of Scanlon and Tadros, begins with the assumption that each person should have the best possible opportunity to avoid harm. A principle that forbid self-defense would not give victims a good opportunity to avoid harm. By contrast, a principle that allows victim to defend themselves by harming culpable aggressors gives victims a much better opportunity to avoid harm. It does so while still providing potential culpable aggressors with a good opportunity to avoid harm. All a person must do to avoid being harmed, according to this principle, is just what we are all already obligated to do. We must simply avoid culpably attempting to harm others.

We can distinguish two closely related roles that choice plays in this argument. First, the value of having choices matters for the premise that we each ought to have the best possible opportunity to avoid harm. The quality of the choices a person is faced with depends, at least partially, on the quality of that person's opportunities. Liability to defensive harming is justified on the grounds that it substantially improves the quality of the opportunities that potential victims are faced with, without substantially decreasing the quality of the opportunities that potential culpable aggressors are faced with. It does not reduce the value of any option that potential culpable aggressors are morally permitted to choose. Choice matters, in this sense, because the quality of the choices each person is faced with is something that liability helps to maximize.

Second, and relatedly, having choices matters because it undermines a possible objection that culpable aggressors could make to being harmed in defense of potential victims. As Scanlon argues, when a person has a good opportunity to avoid being harmed, this undermines that person's ability

to reasonably object to suffering that harm. Culpable aggressors cannot reasonably object to liability, because they are already morally obligated not to culpably attempt harm to others. For this reason, they have a very good opportunity to avoid becoming liable.

For each of these roles, what matters is not just what a person deliberately chooses, but also the options with which that person is faced. As Scanlon puts it, “what matters fundamentally ... is not that the agent made a choice, but that he or she *had* a choice.”⁴² This may not be as significant a departure from other accounts of the significance of choice as Scanlon goes on to suggest.⁴³ To know moral the significance of a choice, we must consider not only what the person chose, but also the other options that she was faced with. This is true on any plausible account of the moral significance of choice.

Moreover, while the focus here is on each person’s opportunities, these opportunities are only morally significant because of each person’s capacity to choose. We each exercise our agency by responding to the reasons that we have and the opportunities that we face. This is what makes us morally responsible. In a specific case, to know the significance of an individual choice, we must look at the opportunities the person faced. Yet we only care about opportunities in each case, because we have a general concern for the person’s capacity to respond to reasons and opportunities.

It is a further question to decide which sorts of responses are morally significant; are only fully informed choices significant, or are failures to realize and remember significant as well? I turn to this further question when I discuss the nature of culpability in Chapters Five and Six. For now, we can put it aside and focus on cases of deliberate choice.

⁴² Scanlon, “Responsibility and the Value of Choice,” 11.

⁴³ For a more detailed argument on this point see Serena Olsaretti, “Scanlon on Responsibility and the Value of Choice,” *Journal of Moral Philosophy* 10, 4 (2013) 465-483.

Putting these ideas together we can say the following; what we choose matters because of the choices we face, and the choices we face matter because of our capacity to choose. For a deeper justification of liability, we must turn to this higher level of abstraction. Above, I gave three suggestions for a deeper justification of the premise that we each ought to have the best possible opportunity to avoid harm. We can now see that these three suggestions are each tied to our capacity to choose, to our being morally responsible agents. First, it is because of our capacity to choose that giving us each the opportunity to avoid harm is likely to reduce the amount of harm overall. Second, it is our capacity to choose that makes it valuable for us to have lives in which our choices are morally significant. Third, it may just be the case that our capacity to choose makes us the kind of creatures whose choices are morally significant. I will not explore these suggestions further here. Instead, I will return to the more practical implications of our choices, in cases of self-defense and liability.

1.1 Accounting for Proportionality

As I have just explained it, the significance of choice argument struggles to properly explain proportionality. Recall that, the fact that a person could easily avoid some action undermines their ability to reasonably object to consequences, such as liability, being attached to this action. This is combined with the premise that people can make no appeal to the costs of avoiding actions that they are already morally obligated to avoid. Both these premises appear plausible, but together they can produce some bad results. There are two reasons for this.

First, because people can make no appeal to the costs of avoiding impermissible actions, there is no inherent limit to the adverse consequences that could be attached to even minorly immoral acts. Suppose I suggest that anyone who culpably attempts to steal anything can be permissibly tortured and killed to prevent the theft. This is a deeply unintuitive suggestion, because it

violates the classic constraint of proportionality. Yet, I might defend this suggestion by pointing out that a person must only avoid stealing, to avoid suffering from my proposal. Since this is something we are all obligated to do anyway, according to the significance of choice argument, no one cannot reasonably object to my suggestion.

Second, because people can make no appeal to the costs of avoiding impermissible actions, there is no distinction made between very culpable acts and only slightly culpable acts. Costs associated with both acts are simply reduced to zero, making them identical from the perspective of the significance of choice analysis. This makes our justification for liabilities insensitive to an important moral distinction.⁴⁴

One possible response to these problems is simply to accept them. Some might claim that people who culpably attempt to harm others are liable to any necessary defensive harm, however great. The constraint of proportionality, on this view, would depend on other moral principles, such as duties of beneficence or the inherent badness of suffering. Some might even claim that it is permissible to inflict any necessary defensive harm, and that proportionality is supererogatory. I will not reply directly to these suggestions, except to say that they appear to me to be second best options. This is because they still leave us with the problem of justifying proportionality. It would be better if we can explain the proportionality relation using the same principles that justify liability.

Another possible response is to appeal to the comparative difficulty of avoiding all minorly culpable choices. This is the strategy adopted by Tadros. He points out that,

It is quite difficult to go through the whole of one's life without acting recklessly or negligently. Many of us, at some point, will be tempted to do so. In contrast, it is easy to go through the whole of one's life without intentionally seriously harming another where one is not justified in doing so.⁴⁵

⁴⁴ Some writers would not consider this a defect. See, for example, Jonathan Quong, *The Morality of Defensive Force* (Forthcoming).

⁴⁵ Tadros, *The Ends of Harm*, 176.

There are two ways of understanding this argument. First, it might be argued that avoiding minorly culpable choices requires more effort. This may be true, but it seems to directly contradict the claim that we cannot appeal to costs we are obligated to bear. Morality can be very demanding, both in the sacrifices it rarely demands and in the attention it always demands. But since we are obligated to bear these costs, the significance of choice analysis claims that we cannot appeal to them.

Second, it might be the probability that we will become slightly culpable that matters. This explanation of proportionality assumes the premise that the more likely we are to commit some act, even if it is something we are obligated not to do, the less adverse consequences can reasonably be attached to it. This is not an appealing premise, because the link between culpability and probability is not consistent. We can easily think of two acts that are equally culpable but not equally likely to occur. For serious crimes, like rapes and murders, attackers are much more likely to have a prior relationship with their victims. However, this fact does not give me any reason to suppose that the perpetrators of these more probable crimes are liable to less defensive harming.

The fundamental problem with this version of the significance of choice analysis is that it incorporates our moral obligations, but it does not account for the fact that the force behind these obligations comes in degrees. It will always struggle, therefore, to properly explain the proportionality relation. The relationship between culpability and liability is too fundamental to be side-stepped or ignored. To solve this problem, we have to go back to Scanlon's original dilemma and find the connection between a bad quality of will and the costs we must bear to avoid a given act.

1.2 The Morally Weighted Cost of Avoidance

The problematic premise in the significance of choice argument is the claim that we can make no appeal to costs that we must bear to avoid acts that are impermissible. For convenience, let

us call this the no-appeal premise. This is an intuitively plausible premise. Yet, it produces such implausible results that we have reason to rethink it.

The fact that we have a moral reason not to do something gives us a reason to give less weight to the costs of avoiding it, when applying the significance of choice argument. I contend that the strength of this reason comes in degrees. After all, some moral reasons are stronger than others. Rather than make no appeal to the costs of avoiding impermissible actions, we should instead weight those costs according to the moral reasons for or against the action. The stronger our moral reasons for not acting in that way, the more the costs of not attacking in that way should be discounted.

We can see this by considering how the significance of choice analysis could be applied to acts that are not impermissible, but are obligatory. For example, suppose we are considering a moral rule that would attach adverse consequences to an act that is morally obligatory. For example, suppose we are considering whether wealthy people who willingly give to charity should be forbidden from using violence to defend the rest of their property against theft. Such a rule seems perverse, but it is instructive to consider why.

The reason this rule seems perverse is that the wealthy are, I am assuming, obligated to give some of their money to charity. We should be particularly averse to attaching adverse consequences to things that people are morally obligated to do. One reason for this is that we want people to do what they have moral reason to do. Another reason for this is that it is particularly unfair for us to collectively to demand some act, and then make it more costly. For both these reasons, it seems that the significance of choice analysis should give extra weight to the costs of avoiding things we are obligated to do.

A similar point can be made about acts that are supererogatory. Suppose that it is supererogatory to devote your life to alleviating poverty. Now suppose that we are considering a

moral rule that would forbid those who devote their lives to alleviating poverty from using violence to defend their property against theft. This rule seems perverse, for the same reasons discussed above. Supererogatory acts are things that morality recommends. We should be particularly averse to moral rules that attach adverse consequences to acts that are supererogatory. The significance of choice analysis should therefore give extra weight to the costs of avoiding acts that are supererogatory.

This mirrors my proposed replacement for the no-appeal premise. The fact that we have a moral reason to do something gives us a reason to give extra weight to the costs of avoiding it, when applying the significance of choice argument. In the same way, the fact that we have a moral reason not to do something gives us a reason to give less weight to the costs of avoiding it. The strength of these reasons comes in degrees, because some moral reasons are stronger than others. What matters is not simply the opportunity each person has to avoid harm, but the morally weighted opportunity to avoid it.

1.3 Support from Other Cases

We can find further support for this amendment by considering other cases in which Scanlon's significance of choice analysis might be applied. These include cases where we must choose which person to save from some impending harm. Above, I mentioned the plausible assumption that a person who has a good opportunity to avoid some harm may have a reduced claim to be rescued from that harm. While this assumption is not a part of the justification for liability, the same sorts of reasons that justify liability may also justify this assumption.

By considering some of these rescue cases we can find further support for my suggestion that the significance of choice argument should be sensitive to the strength of our moral reasons. For example, consider,

Obligation: Joe has a strong moral obligation to move some hazardous waste through the park. If he does not do this, millions of townspeople might die. Sarah does not want to be in the park, but she drives an ambulance and mistakenly believes that someone on the other side of the park needs to be rescued. Zack has no moral reason to be in the park, but he has a powerful desire to go skateboarding.⁴⁶

Now suppose that despite being aware of the danger, Sarah and Zack both enter the park and risk exposing themselves to the hazardous waste Joe is moving. Joe can save one of them, but not both.

What should Joe do? We can use the significance of choice analysis to decide.

Following this analysis, what matters is the costs that each of these people would have to bear in order to avoid exposing themselves to the harm. If we do not account for their moral reasons, then it seems that Zack has the stronger claim to the suit. This is because he would have to reject his powerful desire to go skateboarding, whereas Sarah would prefer not to be in the park at all. This analysis would be lacking. Although Sarah does not want to be in the park, she has a much better evidence-relative moral reason for being there than Zack does. For this reason, it seems to me, we should give extra weight to the costs Sarah would have to bear by avoiding the park. On balance, I think that Sarah has a better claim to the spare hazmat suit than Zack, but the significance of choice analysis can only give us this result by morally weighting the costs of avoiding the park.

We can see the same thing when we consider those who have moral reasons not to do things. Consider:

Obligation (Immoral Variation): Mikey is a convicted sex-offender who is forbidden from visiting the park, but has a powerful desire to go skateboarding. Timmy is a local teen who promised his mother he would not go to the park, but has wandered in anyway. Both these people entered the park despite being warned of the dangerous waste. Only one of them can now be saved.

⁴⁶ This is a variation on a classic series of hazardous waste cases presented by Scanlon in *What We Owe to Each Other*, 257. I use my own versions of these cases to draw attention to a different set of issues. For an overview of the usual context for such cases see Rahul Kumar, "Contractualism and the Roots of Responsibility," *The Nature of Moral Responsibility: New Essays* ed. by Clarke, McKenna, and Smith (OUP, 2015).

In this case, it seems to me that Timmy, rather than Mikey, should be saved. This is because although both Timmy and Mikey were obligated to avoid the park, Mikey had a stronger moral reason to avoid the park than Timmy. Yet the significance of choice account can only reach this result if it replaces the no-appeal premise with morally weighted costs of avoidance. Otherwise, there would be no way for the significance of choice analysis to distinguish between these two, since both were obligated to avoid the park.

The significance of choice analysis only gets the right results when it accounts for the moral reasons people have for their acts and accounts for the relative strength of these moral reasons. The best version of the significance of choice analysis is one that rejects the no-appeal premise and replaces it with moral weighting. This is justified because, on closer inspection, the reasons we have to accept the no-appeal premise can also support moral weighting. It is further justified because moral weighting produces better results than the no-appeal premise.

1.4 Moral Reasons and Culpability

By replacing the no-appeal premise with moral weighting, we incorporate culpability directly into the justification for liability. This is because, as I will argue in Chapters Five and Six, there is a direct link between the strength of the moral reasons you have not to do something and your culpability for doing it. The culpability of a person depends on the sensitivity to morally significant considerations manifested by his actions and intentions. Morally significant considerations are those things that give us moral reasons. So the stronger the moral reasons a person has not to do something, the less sensitivity to morally significant considerations he manifests by doing it.

The kind of moral reasons that matter for moral weighting, and for culpability, are what we might call subjective moral reasons. These are the moral reasons that a person ought to accept and act on given his non-moral beliefs and the available evidence. If I reasonably believe that I ought to

save you from drowning, then we could say that I have a subjective moral reason to save you. This would be true even if you were actually safe, and by jumping in I would risk injuring you.

It would not be better if the significance of choice analysis instead appealed to objective moral reasons, those that hold regardless of the evidence or what you happen to believe. We can see this by returning to *Self-Defense*. Suppose that by attacking Victim, Villain will actually save Victim from another worse attack, in a way that neither of them can foresee. Objectively, we might say, Villain has a moral reason to attack Victim, even though given her epistemic position Villain has a strong subjective moral reason not to attack Victim. In this case, Victim would certainly be permitted to defend herself, even though she might later realize that this was a mistake.

In this way, we can now resolve Scanlon's dilemma. The quality of will people express in their actions is directly related to the subjective moral reasons they have not to perform those actions. The subjective moral reasons people have to perform various acts matters when assessing the moral significance of the costs these people would have to bear to avoid those acts. This modified version of the significance of choice argument can properly explain the relation between culpability and liability.

To complete this analysis, we must see how the appeal to morally weighted costs can explain proportionality in cases of self-defense. We can begin with two assumptions: first, that there is a strong moral presumption against deliberately choosing a distribution of harms that permits more total harm than other available distributions, and second, that there is a strong moral presumption against deliberately inflicting harms on others. Both these assumptions are intuitively plausible, and widely supported.

Liabilities are reasons that can outweigh these plausible assumptions. The stronger the liability, the more harm it will permit us to deliberately inflict, and the further it will permit us to go from a simple consequentialist distribution of harms. Yet liabilities do not trump these assumptions.

We always have some reason not to deliberately inflict harm, and we always have some reason not to choose a distribution that inflicts a greater total quantity of harm. The proportionality relation emerges from the conflict between these three sorts of moral reasons. Liabilities are reasons to discount the interests of certain people, such as culpable attackers, yet these reasons are balanced by reasons we have not to deliberately inflict harms and reduce the total quantity of harm.

Proportionality is a matter of striking this balance.

A justification for liabilities will give us the proportionality relation if it suggests that the strength of the reason we have to discount the interests of culpable aggressors comes in degrees, and if the strength of this reason correlates with culpability. A morally weighted version of the significance of choice analysis can do just that. It provides us with a justification for discounting the interests of those who culpably attempts to harm others, and the strength of that reason correlates with culpability.

In this section, I have argued that the significance of choice gives us a promising justification for self-defense and other cases of liability. It is promising because it can do more than merely identify the necessary and sufficient conditions for liability. It can explain *why* these things can make a person liable to defensive harming. It can do this using simple and widely accepted principles, such as the claim that each person ought to have the best possible opportunity to avoid harm. These principles are appealing to consequentialists, deontologists, and contractualists alike, raising the possibility that the permissibility of self-defence might ultimately be sanctioned by a unified theory of morality.

These are ambitious claims. Indeed, it is ambitious even to claim that the significance of choice argument offers the best justification for self-defense. Defending such a claim would require a negative argument: an argument that shows other potential justifications to be inadequate. I do not

pursue these negative arguments in this thesis. My purpose is simply to elaborate what I take to be a highly promising justification, and to point out a few surprising implications.

2. Causation and Culpability

Perhaps the most striking feature of the significance of choice argument, as I have offered it, is that it makes no reference to causal or moral responsibility. This is striking because these sorts of responsibility are widely assumed to be necessary, and sometimes sufficient, conditions of liability to defensive harming. The significance of choice argument, once elaborated to account for proportionality, raises the possibility that these widespread assumptions are mistaken. In this section I make this case, arguing that the significance of choice argument makes culpability, rather than causal or moral responsibility, the central ingredient in liability.

I concede that moral responsibility depends, to some extent, on the quality of opportunity a person has to avoid certain acts.⁴⁷ The same quality of opportunity, to avoid attacking Victim, may explain both why Villain is liable and why she is responsible. For this reason, liability and moral responsibility will correlate in most cases. Yet the liability need not depend on the responsibility. What matters is the nature of the choice we are faced with, the reasons we have for making those choices, and the costs that can fairly be attached to those choices.

Still, many of the claims that are made about moral responsibility can be supported by the significance of choice analysis. If a person had an excellent opportunity to avoid causing some threat of harm, then we can suppose that she is morally responsible for it and may be liable. It may seem plausible, when choosing between two people, that the one with the better opportunity to avoid creating some threat is the one more responsible for it.

⁴⁷ Frowe develops of an account of moral responsibility along these lines in *Defensive Killing*, Chapter 3.

This seems especially plausible once we account for the moral reasons we have to impose or avoid imposing certain risks. If a person, such as an ambulance driver, is morally obligated to cause some risk of harm, then that person might not be liable.⁴⁸ These are the sorts of claims that are made by the responsibility view, and they can in most cases be supported by the significance of choice argument.

However, we have reason to doubt whether it is responsibility, rather than choice, that makes the difference. People who are obligated to take some risky course of action do sometimes seem, to me at least, to be clearly morally responsible for any foreseeable harms that result. They are morally responsible, yet blameless. It is tempting to deny that these people have liabilities.⁴⁹ Yet it seems odd to claim that an ambulance driver is less responsible for what they do, simply because they have a good moral reason for doing it. Indeed, what matters is precisely the fact that the ambulance driver imposes a risk responsibly, by responding to her moral reasons to impose it.

The significance of choice analysis explains why the ambulance drivers are not liable, using the same argument that explains why more culpable risk imposers can be liable. To incorporate these intuitions into a responsibility view requires a special exception. This suggests that the responsibility view does not quite describe the key ingredient in liability. The significance of choice analysis tells us that choices and culpability, rather than responsibility, generate liability.

In fact, the significance of choice analysis suggests that causal responsibility for a threat should not matter for liability. It is unclear whether causal responsibility is always a necessary condition of moral responsibility. Yet causal responsibility does seem to matter for moral responsibility in a way that it does not matter for the significance of choice argument. Consider the following case:

⁴⁸ McMahan, "The Basis of Moral Liability," 398.

⁴⁹ McMahan, "The Basis of Moral Liability," 398.

Causation: Two equally culpable villains, Attacker and Attempter, each fire a gun at Victim. By an unforeseeable fluke, Attempter's gun jams so that only Attacker could be causally responsible for killing Victim. Victim can defend himself by harming either Attacker or Attempter, or by splitting the defense harm between the two. If Victim does nothing to defend himself, he will be killed.

If Attempter were not present, then clearly Attacker would be liable to be harmed in self-defense. This is because the morally weighted cost to Attacker of not attacking is negligible, while the cost to Victim of allowing himself to be killed is very high. Yet since Attempter is present, it seems to me that he should receive a share of the defensive harm as well.

Here is why. The morally weighted costs of not attacking are identical to the morally weighted costs of not attempting. It is no more burdensome for Attacker to decide not to fire his gun in *Causation* than it is for Attempter to refrain from firing his gun. If anything, the costs of not attacking might be considered higher, if they include forgoing the pleasure of a successful attack. On balance, I think it is better to focus on the subjective costs, those a person can anticipate given their epistemic position. From this perspective, Attempter and Attacker are exactly equal. Moreover, by permitting the Victim to defend herself by harming either Attacker or Attempter, we are improving her opportunity to avoid being harmed in a wide range of similar cases. Victim's opportunity to avoid harm is improved in a way to which neither Attacker nor Attempter can object.

If we proposed a rule according to which only Attacker could be harmed in self-defense, it would be reasonable for Attacker to object. He could reasonably claim that the costs he would have to bear to avoid attacking are not smaller than the costs that Attempter would have to bear to avoid attempting. It is unfair, from the perspective of the significance of choice analysis, to attach adverse consequences to attacking alone, rather than to attacking and attempting. There may be other reasons for thinking that causation matters for liability. However, the significance of choice analysis does not provide them.

Rejecting causal responsibility, as a necessary condition of liability, is intuitive in a wide range of cases.⁵⁰ As Gerhard Overland points out, it is particularly intuitive to suppose that causal responsibility is not necessary in cases that involve major culpability.⁵¹ Consider,

Firing Squad: A firing squad is preparing to unjustly execute an innocent victim. We can save the victim, but only by shooting any member of the squad. We know that, unbeknownst to them, only one member of the squad has a loaded gun. This member happens to be the least culpable, since he is acting under considerable duress. The other members of the firing squad are fully culpable and would be bitterly disappointed if they knew that their guns were not loaded.⁵²

My intuition, in this case, is that we ought to shoot any of the squad members other than the squad member who has the loaded gun. This is because he is acting under duress. He had a substantially worse opportunity to avoid attempting to harm the victim, and can therefore more reasonably object to liability.

It is not my purpose, in this chapter, to make an intuition based argument for the plausibility of rejecting causal responsibility as a necessary condition of liability. I turn to these issues in Chapter Four. In this chapter, my focus is on elaborating a choice based justification for liability. My point is simply that we can justify liability without reference to causal or moral responsibility.

By contrast, the significance of choice account makes culpability central to the justification of liabilities. Once we take account of the morally weighted costs of avoiding impermissible acts, we can see that making highly culpable choices provides a clear justification for liability. The more culpable the choice, the stronger the moral reasons the person had not to choose it, and so the less that person can reasonably object to being liable as a result. They had a good opportunity to avoid

⁵⁰ A classic example is the ‘basement window’ case from McMahan, “The Basis of Moral Liability,” 391. See also, Gerhard Overland, “Moral Taint: On the Transfer of the Implications of Moral Culpability,” *Journal of Applied Philosophy* 28, 2 (May 2011) 122-136; Lars Christie, *Killing One to Save Another* (PhD Thesis, University of Oslo: 2015); and Victor Tadros, *The Ends of Harm* (OUP, 2011).

⁵¹ Overland, “Moral Taint,” 122.

⁵² This is a variation of a case presented by Lars Christie in *Killing One to Save Another*, 122.

the choice, since they were morally obligated not to make it. In this way, as I have argued, the more culpable the choice, the stronger the liability.

Excuses, such as ignorance and duress, show that the person did not have a good opportunity to avoid his choice. Thus, people with excuses will have weaker liabilities, or no liabilities at all. Culpability can be understood as a measure of the quality of someone's sensitivity to action-guiding moral reasons. In this way, the justification for liabilities is directly tied to the essence of culpability.

None of this implies that causation is irrelevant to liability, or that other considerations, such as moral responsibility cannot also be sufficient conditions of liability. In the next section I discuss the case for non-culpable moral responsibility as a basis of liability. I do not deny that this is possible. However, I contend that the case for such liabilities is more complicated and limited than it might first appear.

3. The Significance of Risk Imposition

I have offered a justification for liability to defensive harms in which causal responsibility does not feature. Indeed, this justification gives us reason to suppose that causal responsibility for a threat of harm would be unfair as a necessary condition for liability. This is significant because causal and moral responsibility are widely considered to be essential parts in the justification for liability. In this section I consider whether there are other choice based justifications for liability that make causal and moral responsibility essential.

We can begin with the concession that causal responsibility is clearly morally significant in a variety of ways. Judith Thomson writes,

You and your neighbor work equally hard, and equally imaginatively, on a cure for the common cold. Nature then smiles on you: a sudden gust of wind blows your test tubes together, and rattles your chemicals, and lo, there you have it. Both of you acted well; but who is to be in pocket for the profits? You are.

Why? That is as deep and difficult a question as the one we are attending to ... "B is responsible for the damage to A's fence; so B should repair it." "The mess on A's floor is B's fault; so B should clean it up." Or anyway, B should have the fence repaired, and the mess cleaned up.⁵³

I am not prepared to answer the 'deep and difficult' questions raised by this sort of luck. It does seem that you, and not your neighbor, has gotten lucky. It also seems that you are entitled to this luck.

The importance of luck strikes me, in this context at least, as a consideration that should be weighed against the importance of fair treatment based on culpability or desert. The fact that you, and not your neighbor, got lucky is a reason to let you pocket the profits. Yet, the fact that your neighbor worked just as hard, and was just as deserving, is a reason to think she deserves something too. Both are important and different people may prefer different compromises between the two.

One reason moral responsibility might matter in these cases is that it can establish a certain sort of ownership. In fact, ownership may provide a good metaphor for moral responsibility. Even if causal responsibility is not a necessary condition of moral responsibility, the acts and events I am morally responsible for are things that in a certain sense belong to me.

Moral responsibility can also establish a more literal sort of ownership. Perhaps this is what happens when my experiment is successful and my neighbor's fails. My experiment and its results are my property. This makes it hard to justify taking some of the profits and giving them to my neighbor. Yet it would not explain the intuition that if B made a mess in A's kitchen, then B should clean it up.

The examples Thomson gives are somewhat quotidian, and they do not tempt us to imagine that B is highly culpable. B's obligation to clean the kitchen clearly depends on something other than his culpability. If I let someone use my kitchen, it is on the tacit condition that if they make a mess

⁵³ Judith Thomson, "Remarks on Causation and Liability," *Philosophy and Public Affairs* 13, 2 (Spring, 1984) 106.

they will clean it up. This example can be used to make a broader point, which might offer a responsibility based justification for liability, in cases that do not involve culpability.

It makes sense to suppose that B's permission to use the kitchen was conditional on B's willingness to clean up the resulting mess. This condition can explain why B, and not A, is the one who should clean up the mess. Only by doing so can B make sure he leaves the kitchen clean, and thus use the kitchen permissibly. A is fortunate to have already satisfied this duty by not making a mess in the first place.⁵⁴

It may be that whenever we impose risks on others, it is only permissible on the condition that we accept liability in the event that those risks result in threats of harm. Only by doing so can we satisfy our obligations not to cause harm to others. This is another way of saying that we could not reasonably object to being harmed in order to prevent a risk we imposed from harming another. This argument could explain why moral responsibility, even without culpability, can generate liabilities. It can be illustrated with a case devised by Ketan Ramakrishnan:

Internalizers: In the future we can buy Risk Internalizers. These clever devices allow us to internalize any risks that we impose on others. If a risk we have imposed is about to result in harm, the Internalizer ensures that we are harmed instead, and the person we imposed the risk on is saved.⁵⁵

Supposing they are relatively cheap, are we obligated to buy Internalizers, and must they always be turned on? Surely, would be very selfish to refuse this futuristic moral convenience and make others bear the risks we impose. If we are obligated to use Internalizers when we can, then we cannot object to being held liable today, on the basis of our moral responsibility for threats to others.

Perhaps, like cooking in someone else's kitchen, we cannot permissibly venture into the world, and risk harm to others, if we are not prepared to clean up after ourselves. This would

⁵⁴ I discuss the relationship between duties and liabilities in Chapter 3.

⁵⁵ In conversation.

explain why it is permissible to kill those who permissibly impose small risks on others, to save those they blamelessly threaten.⁵⁶ Third parties ready to enforce such liabilities might be a second best alternative, until the Internalizers arrive. This argument is persuasive. However, in the next section, I argue that it faces a few complications. Once these complications are accounted for, we find that this justification for liability is not as different from the significance of choice argument as it might first appear.

3.1 Imposing and Assuming Risk

To understand the moral significance of risk imposition, we must distinguish cases of risk imposition from cases of risk assumption. When a risk is also assumed, it no longer seems like the sort of risk we ought to internalize, or for which we can be liable. Consider this case from McMahan:

McMahan:

NASCAR: A spectator leaps down from the stands, onto the track, to get a better photo of a NASCAR race. As a result, he will be killed by the next car unless it is blown-up and the driver is killed.⁵⁷

I agree with McMahan that it is not permissible to kill the driver in this case. This is so despite the fact that choosing to be paid to drive very quickly around a track does impose risks on the people on and around the track. Yet the photographer's reckless act means that this case is more one of risk assumption than it is of risk imposition. Given the choices he has made, the photographer cannot reasonably demand that the driver be sacrificed.

This case can be accommodated into a responsibility based view of liability, since it seems to be the photographer, and not the driver, who is most morally responsible for the impending crash.

⁵⁶ For a forceful argument that permissibly imposing small risks on others can be a source of liability see McMahan "The Basis of Moral Liability."

⁵⁷ In Conversation

This tells us that it can be permissible to impose risks on others, without internalizing them, as long as those risks are also assumed by those others. Risk assumption is just as important and widespread as risk imposition. It should have a central place in any moral responsibility based view of liability.

In many ways risk assumption is simply the flipside of risk imposition. For now, let us assume that a risk is imposed when a person knowingly or negligently causes an increased risk of harm to someone else. A risk is assumed in the same way, when a person knowingly or negligently causes an increased risk of harm to themselves. In both cases the foreseeability of the risk seems to be key. We care about what the person knew, or ought to have known.

Risk assumption complicates the conditional permissibility argument. We cannot assume that if a risk has been imposed, then it ought to be internalized and can be the basis of liability. Instead, we must distinguish between those risks that have been assumed, those that have been imposed, and those that are both assumed and imposed. These cases will not be clear cut. Consider,

Night Driving: A conscientious driver is slightly exceeding the speed limit while driving home at night. At the same time a pedestrian in dark clothes is crossing the road, outside of a crosswalk. As a result, the pedestrian will be killed unless the car is blown up and the driver is killed.⁵⁸

This is a hard case, yet it is also the kind of case that actually occurs. Perhaps the driver ought to drive more slowly, and should know that speeding increases the risk that he will not see a pedestrian. Perhaps the pedestrian should cross at the crosswalk, and should know that Jay-walking at night is risky. Does it matter if the speed limit is unreasonably low, or if the crosswalks are inconveniently placed?

To make sense of moral responsibility based liability, we need an account that can make sense of these cases. There may still be ambiguous or borderline cases, but we at least need to know

⁵⁸ This is a variation of the influential ‘conscientious driver’ case presented by McMahan in “The Basis of Moral Liability,” 393. In McMahan’s version the responsibility is entirely with the driver.

the sorts of things that matter. We also need to know what to do in borderline cases. Suppose we decide that the speeding driver imposed slightly more risk than the Jay-walking pedestrian assumed. This slight difference hardly seems like enough to overcome the otherwise formidable distinction between deliberately killing the driver, and allowing the pedestrian to be killed.

In answering these questions, it may be helpful to consider why we should care about risk assumption at all. Risk assumption matters in the first instance because, if it did not, too great a burden would be placed on the risk imposers. We can see this in *NASCAR*. If the drivers could be killed to save anyone who jumped onto the track, they would not be able to compete without a great risk to themselves. In an extreme case, a malicious person could leap in front of a car simply in order to have a reason to kill the driver. The drivers might decide not to race at all.

Almost any act has the potential to harm someone who is sufficiently careless and unlucky. Just flipping a light switch risks killing someone who has decided to stick his finger in the socket. If risk assumption could not mitigate responsibility and liability in these cases, we would all have to be much more careful. We would have to avoid acts that are much more valuable than *NASCAR* driving. The first reason to care about risk assumption, is to protect the risk imposers from unreasonable danger and restriction.

Imposition and assumption have a way of circling back on each other, and freedom for the risk imposers is also freedom for the risk assumers. If risk assumption could not mitigate responsibility and liability, then we would surely be under much more stringent duties to avoid bringing risks on ourselves. This is because each time we increased the chance of someone else harming us, we would also increase the chance that the imposer would be harmed to save us.

We can imagine this more vividly by picturing a world in which everyone carried Internalizers at all times. Such a world would severely limit what we could safely do without risk of being harmed by our Internalizers. Many valuable acts would be put out of reach. Yet such a world,

it seems to me, would also place severe moral restrictions on what we could do. Many ordinary risky acts, like Jay-walking at night or walking quickly past ‘Slippery When Wet’ signs, might become impermissible. They would be impermissible because we would be unable to take these risks onto ourselves. In this way, risk assumption is valuable for the risk assumers as well, because it allows them a much greater range of permissible action.

This circling back raises a separate problem for the designer of the Internalizer. Once the Internalizer is switched on, risks that are assumed become risks that are imposed. Consider *NASCAR*. The photographer leaps onto the track and assumes the risk of being killed, but the driver is carrying an Internalizer, so the photographer has actually imposed a risk onto the driver. If the photographer is also wearing an Internalizer, this should then save the driver, making the photographer once again the risk imposer. Without distinguishing assumers from imposers, this cycle could continue indefinitely. This gives us a second reason to care about risk assumption; we cannot make sense of risk imposition without it.

It seems that we are better off, in our capacity as both assumers and imposers of risk, if we are not obligated to carry Internalizers. In addition, it may be incoherent to suppose that all risks should be internalized. To understand moral responsibility, and how it might generate liability, we need to know more about the distinction between imposing and assuming risk.

To make this distinction, we could simply appeal to how risky people believe, or ought to believe, their acts to be. We could ask how much the Jay-walker ought to believe he has increased his risk of being run-over, and compare this to how much the speeding driver ought to believe he has increased his risk of hitting a pedestrian. This is one way of telling if the case is more one of risk assumption or risk imposition. Yet it is not clear, without further argument, why this difference in evidence should be morally significant. Moreover, I do not think it will be able to capture our

intuitions in these cases. While the sheer risk assumed or imposed matters, there are other things that matter as well.

For example, as we have seen, moral reasons can make a substantial difference. If the speeding driver is driving an ambulance, this makes it harder for us to judge that he is liable. Similarly, if the pedestrian is rushing across the street to save someone, then it is harder to claim that he has willingly assumed the risk. In the same way, if the speeding driver is a criminal fleeing from a crime, we are more inclined to consider him liable. If the pedestrian is crossing the street illegally, then we are more willing to claim that he has assumed the risk of being hit.

Other non-moral reasons may matter as well. Suppose that, given where he lives, the pedestrian is unable to leave his house without Jay-walking. This makes it harder to treat him as if he has assumed a risk by crossing the street. By contrast, if the Jay-walking could be easily avoided, if it were just a result of laziness, then it is easier to make the case that the pedestrian has assumed the risk. In all these cases I remain uncertain about how we should ultimately respond. My point is to draw attention to the numerous factors that could go into deciding whether we treat a case as one of assumption or imposition.

One final factor is the extent to which a person has adhered to, or transgressed, tacit or explicit conventions governing who and what ought to be where. These conventions matter because they give people good opportunities to avoid harm. This is vividly illustrated in *NASCAR*. Let us assume that getting a good picture is just as important for the photographer as racing is for the driver. We can also assume that the driver knows that photographers sometimes leap onto the track, and so correctly believes his driving to be just as risky as the photographer believes his jumping onto the track to be. From the perspective of reasons for action, and knowledge, these people are the same. Still, this is clearly more a case of risk assumption than risk imposition. This is because the conventions of the racetrack allow the drivers, and not the photographers, onto the track.

Putting all these factors together, we can see the underlying theme that connects them. We care about the person's evidence, the reasons they have for action, and whether they followed norms governing who ought to be where, all for the same reason. All these factors matter for the quality of opportunity each person had to avoid creating the risk, either as an imposer or an assumer.

If the assumer had a very good opportunity to avoid creating a risk to himself, while the imposer was obligated to impose the risk, then it makes sense to treat the case as one of risk assumption. In the same way, if the assumer had little choice but to take on the risk, while the imposer created the risk by doing something he was already obligated not to do, then it is easy to treat the case as one of risk imposition.

It is for this reason that the evidence people have about the riskiness of their acts matters morally. It matters once it is combined with the intuitive premise that the quality of opportunity a person has to avoid creating a risk to himself, or others, matters for whether we can fairly treat the case as one of risk assumption or imposition. Yet, this premise is not limited to beliefs and available evidence. It also takes account of the reasons people have for certain acts and the morally weighted costs of avoiding those acts.

The quality of opportunity a person has to avoid creating a risk does more than distinguish between cases of imposition and assumption. It also explains why we should care about these categories when thinking about liability. It is plausible to suppose that a person cannot reasonably object to internalizing the risks of an act he could easily avoid. Particularly when that act imposes risks that others cannot easily avoid. By contrast, if the act is morally obligatory, or would be very costly to avoid, then the person could more reasonably object that others should have to share the risks.

In this way, despite initial appearances, the case for risk imposition as a source of liability turns out to be an extension of the same significance of choice argument discussed above. If a risk

imposer has a good morally weighted opportunity to avoid imposing some risk, then this may undermine her ability to reasonably object to internalizing this risk. Similarly, if a risk assumer has a good morally weighted opportunity to avoid assuming some risk, then this may reduce her claim to be rescued.

This is significant, because if it is the quality of a person's choices and opportunities that distinguish imposers from assumers, and explains why these categories might matter for liability, then we have reason to doubt the significance of causal responsibility, even in cases that do not involve culpability. This is because the causal consequences of an act do not alter the quality of opportunity the person had to avoid causing those consequences. Once again, it seems to be choice, and not causation, that justifies liabilities.

It is still plausible that imposing a risk on others can sometimes justify a liability. However, we can now add two qualifications to this justification. The first is that many risks that are imposed are also assumed. If the case is more one of risk assumption, then the imposer need not accept liability. Distinguishing these cases requires us to look at the quality of opportunity these people had to avoid imposing or assuming the risks. The second qualification is that causal responsibility may not be indispensable to this justification for liability. This is because it is the quality of the person's choice, and the opportunity she had to choose otherwise, that is morally crucial.

It is sufficient for my purposes to argue, as I did in parts two and three, that culpability can justify liability without causal responsibility. Perhaps causal responsibility is crucial in cases of non-culpable risk imposition. Yet given the reasons we have for caring about risk imposition and assumption, I am unsure why this should be the case. Consider an equally far-fetched alternative to the Risk Internalizer,

Internalizer 2.0: These devices provide a service similar to that of the original Risk Internalizers. They save those threatened by risks that have been imposed but not assumed. However, unlike the original Risk Internalizers, the 2.0 model does not simply harm whichever person imposed the risk in question. Instead, the 2.0 model distributes this harm between risk imposers in proportion to the degree of risk they imposed and the quality of opportunity they had to avoid imposing it.

I am not convinced that we must all buy these new Internalizers either. However, there are some substantial advantages over the original model. Distributing the harms allows risk imposers to act as each other's insurer, leaving fewer people to suffer massive harms.

It also brings the distribution of these harms more in line with the reasons we have to care about risk imposition in the first place. Risk imposition is a grounds for liability when the risk is not assumed and the imposer had a good opportunity to avoid imposing the risk. The person's choices, and the evidence and reasons he had in making those choices, are of central moral importance. We at least have reason to doubt the centrality of causal responsibility in cases of liability for risk imposition.

4. Conclusion

In this chapter, I have been considering the significance of choice, for the justification of liabilities. In other words, I have been considering how our choices can make us morally susceptible to defensive harms: potential exceptions to the moral presumption against deliberately harming others and the presumption in favour minimizing the overall amount of harm. In the first half of the chapter I argued that culpably choosing to harm another can provide a clear justification for liability. Perhaps surprisingly, this justification does not depend on causal responsibility, and even suggests that causal responsibility should not make a difference to liability. By contrast, culpability seems to be directly linked to liability.

I then went on to consider other ways that our choices might make us liable to harm. Those who knowingly or negligently impose risks on others may thereby acquire a liability, even if the risk imposition was not culpable. This is because it would not be permissible to impose those risks on others without being willing to internalize the risks and accept liability. Yet this source of liability is more complex and limited than it might first appear.

We must distinguish between cases of risk imposition and risk assumption. To make this distinction we should appeal to the person's choices, and the opportunity they had to avoid imposing or assuming the risk. This suggests, once again, that choice may be more significant for liability than causal or moral responsibility. Moreover, not all risk imposition can be a basis for liability, even if that risk is not assumed. We need to consider what the risk imposer owes to others, and whether the permissibility of imposing the risk is conditional on accepting liability. These last conclusions are not definite. There may be many other things that make us liable, or contribute to the moral significance of our choices. In the next chapter I consider the possibility that it is our enforceable duties that explain and justify our liabilities.

Chapter Three:

Duties and Liabilities

Introduction

A duty, at first glance, looks much like a liability. Both can be reasons for costs to fall on one person rather than another. If doing my duty requires me to bear a cost, then I have a moral reason to bear this cost. I have a stronger moral reason to bear this cost, other things being equal, than another person whose duties do not require her to bear that cost. It seems plausible that in at least some cases third parties ought to direct that cost towards me, rather than towards somebody else, since I have a stronger moral reason to bring it on myself. This is particularly plausible in the case of enforceable duties: those duties that it is permissible to force the person to fulfill. Perhaps, as Victor Tadros has persuasively argued, we can look to enforceable duties to explain and justify liabilities.⁵⁹

In this chapter I explore the relationship between duties and liabilities. In the first section, I argue that duties do not simply generate liabilities. The relationship between duties and liabilities is closer to what a scientist might call correlation but not causation. In the second section, I consider what the duty not to harm others can tell us about liability to defensive harms. Thinking about our duties suggests that culpability is far more important for liability than causal responsibility. Finally, in the third section, I consider the implications of my arguments for the relationship between our duties, liabilities, and the Means Principle.

⁵⁹ Victor Tadros, *The Ends of Harm* (OUP, 2011) and “Causation, Culpability, and Liability,” *The Ethics of Self-Defense* ed. by Christian Coons and Michael Weber (OUP, 2016) 110-126.

1. The Relationship Between Duties and Liabilities

When a person, A, has a special duty to protect another person, B, from harm, diverting harms from B to A may be permissible because this allows A to fulfil her duty. As with other sorts of liability, it is not good for harms to befall A, it is just that it is better for harms to befall A than B. It is also better for A to be harmed in the protection of B than it is for a bystander to be harmed in the protection of B. This kind of duty can be generated simply by virtue of the relationship between A and B. Consider,

Father. Villain has kidnapped a small child, but is now cornered by the police and about to throw the child off a balcony. The child's father is nearby, as is an innocent bystander. Both are unconscious. You can a) throw the father underneath the balcony to cushion the child's fall, in which case the child will be unharmed and the father will suffer some broken bones, b) throw the bystander underneath the balcony, in which case the child will be unharmed and the bystander will break some bones, or c) do nothing in which case the child will be killed.

My intuition, in this case, is that (a) is clearly preferable. Let us assume, for the sake of argument, that the father would have a duty to throw himself under the falling child if he were able. This duty might be thought to explain why it is preferable, when the father is unable to act, to re-direct the harm towards the father. Doing so allows us to bring about the distribution of harm that the father would have been under a duty to bring about had he been conscious.

There are other special duties, besides those owed from parents to children, which may operate in the same way. Such duties might be generated by promises or other obligations. Here is an example:

Royal Guard: The members of the Royal Guard have sworn to protect the King. Villain attacks the King by rolling a large boulder down the corridor towards the throne room. You have only three options, a) divert the boulder into the waiting room, in which case it will break the ankle of a member of the public, b) divert the boulder into the guard room, in which case it will break the ankle of a member of the Royal Guard, or c) do nothing, in which case the boulder will roll into the throne room and squash the King.

The promise that the members of the Royal Guard have made to protect the King put them under a duty to bear costs in order to protect the King. Members of the public, by contrast, are mere innocent bystanders. My intuition, therefore, is that (b) is the best option.

In these cases, we can see that there is a strong connection between duties and liabilities. There are several ways in which we might understand this connection. It has been powerfully argued that all enforceable duties entail liabilities, and that duties can therefore be used to explain and justify our intuitions about liability.⁶⁰ In this section, I argue that this argument, while very appealing, overstates the strength of the connection between duties and liabilities. We have reason to doubt that all enforceable duties generate liabilities and that duties are what justify liability.

To see this, we should begin by examining the ways in which duties might be thought to generate liabilities. I argue that making this case requires us to accept some otherwise implausible assumptions and to posit some otherwise unintuitive duties. Doing so is not justified because there is a better and much simpler explanation available. The connection between duties and liabilities is better understood as a case of what scientists call correlation but not causation.

1.1 Simple Entailment

It might be thought that the presence of an enforceable duty simply entails or generates a liability. This seems plausible, and is admirably simple, but it falls afoul of an equally simple objection. It is not at all obvious, and is at the very least controversial, that people can have duties to do things that they are themselves unable to do. This is because it is generally accepted that in order for someone to have a moral reason to do something, they must be able to do that thing. The slogan ‘ought implies can’ captures this supposition.

⁶⁰ This view is defended by Victor Tadros. See *The Ends of Harm*, Chapter 6.

Consider the case of the father. We can assume that the father has a general duty to protect his child. We want to know whether this general duty makes it permissible to throw the unconscious father underneath the falling child. To answer this question, we must remember that the father is unconscious, and so is totally unable to throw himself underneath the falling child.

If it is true that 'ought implies can' it seems that the father's general duty to protect his child does not put him under a duty to throw himself under the falling child. This can be the case even if it remains true of the father, while he is unconscious, that he has a general duty to protect the child from harm. This general duty is a description of the father and his relationship with the child that is unaffected by the father being temporarily unconscious. The father retains a general duty to save his child from harm, because it may still be possible for him to do so in the future. What is affected by the father's incapacity is the extent to which this general duty can put the father under a specific duty to save the child from falling in this case.

Here is an analogy. We accept that the father has a general duty to protect his child. One thing that would certainly help is if the father could use a time machine to go back in time and prevent his child from being kidnapped. But this is impossible. Thus, it is at the very least strange to suppose that the father's general duty to protect his child puts him under a specific duty to go back in time and prevent the child from being kidnapped. While the father is under a duty to protect his child, he is not under a duty to go back in time.

This is analogous to the situation in *Father*. The father has a general duty to protect his child. However, since he is unconscious, he does not have a specific duty to throw himself under the falling child. Positing such a duty is no more plausible than supposing that the father has a duty to travel back in time. It may remain true of the father, while unconscious, that he has a general duty to protect his child. In the same way, it remains true that the father has a general duty to protect his

child despite the fact that he lacks a time machine. In neither case, however, is it intuitive to suppose that the father has a moral duty to perform a specific action that he is totally unable to perform.

Of course, this does not mean that it would not be permissible to throw the father under the falling child. It just means that we cannot easily explain the permissibility of throwing the father under the falling child purely in terms of his enforceable duties. This is because the father does not have an enforceable duty to throw himself under the falling child. We might posit such a duty if this is the best way to explain liabilities. We should acknowledge, however, that the intuitive case for such a duty is not good.

Perhaps we could claim that general duties can entail liabilities even in the absence of specific duties. This is not impossible. However, it is strange to suppose that I could be liable to bear some harm because of my duty, despite the fact that I am not actually under a duty to bear that harm. We would be forced to claim that the father's duty to the child makes him liable to be thrown under the child, despite the fact that the father does not have a duty to throw himself under the child. This strangeness gives us a reason to consider alternatives.

1.2 Hypothetical Consent

A second option is to consider what the people in these cases would or should consent to if they were capable.⁶¹ If people who would be obliged to bring harms on themselves if they were able can be assumed to consent to others imposing those harms on them, this hypothetical consent might then explain why these people are liable. Even if 'ought implies can,' and people who are unconscious or otherwise unable to bring harms on themselves cannot be said to have specific

⁶¹ I take this to be the main argument used by Tadros to justify and explain the relationship between duties and liabilities. See *The Ends of Harm*, Chapter 6.

duties, it is still true that these people are liable to be harmed because they would or ought to consent to those harms were they able to on account of their duty.

One way of understanding this argument is to see the intervening person as a tool the duty holding person could use to fulfill their duty. While it may be impossible for the father to throw himself under the falling child, it is possible for you – the intervening person – to do so. This means that the father can still fulfill his duty simply by empowering you, through real or hypothetical consent, to throw him under the falling child.

This is a promising solution, and it certainly seems intuitively plausible. Unfortunately, this way of understanding the relation between duties and liabilities faces a dilemma. First, let us assume that the father would consent to be thrown under the falling child if he were able. If he can be assumed to hypothetically consent to being thrown under the falling child, this would certainly explain why it would be permissible to throw him under the falling child.

It is less clear, however, whether this would make him liable to be harmed to save the child. When an unconscious accident victim is given emergency surgery this is made permissible by the fact that the victim can be reasonably assumed to hypothetically consent to the surgery. This does not mean, however, that the accident victim is liable to be operated on. Indeed, Tadros himself defines liability as a consideration that makes it permissible to impose harms on people who do not consent to them.⁶² Real or hypothetical can give us an independent reason to suppose that imposing the harm is permissible. This reason need not depend on duties or liabilities.

It might be objected that hypothetical consent alone cannot justify imposing harms that are supererogatory.⁶³ If this is true, then the father's hypothetical consent can only justify throwing him under the falling child if the father would have a duty to throw himself under the child. In this way,

⁶² Tadros, "Causation, Culpability, and Liability," 113.

⁶³ Tadros makes this point in *The Ends of Harm*, 135.

duties might still play a role in the justifying the permissibility of throwing the father under the falling child.

I am skeptical of the claim that hypothetical consent cannot justify imposing harms that are supererogatory. When a person cares deeply about something, we should respect what we know that person would want. This is true even if the person would want to make a supererogatory sacrifice.

Consider:

Uncle. An uncle has devoted his life to caring for his orphaned nephew. A villain has kidnapped the nephew and thrown him off a building. We can save the nephew, but only by throwing the unconscious uncle under it. If we do this the nephew will be saved but the uncle will suffer some broken bones.

Let us suppose that, were he conscious, it would be supererogatory for the uncle to throw himself under the nephew. He is under no duty to save the nephew at such a cost to himself. Yet, we also know that the uncle would not hesitate to sacrifice his bones in order to save the nephew. If he could, the uncle would beg us to throw his body under the falling child.

In this case, it seems to me, we have a good reason to respect the uncle's hypothetical wishes. Hypothetical consent alone can explain the permissibility of sacrificing a person like the father or the uncle, without reference to duties or liabilities. To know whether a person can be liable to be harmed on account of his duty, we need to consider how things would stand if the person would not consent to this were he able.

Suppose, therefore, that the father is frequently negligent in his duties towards the child. Perhaps he has even told us that he would never consent to be harmed even in defense of his child. In this case, although it may be true that the father would have a duty to throw himself under the child if he were able, we have good reason to suppose that he would refuse to do this.

It is tempting to assume that a person can be treated as if he had consented, on the grounds that he ought to consent. Even if this is a good assumption, it cannot explain cases like *Father*. This

is because it simply throws us back to the initial difficulty caused by ‘ought implies can.’ If the person is unconscious, and unable to give consent, then if ‘ought implies can’ it is not in fact true of that person that they ought to consent. It may be true of the father that he would be under a duty to consent if he were conscious, but that does not mean that he is under a duty to consent while he is unconscious.

In *The Ends of Harm*, Tadros argues that we cannot appeal to the absence of hypothetical consent as a way to justify not imposing harms on people who have a duty to bare those harms. He writes,

The fact that were I able to I would violate an enforceable duty by refusing to authorize you to use my body cannot compel you not to use my body ... If actual consent is not required for you to use my body as a means to save the child [then] hypothetical authorization is not required for you to use my body as a means to save the child where I am unconscious.⁶⁴

This seems correct. We should not refrain from sacrificing the father to save his child just because we know that he would refuse to do this if he were conscious. Hypothetical consent is, to this extent, irrelevant to the permissibility of sacrificing the father. Yet this does not explain why we should sacrifice the father. We are left looking for an explanation of the connection between duties and liabilities.

1.3 Enforceability

A third possibility is that it is the enforceability of these duties that connects them with liability.⁶⁵ We might suppose that when a duty is enforceable it is permissible to impose harms on people that they would have a duty to bring on themselves, if they were able. This might be

⁶⁴ Tadros, *The Ends of Harm*, 135

⁶⁵ Tadros makes it clear that only ‘enforceable’ duties produce liabilities. If ‘enforceable’ is stipulated to mean resulting in a liability, then some further explanation is required. We would need to know why some duties are enforceable and others are not.

tempting, but it would expand on the standard meaning of ‘enforceable.’ In most contexts, a duty is enforceable if it would be permissible to compel a person to fulfill that duty. We will have to see whether this further step is justified.

There are several reasons for doubt. First, when a person is capable of fulfilling their duty, we should first give them an opportunity to do so. Suppose that the father in *Father* is awake, aware of what is going on, and capable of throwing himself under the falling child. Time allowing, we should wait to see what the father will do before throwing him under the child ourselves. It would be wrong of us to push the father before he has had a chance to throw himself. Similarly, if the father is awake, but unable to throw himself under the child, we should first seek his permission before throwing him. We might suppose that enforceable duties cannot be enforced until the person has had a fair opportunity to fulfill the duty willingly.

This raises the possibility that it is culpability, and not the duty itself, that justifies imposing harms on people who have a duty to bear them.⁶⁶ Suppose that the father is able to throw himself under the child, but refuses to do so. It would then be permissible to throw him under the child without his consent. Yet we might be able to explain this simply by appealing to the father’s culpability. By culpably failing to throw himself under the falling child, the father has made himself liable to be sacrificed to protect the child. As we saw in Chapter Two, culpability is an important source of liability. Thus, if the father refused to do his duty, culpability alone could then explain the permissibility of throwing the father under the child.

The father’s duty has only an indirect role to play, by establishing the conditions in which the father might make himself culpable. This is particularly problematic when we return to the original case in which the father is unconscious. In this case it would be absurd to suppose that the father is culpable for failing to save the falling child. If culpability is part of what makes duties enforceable,

⁶⁶ This is Jeff McMahan’s preferred response to Tadros’s argument. In conversation.

then we have reason to doubt that the father's duty to his child can be permissibly enforced in this case.

A second reason for doubt arises from the difference between compelling a person to fulfill their duties, and imposing costs on people that they would have a duty to bring on themselves in some circumstances. These two things are very different. While enforceability clearly entails the permissibility of the former, there is no necessary connection between enforceability and the latter. In other words, there is no necessary connection between the costs that a person would have an enforceable duty to bring on themselves if they were able, and the costs it is permissible to impose on that person.

We can see this by returning to the example of the *Royal Guard*. In this case the members of the Royal Guard have a duty to protect the King because they have promised to do so. Let us now be more specific. Suppose that the guards all explicitly promise to sacrifice themselves to save the King whenever they are able. However, the guards all explicitly assert that they do not consent to have others impose costs on them to protect the King. There is nothing incoherent or illogical about the combination of this promise and this assertion. It might even make sense to suppose that the guards are willing to heroically sacrifice themselves but are unwilling to be passively sacrificed.

Now suppose that the Villain rolls his boulder down the hall towards the King, and the only way to save the King is to divert the boulder into the guard room. In this case, because of their promise, the guards would all have a duty to divert the boulder into the guard room if they were able. Moreover, we can assume that this duty is enforceable. If the guards refused to sacrifice themselves, it would be permissible for us to threaten them or otherwise compel them to fulfill their duty. Yet if none of the guards is in a position to divert the boulder into the guard room, and if we can assume that none of them is able under the circumstances to authorize someone else to divert it,

then it would not be permissible for a third party or for the King to do so. This is because of the terms of the promise and assertion that establish the limits of the guards' duty to the King.

This is an admittedly contrived example. Yet it is nevertheless significant. It shows that there is no necessary connection between the costs we would be obliged to bring on ourselves if we were able, and the costs others can permissibly impose on us. While other cases, such as *Father*, suggest that there is a strong correlation between these two things, we have reason to doubt that the relation between duties and liabilities is deeper than this correlation. The mere possibility of cases like this show that enforceable duties do not entail liabilities.

1.4 Correlation but Not Causation

I agree that there is a close connection between enforceable duties and liabilities. The fact that a person is, or would be, under a duty to impose a harm on themselves can be a good clue that the person is also liable to suffer that harm. This liability does not follow, however, from the simple fact that the person has a duty. Rather it follows from whatever underlying factors generate that duty. In *Father*, the father's relationship to the child can explain both his duty to the child and his liability to be harmed in defense of the child. This relationship, unlike his specific duties, clearly endures even when the father is unconscious. There may be other cases, such as *Royal Guard*, where the underlying factors that generated the duty do not also generate a liability, or more accurately where the scope of the liability and the scope of the duty diverge.

We can further explain this account by examining a case presented by Tadros:

Pond Rescue: A child is drowning in a pond. To save the child from drowning I must enter the pond. If I do this, the child will be saved but I will lose a finger.⁶⁷

⁶⁷ Tadros, *The Ends of Harm*, 130.

Let us suppose that I am obliged to rescue the child, even though it will cost me my finger. I am permitted to give my own interests extra weight when deciding what I should do. If saving the child required me to sacrifice my own life, or even my leg, we can assume that I would not be obliged to do it. However, I cannot completely ignore the interests of others. In this case, the harm that would befall the child if I did nothing is so much greater than the harm I would suffer by preventing it that I am obliged to save the child.

Now suppose that I am unconscious, and therefore unable to jump into the pond or consent to being thrown into the pond.⁶⁸ Nevertheless, it is intuitive to suppose that I may be permissibly thrown into the pond, at the cost of my finger, if this is the only way to save the child. The same explanation can be given for why this is the case. When deciding what it is permissible to do to me we should be particularly averse to using my body to save others without my consent. This is because I should be allowed to decide how my body is used, and I should be allowed to use my body to pursue my own ends. However, we must also pay attention to the interests of the child. Even if we give my interests extra weight, on the grounds that we would be using my body without my consent, we might still conclude that preventing the harm of death from befalling the child is more morally important than allowing me to decide whether my finger is sacrificed to save the child.

A nearly identical moral calculus occurs in each of these cases. The fact that it is more morally important to save the child from death than to allow me to keep my finger is exactly the same reason that I am obliged, and not merely permitted, to sacrifice myself to save the child if I am able. In each case this imposition of costs can be justified by a sufficiently weighty moral reason: namely, the badness of death and the comparative insignificance of a lost finger.

In this way, the harm I have a duty to bring on myself may correlate with the harm it is permissible to inflict on me without my consent. This need not mean that my duty is what explains

⁶⁸ This variation is considered by Tadros in *The Ends of Harm*, 131.

this permission. As I have discussed, some hard cases show that this explanation is problematic. Rather, the simplest explanation is that my duty to bring a harm on myself, and the permissibility of imposing that harm on me, can each be explained by the same underlying moral reasons.

To be clear, I do not mean to claim that a liability is what makes it permissible to sacrifice me to save the child in *Pond Rescue*. I find it clearer to suppose that a liability is a consideration that makes a person particularly susceptible to bear costs. In rescue cases of this sort there is nothing special about the person that makes them particularly susceptible to bear costs. The significance of this case is that it shows how people can be permissibly sacrificed even when they are not particularly morally susceptible to bear costs. My point here is merely to show that we can explain the link between what I am obliged to do and what can be permissibly done to me by appealing to underlying moral reasons, without supposing that my obligations explain or justify these permissions.

A similar point can be made using cases that involve liabilities. Consider, for example, the case of the father. In this case we have assumed that the parental relationship the father has with his child puts the father under an obligation to save the child if he is able. This is an intuitive thought, though it is not entirely clear why this should be. Perhaps parents owe their children protection because parents are responsible for creating such vulnerable creatures.

Whatever the underlying reason, this obligation is not unlimited. It would not be wrong of the father to allow his child to suffer a strong pinch if the only way to prevent the pinch would cost the father a week of torture. We can imagine the sort of deliberation the father ought to undertake when deciding whether he ought to accept some harm so as to save the child. He must weigh the moral reasons generated by his relationship to the child, as well as the comparative badness of different quantities of suffering, and give each of these considerations its proper weight.

A similar deliberation is what we should undertake when deciding whether to throw the father's unconscious body under the falling child. We take the same morally significant considerations into account and give each its proper weight in order to decide what we ought to do. This explanation is more plausible because it does not require us to solve difficult questions about whether 'ought implies can' or accept the premise that it is always permissible to inflict a harm on a person if that person would be obliged to inflict the harm on himself if he were able. This approach is equally parsimonious because the same underlying moral reasons that can explain both duties and liabilities are needed anyway to explain duties, even if those duties can then explain the liabilities.

Moreover, the correlation view I have proposed has the advantage of flexibility. It can explain why some liabilities correlate neatly with duties, but it can also explain why they sometimes diverge. Overall, this flexibility allows the correlation view to be more parsimonious because it does not require us to posit otherwise unintuitive duties to explain intuitive liabilities. It is intuitive to suppose that highly culpable people can be permissibly sacrificed to save innocent people from harm. It is not nearly as intuitive to suppose that highly culpable people have an equally strong duty to sacrifice themselves for the sake of innocent people. Culpable people will have some duties, but I am not confident that these duties will always align perfectly with their liabilities. We can at least conclude that a specific duty is neither a necessary nor sufficient condition for a liability.

2. Duties and Self-Defense

Enforceable duties do not entail liabilities, so we cannot simply rely on duties to explain and justify our theory of liabilities. However, there does seem to be a close connection between the sorts of things that generate liabilities and the sorts of things that generate duties. This means that investigating our duties can still be a good starting point for a theory of self-defense. In this section I will argue that in self-defense cases our actual duties and our liabilities often diverge. Nevertheless,

thinking about duties can provide further support for the culpability based theory of liabilities I have been developing.

Suppose we are all under a strict enforceable duty not to cause unjustified harm to others. This duty seems like a plausible starting point for a theory of self-defense. If an attacker does launch an unjustified threat of harm to the victim, then the attacker is under a duty to prevent this threat from harming the victim. This is simply an extension of her duty not to cause unjustified harm. The attacker's duty may not simply entail a liability, but it could be that the same thing that generates this duty also generates her liability. Our duty not to cause harm to others could provide a good starting point for explaining liabilities.

2.1 Preventing Innocent Threats

Many of the reasons we have not to cause unjustified harm apply even when we are blamelessly ignorant. Consider,

Light Switch: A person, Switcher, innocently turns on a light switch when she gets home from work. Unforeseeably, turning this switch will cause Victim to be electrocuted unless Switcher is sacrificed instead.

We can assume, for the sake of argument, that Switcher had a duty not to flip the switch, even though she could not possibly be aware of this fact. Perhaps Switcher now has a stringent duty to prevent the threat of unjust harm she has caused. Such a duty seems like a simple extension of her fact-relative duty not to cause harm in the first place. If liabilities tend to track duties, this suggests that Switcher should be liable to be sacrificed. This line of reasoning suggests that causal responsibility, and not culpability, is what matters.

My intuition in *Light Switch* is that Switcher is not liable to be harmed in this case. This can be explained using the arguments from Chapter Two. However, we can also use duties to help explain why this is the case. Consider the following pair of cases:

Floodgates I: Agent is in the control room of a factory, and is in the process of opening the floodgates. Although she could not possibly be aware, Victim is trapped in the discharge pipe and will be drowned if Agent opens the gates. Agent must pull two levers to open the gates. After she has pulled the first lever, Agent learns that Victim is trapped in the pipe.

In this case it is clear that Agent ought not to pull the second lever. Given what she now knows, pulling the lever would amount to knowingly killing Victim. Agent has a strong duty not to do this. If necessary, she must allow herself to be badly hurt or even killed to avoid pulling the second lever.

Things look very different when we turn to:

Floodgates II: Same as *Floodgates* except that opening the gates only requires pulling one lever. There is then a 10-minute delay during which pushing the Emergency Stop will prevent the gates from opening. Just after pulling the lever, Agent learns that Victim is trapped in the pipe.

Agent clearly has a duty to press the Emergency Stop. However, this duty does not appear nearly as strong as Agent's duty not to pull the second lever in *Floodgates I*. If pushing the Emergency Stop would cause Agent to be badly hurt or killed, then it would certainly be permissible for her to refrain from pushing it. Agent's duty in this case is more like a duty to rescue Victim from harm. This is despite the fact that Agent's action, in pulling the lever, was causally responsible for the threat that Victim now faces.

Floodgates II shows that our duty to prevent the threats that we cause is not always the same as our duty to not create threats of harm. Pushing the Emergency Stop is the only way Agent can prevent herself from being causally responsible for an unjust harm to Victim. This alone does not tell us how strong Agent's duty to press the Emergency Stop is. Agent's duty to press the Emergency Stop is not just an extension of her duty not to cause threats of unjust harm. If it were, then her duty to press the Stop would be as strong as her duty not to pull the second lever in

Floodgates I. We need to look at the epistemic position of Agent, and the quality of the opportunity she had to avoid creating the threat of harm.

Since our duties and liabilities tend to go together, this conclusion may tell us something about liabilities. We can conclude that we can only impose small harms on Agent in *Floodgate II*, if this is necessary to save Victim. The harm we can permissibly apply may be roughly similar to the harm we could expect Agent to bear in order to rescue another in need. On the other hand, if Agent tries to pull the second lever in *Floodgate I*, despite knowing that doing so will kill Victim, then we can permissibly impose large harms on Agent in order to save Victim.

Returning to *Light Switch* we can now give a duty based explanation for why Switcher is not liable in this case. *Light Switch* is clearly analogous to *Floodgates II*. If our conclusions are correct in *Floodgates II* then it is impermissible to sacrifice Switcher to save Victim in *Light Switch*. The only harms we can apply to Switcher are those roughly equivalent to the sacrifice required by a duty of easy rescue.

It should not matter whether Switcher is about to flip the switch or has just flipped the switch. These cases are so similar that it would be unreasonable for Switcher to be permissibly sacrificed if she were about to flip the switch but unable to be sacrificed if she had already switched it. The same moral considerations clearly apply in each case.

Thus, even if Switcher is about to flip the switch and electrocute Victim, it is impermissible to kill Switcher to save Victim. This is because, by virtue of Switcher's blameless ignorance, her duty to prevent herself from killing Victim in this case is roughly equivalent in strength to her duty to rescue Victim from harms. The comparative weakness of this duty suggests that Switcher is not liable to be killed to save Victim. This is despite the fact Switcher has a very strong fact-relative obligation not to flip the switch.

In this way, despite initial appearances, our duties not to harm others do not simply extend to explain our duty to prevent ourselves from causing harm to others. In cases where people cause threats of which they are blamelessly unaware, these two duties come apart. To the extent that duties correlate with liabilities, this suggests that the quality of opportunity we have to avoid creating threats to others matters when determining our liabilities. Thinking about duties can provide further support for the conclusions we reached following the significance of choice argument in Chapter Two. However, there may be limits to how far this support goes.

2.2 Causal Responsibility

While our duties support the view that causal responsibility for a threat is insufficient for liability, they also seem to suggest that causal responsibility is necessary for liability. This is because my duty to prevent threats that I culpably cause is much stronger than my duty to save victims from other threats. If I culpably order an assassin to shoot someone, then I have a very strong duty to stop the assassin from killing that person. The strength of this duty should be roughly the same as the strength of my duty not order the assassin to kill them in the first place. By contrast, if I just happen to see someone about to be assassinated, my duty to rescue that person is much weaker, equivalent to my duty to rescue any person.

This seems true even if I have myself just culpably ordered an assassin to kill someone. Although I am highly culpable for trying to cause the very same sort of harm, it is not plausible to suppose that I suddenly acquire a very strong duty to rescue others from harm. At least, such a duty is not part of common sense morality in these cases. To the extent that our duties tend to correlate with our liabilities, this suggests that causal responsibility is a necessary condition of liability.

This alone does not prove that we should reject the culpability based view I have proposed. As I argued in the last chapter, the justifications for liability do not require any sort of causal

responsibility. The significance of costs argument even suggests that it would be wrong to appeal to causal responsibility in these cases. These arguments give us good reason to suppose that casual responsibility does not matter, regardless of what our duties might suggest. After all, the relation between our duties and our liabilities is one of correlation, not entailment. This may simply be a case where our duties and liabilities diverge.

Indeed, the ability to allow for this divergence may actually be a strength of the correlation view. This is because, as I will argue, in many cases it is highly intuitive to suppose that causal responsibility is not a necessary condition of liability. It is much better if we can accept these intuitions without positing unintuitive duties. We can see this by considering a series of cases originally posed by Tadros:

Hit Man: Villain culpably orders Hit Man to kill Victim. The only way Victim can save herself from Hit Man's bullet is by using Villain as a human shield.⁶⁹

In this case it is perfectly clear that Victim can sacrifice Villain to save herself from being killed.

Now consider:

Double Hit Man: Villain culpably orders Hit Man to kill Victim. At the same time another Villain orders another Hit Man to kill Victim. The only way Victim can save herself is by using both Villains as a human shield.⁷⁰

Once again, my intuition is that Victim can permissibly use both Villains as a human shield. Yet if causal responsibility for a threat were a necessary condition of liability to prevent that threat, then Victim could only save herself if she were sure that the bullet from each Hit Man would be stopped by the body of the Villain that hired him. This seems like an unreasonable burden to place on Victim's defensive action. If we examine the ballistics and learn that the Villains were actually hit by

⁶⁹ This is an adapted version of Tadros's influential case of the same name. See *The Ends of Harm* 187.

⁷⁰ Again, this is an adapted version of Tadros's case. *The Ends of Harm* 192.

the wrong bullets, this should not lead us to conclude that Victim has wrongfully sacrificed the Villains.

At first glance, these intuitions present a problem for the duty view of liability. This is because although causal responsibility does not seem to limit the liability of the villains in this case, it does seem to limit their duties. Each Villain has a very strong duty to prevent the threat that they are culpably responsible for. However, they only have a relatively weak duty to rescue the Victim from a threat they did not cause.

Tadros offers two solutions to the problem posed by *Double Hit Man*.⁷¹ The first is to suggest that people who culpably attempt to harm others thereby acquire a range of duties to save others from harm. In this way, the two Villains might actually have strong duties to save the Victim from threats they did not cause, simply by virtue of their culpability. This approach is reasonable since liabilities correlate with duties. The fact that the two Villains are liable to be killed to save Victim from either threat might suggest that these Villains actually do have duties to prevent either threat. The weakness of this solution is that we do not have much other evidence for the existence of these duties. Positing them is therefore a somewhat ad hoc solution to the problem.

The second solution is to appeal to the duties the two Villains might have to make an agreement with each other. We need only assume that each Villain has a powerful duty only to prevent the threat that she caused. Now consider what we can expect each Villain to do to fulfill this duty. Suppose the Villains realize that neither can stop the bullet that they are causally responsible for. However, they could throw themselves in front of each other's bullets, and thereby save the Victim. Tadros suggests that in this case the two Villains ought to make an agreement to sacrifice themselves to prevent the bullets from killing the Victim.

⁷¹ See Tadros, *The Ends of Harm*, Chapter 8 part IV.

Making this agreement is the only way they can each fulfill their duty to save Victim from the bullet that they are causally responsible for. In this way, the duty to prevent the threat that they caused can give them a further duty to prevent a threat that they did not cause. Tadros concludes that we can permissibly treat the two Villains as if they had made such an agreement, since they have a duty to make this agreement if they could.

This is an ingenious solution to the problem, but it faces a few substantial objections. First, it might be the case that the two Villains cannot in fact communicate with each other in order to make this agreement. This inability to communicate should not prevent them from being sacrificed to save the Victim. Yet, if ‘ought implies can’ then it does prevent them from actually having a duty to make an agreement to prevent each other’s threats.

Second, even if the two Villains could communicate, I am not sure that they actually have a duty to literally make such an agreement. If it simply did not occur to the two Villains to make such an agreement, I do not think that they would be culpable for this failure. Thus, while I agree that it is permissible to treat the villains as if they had made such an agreement, I am not sure that they have an actual duty to make such an agreement.

Third, the case only has to be slightly modified to prevent this solution from working. Suppose that moments before making the agreement it becomes clear that one of the bullets is going to miss. In this case the Villain whose bullet is going to miss no longer has a reason to accept the agreement. Yet, to my intuitions at least, it should still be permissible to use this Villain to save the victim.

Alternately, suppose that the two bullets will arrive at different times, so that one Villain will have to be sacrificed slightly before the other. After the bullet he is causally responsible for has been stopped, what reason does the second Villain have to stick to his agreement and stop the remaining

bullet. This villain's duty to stick to the agreement is no longer an extension of his duty to prevent the threat that he caused, since that threat has already been prevented.

At best, the Villain's duty to stick to the agreement is an extension of his duty to keep his promises. This may be a weaker duty that Villain need not sacrifice his life to avoid breaking. Moreover, in realistic cases it is unlikely that the Villains will have literally made a promise to each other, or even have an opportunity to do so. It seems particularly unlikely that we can permissibly kill a person in order to enforce their duty to keep a promise that we merely assumed they would be obligated to make if they could.

Thinking about the agreements that the villains might be obliged to make offers a promising solution to the problem for the duty view posed by *Double Hit Man*. Yet we have reason to doubt whether this solution is successful. While the two Villains ought to agree to sacrifice themselves to prevent each other's threats, it is not clear that they actually have a duty to literally make such an agreement in real cases. There will be some cases where literally making such an agreement is not possible. Moreover, there are variations of *Double Hit Man* where it still seems permissible to sacrifice the two villains, despite the fact that the two villains have even less reason to make and keep an agreement with each other.

2.3 Hypothetical Duties to Agree

We can reinterpret this account of *Double Hit Man* to avoid these problems and to provide further support for the culpability based view of liability I have defended. Tadros is committed to supposing that the villains have an actual duty to make an agreement with each other, because on his view it is our actual duties that generate our liabilities. This is what gets the account into trouble. In many cases, is not clear that the villains have such a duty. However, there is certainly a sense in which the villains ought to agree to prevent each other's harms if such an agreement is proposed.

Perhaps we do not need to suppose that this is an actual duty, meaning an obligation to do or refrain from doing something in the actual world. Rather, we could focus instead on what we might call a hypothetical duty to agree. One way to model moral deliberation is to imagine a conversation between the different people involved. A hypothetical duty to agree is a claim about the sorts of moral rules that people ought to accept in this kind of deliberation. This does not mean that these people have a duty to make an actual agreement. It simply means that in theoretical deliberations they would be obligated to accept such an agreement. In other words, such an agreement is not something that the villains can reasonably reject.

Moving from actual duties to hypothetical duties to agree can resolve the first two objections I introduced above. Whether or not the villains can actually make an agreement with each other does not matter because we are no longer interested in their actual duties. Each villain's duty to prevent her own threat gives her a hypothetical duty to agree to being used to stop the other's threat. This is because agreeing to swap threats may be the only way each villain can prevent herself from culpably causing unjust harm to the victim. In this way, each villain's duty to stop her own bullet can help explain why she cannot reasonably object to being used to stop the other's bullet.

This can be the case even if it turns out that one of the bullets is going to miss. Suppose the villains know that one of the bullets might miss, but they do not know which it will be. They still ought to agree to be sacrificed, if necessary, to prevent each other's bullet. Making such an agreement in advance is the only way they can both be sure of not killing the victim. Neither of them can permissibly rely on the hope that his bullet will miss. Thus, they both have a hypothetical duty to agree to be sacrificed to stop either bullet, even if one of them is going to miss, because making this agreement gives each of them the best chance of not killing the victim.

In addition, we do not need to accept that a villain who sees that his bullet is going to miss still has a duty to save the victim. All we need to accept is that this villain at some point had a

hypothetical duty to agree to be sacrificed, if necessary, under such circumstances. He has to agree to this in order to avoid the possibility that his bullet will not miss and the other villain will not be sacrificed to stop it. In the same way, it does not matter if the bullets will arrive at different times, because an arrangement that sacrifices both villains one at a time is not something that either villain can reasonably object to. By switching from actual duties to hypothetical duties to agree, we can resolve the three objections I raised to this solution to *Double Hit Man*.

This argument can be extended to provide further support for my culpability based view of liability. After all, it does not matter that the of the two villains in *Double Hit Man* are threatening the same person. If there were two victims, who could only be saved by swapping the two villains, this swapping is not something that either villain could reasonably reject. This is because the villains would have hypothetical duty to agree to such a proposal.

We can go further still, and see that all people who culpably threaten to harm others can improve their chances of not causing unjust harms by agreeing in advance to a principle that allows any culpable threatener to be sacrificed, if necessary, to save any potential victim. Such a principle is not one that any culpable threatener, or potential culpable threatener, could reasonably reject. Each person has a hypothetical duty to agree to this swapping arrangement, because they each have a stringent duty to prevent themselves from causing harm. Making this arrangement gives them each the best chance to avoid causing harm.

The amount of harm each culpable person cannot reject, as part of such an agreement, should be roughly equal to the amount of harm they are obliged to accept in order to prevent themselves from causing harm. Thus, culpable people must accept large harms, although the amount of harm they can reasonably reject in a given circumstance depends on their level of culpability. By contrast, as we have seen, non-culpable people need accept no more harm than they must accept in order to provide easy rescues.

In none of these agreements does actual causal responsibility matter. This is surprising because each person only has a stringent duty to prevent themselves from causing harm. However, this stringent duty gives them each a hypothetical duty to agree to a system where culpable threateners can be sacrificed to prevent threats that they did not cause. In fact, the duty to prevent themselves from causing harm gives each culpable person a hypothetical duty to agree to a system where a culpable threatener can be sacrificed even if that threatener's own threat is going to miss or has already been blocked. These swapping agreements are not something that anyone can reasonably reject because they give us all the best chance to fulfill our stringent duty to prevent ourselves from culpably causing unjust harms to others.

These hypothetical duties to agree can also show us how our duties and liabilities can sometimes diverge. In cases like *Double Hit Man* each villain only has an actual duty to prevent the threat he has caused. Nevertheless, partially because of this duty, each villain must accept that he is liable to be harmed to prevent either threat. Causation clearly matters for our duties to avoid harming others, but it need not matter for our liabilities. Such a divergence is possible because duties do not entail liabilities, they merely tend to correlate with them.

Inevitably, some threateners will get lucky and others will get unlucky. Depending on the circumstances, some people might have the threats they cause blocked by others without being sacrificed themselves. Other people might miss and still be sacrificed to prevent a threat they did not cause. However, the possibility of this luck is something that these people cannot reasonably object to. This is because they all have a hypothetical duty to agree to a system of swapping harms, even though this system will occasionally result in this kind of luck. They must agree to this system because they cannot permissibly accept the risk that they will culpably cause a harm to others in order to avoid the risk of being harmed to prevent a threat they did not cause. In other words,

culpable people must accept an increase in the chances that they will suffer defensive harms, in return for a decrease in the chances that they will culpably cause harm to others.

I do not wish to rely too heavily on this argument in defending the culpability based view of liability. This argument is not necessary, because the significance of choice argument I introduced in Chapter Two is sufficient to show that culpability can explain and justify our liabilities. Indeed, hypothetical duties to agree may be best understood as a further support for the significance of costs argument. The fact that culpable people have a stringent duty to prevent themselves from causing harm can give us an additional reason to suppose that culpable people cannot reasonably object to a moral rule that treats them liable to defensive harming.

The main role of the argument in this section is to confront the proponents of the duty view with a dilemma. If they reject hypothetical duties to agree, and stick to actual duties, then they are left without a good way to explain the relationship between duties and liabilities. They are also stuck with accepting unintuitive conclusions in cases like *Double Hit Man*. On the other hand, if they accept hypothetical duties to agree then they should accept that culpability is the key ingredient in liability to defensive harming, and that our actual duties and our liabilities sometimes diverge.

3. Demandingness and the Means Principle

It might be objected that in my discussion of the duty view of liability I have neglected to consider its key advantage. This is its purported ability to explain neatly the relationship between liabilities and the Means Principle. Accepting that actual duties merely correlate with our liabilities may force us to forego this advantage. This might simply be a cost we have to accept. Morality may not be as tidy as we would like it to be, and our theories should not make it appear more tidy than it is. However, I believe that my culpability view allows for essentially the same approach to the Means Principle as the duty view.

I will not try to offer a theory or definition of the Means Principle. I will only try to describe it in general terms that should be applicable to most theories. The Means Principle suggests that it is particularly morally problematic, and in most cases wrong, to use another person's resources without their consent. A complete theory of the Means Principle should tell us first, what each person's resources are, second, what sorts of acts and omissions count as using these resources, and third, how strong the prohibition is on using these resources without consent. Almost all recent work on the Means Principle has focused on the perplexing questions raised by the second of these three parts. This can leave us with the impression that the Means Principle is uniquely concerned with the distinction between using and not-using. However, it is important to bear the other parts in mind as well.

Tadros argues that when we have a duty to use our resources in a certain way it does not violate the Means Principle if another person takes those resources to use them in that way.⁷² Thus, if duties entail liabilities, then we have a neat way of showing how liabilities can be an exception to the Means Principle. This solution allows us to defend victims and impose punishments without violating the Means Principle. Unfortunately, as I have argued, it is not the case that actual duties entail liabilities. Moreover, as cases like *Royal Guard* show it is not always the case that we can simply use another's resources in a way that they would have a duty to use them.

However, I believe that the correlation view takes basically the same approach to the Means Principle as that provided by the duty view. Tadros begins this argument with our duties already established. We have an obligation to save the child in *Pond Rescue*, and conclusions about our liabilities are derived from there. Whatever these liabilities are, on Tadros's view, enforcing them does not violate the Means Principle. To see the similarity with the correlation view, we only need to look one step further back in our deliberations.

⁷² Tadros, *The Ends of Harm*, Chapter 6.

Consider the step where we decide what our duties are. In *Pond Rescue* we must weigh the cost of scarifying a finger against the cost of allowing a child to drown. Intuitively, we might conclude that we are obligated to sacrifice a finger, but not a leg, to save the child from drowning. We are not obliged to rescue the child at anything near the cost of our own life. Clearly there is some extra-utilitarian principle at work in our thinking about this case. Let us call this the Demandingness Principle. This principle suggests that it is particularly problematic to require a person to use her resources in a way that, were she not required, she would not consent to.

When deciding what our duties are, in cases like *Pond Rescue*, this principle is an input into our deliberations. In requiring the person to make some sacrifice in *Pond Rescue* we are in some sense going against Demandingness Principle, but this is justified by the balance of reasons. We might reasonably conclude that the person is obligated to save the drowning child at the cost of a finger, but not a leg.

Now consider the correlative liability. Tadros argues that we can forcibly sacrifice a person's finger, but not her leg, without violating the Means Principle. He explains this by supposing that the person's duties generate her liabilities. But we do not need this intervening step to reach the same conclusion.

We can reach the same conclusion if we recognize that the Demandingness Principle has a similar force in our deliberations about duties as the Means Principle has in our deliberations about liabilities. When a person is liable to some harm, we can impose that harm without violating the Means Principle. In doing so we are in some sense going against the Means Principle, just as imposing a duty goes against the Demandingness Principle, but this can be justified by the balance of reasons. It is not desirable to disguise the conflict between reasons that is inherent in our moral deliberations. We should acknowledge that imposing liabilities can go against the Means Principle, just as imposing duties goes against the Demandingness Principle.

We should not be surprised that the same sorts of considerations that can outweigh the Demandingness Principle, when deciding the extent of our duties, can be expected to outweigh the Means Principle when deciding the extent of our liabilities. After all, the two principles are very similar. All duties are duties to use one's resources in a certain way regardless of whether one would otherwise chose to do so. We try to limit the demandingness of these duties, but there is a limit to how strong this limit can be. The Demandingness Principle seems to involve all the same elements as the Means Principle. It also seems to play a similar role in our moral deliberations.

Perhaps this is because these are just two versions of the same principle, or perhaps it is because they are both further justified by the same underlying consideration. Whatever the deeper explanation, there is clearly a close relation between these two principles, both in their reason giving force and in the direction that force is applied. More work should be done to investigate this connection, and see what underlying considerations support these sister principles.⁷³

This is one way of understanding Tadros's key insight in *The Ends of Harm*. Tadros recognizes a close correlation between the demandingness of our duties and the extent of our liabilities. He describes this close connection by supposing that our duties generate our liabilities. Yet, for the somewhat technical reasons described above, this is not quite correct. Instead, what I think Tadros has recognized is the close connection between the Demandingness Principle and the Means Principle.

This reinterpretation is supported by the conclusions about duties and liabilities I reached above. In multiple cases we saw that our actual duties can diverge from our liabilities. Yet they diverge because our actual duties sometimes did not apply, or were sometimes limited to more specific goals. By contrast, as we saw, our actual duties and liabilities nearly always correlate in their potential demandingness, even if they do not always correlate in their application. This conclusion

⁷³ I have undertaken some of this work in my paper "Using, Demanding, Doing and Allowing."

can tell us about the relationship between duties and liabilities, but it can also help us to better understand the Means Principle.

4. Conclusion

There is clearly a very close relationship between duties and liabilities. In this chapter I argued that our duties do not simply entail, or generate, our liabilities. Instead, I suggested that many of the same underlying moral considerations that generate duties also generate liabilities. For this reason, our duties and liabilities tend to correlate in their stringency, even if they are not always the same in their scope. The correlation between duties and liabilities means that duties can give us a clue as to the strength and extent of liabilities to defensive harming. In part two of this chapter I argued that thinking about duties can give us reason to doubt the significance of causal responsibility in cases of liability to defensive harming. This adds to the reasons provided by the significance of choice argument. In the next chapter, I consider the relationship between causation, culpability, and liability more fully. I argue that causal responsibility for a threat of harm is not a necessary condition for liability to be harmed in prevention of that threat.

Chapter Four:

Culpability, Causation, and Liability

Introduction

Through our actions, we can cause things to occur in the world. This must be a tacit premise in any discussion of defensive harming. Causal responsibility, the way in which we must own those things we knowingly cause to occur, is a central part of our moral lives. It is part of the way we think of ourselves, and it carries a host of moral implications. Any account of morality, or of defensive harming, must reserve a central place for causation and causal responsibility. Still, as I started to argue in the last few chapters, causal responsibility may be far less important for the justification of liability than many people currently believe.

In this chapter I discuss what I call the Causal Connection Requirement. This is the widely held assumption that a person can only be liable to be harmed in the prevention of some threat if that person is causally responsible for that threat. I sometimes abbreviate this by describing it as the assumption that causal responsibility is a necessary condition of liability. In the first two parts of this chapter I argue that the Causal Connection Requirement is false. In part one, I offer a series of different arguments against this requirement. In part two, I offer a further argument based on the preemptive nature of much defensive harming. In these cases, a person may be permissibly harmed although they have not yet caused a threat of harm. I argue that it is particularly implausible to follow the Causal Connection Requirement in these cases.

It is sometimes argued that the Causal Connection Requirement should be maintained, despite its counterintuitive implications, because it is needed to restrict the scope of liability. If culpability alone is sufficient to generate liability, it is worried, then a single instance of culpability can result in a lifetime of liability. In the last two sections of this chapter I discuss and reply to this

objection. I argue that we should not adopt a counterintuitive restriction on liability until we have looked more closely at culpability itself. There is good reason to believe that the nature of culpability can provide a natural limit to the scope of liability.

1. The Causal Connection Requirement

According to the Causal Connection Requirement, causal responsibility for some threat is a necessary condition of liability to be harmed in the prevention of that threat. This requirement is a widely accepted and often unchallenged assumption in the ethics of self-defense. It is only on closer inspection, and in more complex cases, that this initial plausibility disappears. In this section I will offer several arguments to show that, despite its initial common-sense plausibility, the Causal Connection Requirement should be rejected. Taken together, these arguments make it clear that we should reject the Causal Connection Requirement.

1.1. Swapping Cases

In Chapter Two, I argued that choices, particularly culpable choices, can generate liabilities. When two people have attempted the same sort of action, and are equally culpable, we should treat each of them in the same way. This is a simple principle of fairness, and it is captured by proportionality. It should not matter whether each culpable person only receives a defensive harm that prevents a threat that he caused. We have already seen this in cases like *Firing Squad* and *Double Hit Man*.⁷⁴ Ensuring a proper pairing of villains and threats seems, in these cases, like a moral technicality. What matters is saving the life of the victim, and not imposing any disproportionate or unnecessary harms on the villains. If doing this requires the villains to swap threats, then this is what we should do.

⁷⁴ See my discussion of these cases in Chapter Two and Chapter Three respectively.

In the last two chapters I considered the theory behind these cases in some detail. It might be thought that it is permissible to swap the villains in *Double Hit Man* because each villain has a duty to agree to this swapping if it is the best way to save the victim. I argued that such a duty need not be an actual duty but could be understood as a claim about what sorts of moral rules the villains cannot reasonably reject. The villains might have a duty to agree to swapping because this gives them the best chance to avoid culpably causing harm to others.

My preferred view is that villains cannot reasonably object to a rule that allows them to be swapped, for the same reasons that they cannot reasonably object to a rule that allows self-defense. This is because the morally weighted costs of avoiding liability under such a rule are very low, while the costs to victims of forbidding self-defense and swapping are very high. There should be no difference between successfully attempting to pose a threat of harm and unsuccessfully attempting to pose a threat of harm. This is because the morally weighted costs of not attempting to pose a threat of harm are the same as the morally weighted costs of not successfully attempting to pose a threat of harm. These theoretical arguments are another way of showing the Causal Connection Requirement is false.

However, in this chapter I am concerned with our intuitions in specific cases, rather than with more general moral theories. We can start with a simple case. Consider,

Faulty Detonator: Two rival villains, Apparent and Imminent, are each culpably attempting to kill Victim. Unknowingly, Apparent is using a faulty detonator. Killing either villain could save Victim. However, from your vantage point, the only way to save Victim from Imminent is by shooting Apparent. Shooting Apparent will set off an explosion that will allow Victim to escape.⁷⁵

To my intuitions, we ought to sacrifice Apparent to save the Victim in this case. This is so despite the fact that only Imminent has a working detonator. Allowing Victim to be wrongfully killed in this

⁷⁵ This is an adaptation of the “Basement Window” case from McMahan, “The Basis of Moral Liability,” 391. See also, McMahan, “Self Defense and Culpability,” *Law and Philosophy* (2005) 757.

case would be very bad. Since it would be permissible to kill Imminent to save the victim, and Apparent and Imminent are equally culpable through attempting to pose identical threats, it ought to be permissible to sacrifice Apparent instead.

One way to show this is to suppose that you do not know which of the villains poses the real threat. Perhaps you know that one of them is using a faulty detonator, but you do not know which one. Before saving the victim, should you spend any time trying to determine which villain has the working detonator? Let us suppose that you could find out which poses the real threat without endangering the victim but at some small cost to yourself. Even in this case, it seems to me, there is very little reason for you to bear this cost.

In contrast, the culpability of the people clearly matters a great deal. Suppose you learn that one of villains is actually an innocent construction worker who is tasked with destroying Victim's house. Perhaps this worker has been somewhat negligent, in failing to triple check that the building is empty, but he is not trying to murder anybody. This realization changes things considerably. You might not want to expend much effort determining which person has the working detonator. You ought, however, to find out which person is the highly culpable one. It is far better to kill a highly culpable attempting murderer, who poses a merely apparent threat, than it would be to kill a merely negligent person who poses a real threat.

This intuition is entirely explained by the proportionality relation. If you had a choice between killing a highly culpable person, and killing a merely negligent person, to prevent the same threat, you should clearly choose to kill the culpable person. This is because it is less bad to harm a more culpable person, relative to the same harm prevented. There is a clear relationship, captured by proportionality, between the badness of preventive harms applied to a person and that person's degree of culpability. There is nothing about this relationship, moreover, to suggest that causal responsibility is indispensable. Sometimes, as in *Firing Squad* and *Faulty Detonator*, causal

responsibility depends on pure coincidence and does not track anything very morally significant about the people involved. In such cases, we have strong reasons to suppose that causal responsibility is unimportant. In contrast, culpability reliably makes a crucially important difference to the apparent badness of preventive harms.

There is an intuitive principle at work behind all these cases swapping cases. We can generalize our conclusions from these swapping cases into a principle about proportionality:

Swapping Principle: if it is narrowly proportionate to inflict on some person a preventive harm of magnitude x , to prevent some harm of magnitude y , on the basis of A 's degree of culpability z , then it is also narrowly proportionate to inflict x to prevent y on any person with culpability of z .

It is this sort of Swapping Principle that seems to be driving our intuitions in *Double Hit Man*, *Firing Squad*, *Faulty Detonator* and many other cases besides. Combined with our earlier conclusion that culpability is the basis of liability, the Swapping Principle gives a clear and intuitive account of all these cases.

So far I have only been discussing causal responsibility, since this is one of the most widely accepted constraints on liability to defensive harming. To keep the focus on causal responsibility I have kept other factors constant. In all my cases so far the villains have each been using the same kind of weapon to target the same victims. Yet, for the same reasons that causation does not matter, the type of weapon or identity of target does not matter either. These factors only make a difference inasmuch as they matter for culpability.

Suppose that in *Faulty Detonator* you do not know whether Apparent is attempting to murder the same victim as Imminent, or one of the many neighbors who live on the same street. If the identity of the victim mattered, it would be permissible to kill Apparent if she was targeting the same victim as Imminent, but impermissible to kill Apparent if she was targeting a neighbor. Yet, to me at

least, it does not appear necessary that we find out who Apparent is targeting before we can intervene.

The fact about Apparent that we really care about is his culpability. If we suppose that Apparent would be justified in killing his other possible target (perhaps he might be trying to save the victim by killing Imminent!) then whom Apparent is targeting once again matters. In this case, it would be permissible to kill Apparent to save the victim if Apparent was targeting Victim, but impermissible to kill Apparent to save Victim if Apparent was targeting Imminent. It is culpability, in this case, not the identity of the victim per se, that matters for liability.

In same way, it should not be controversial, though I think it is worth pointing out, that the permissibility of harming a villain to save a victim does not depend on similarity between the actual threat to the victim and the type of threat the villain is attempting to create. If we suppose that Apparent is trying to shoot the victim, while Imminent is using explosives, this surely does not make a difference to Apparent's liability. The type of threat only matters when it makes a difference to culpability. If Apparent is trying to break the victim's arm, while Imminent is using explosives to blow him up, this will make a difference for culpability and hence for liability.

I concede that the Swapping Principle looms much larger in my intuitions when the actions of the people being compared are nearly identical. In these cases, such as *Faulty Detonator*, the Swapping Principle appears incontrovertible. However, the greater the difference between the actions of the culpable people, the more unclear things become. This could suggest that a similarity in target or type of threat really does matter for liability. Yet when I test each of these factors in turn, I find that they make little difference.

The only time the identity of the target or the type of threat makes a clear difference, is when it matters for culpability. I suspect, therefore, that the reason sheer similarity matters in my intuitions, is that it makes the basic fairness of the Swapping Principle more obvious. The greater the

differences between the two people being compared, the more I must rely on the stipulation that they are equally culpable. This phenomenon is not unique to the Swapping Principle. Moral reasoning relies heavily on analogy, and analogies are clearer the more even irrelevant details are kept the same.

Adopting the Swapping Principle benefits both culpable aggressors and victims overall. Of course, in some cases, such as *Faulty Detonator*, it means that more culpable aggressors can be permissibly harmed. This does not mean, however, that swapping will be worse overall for culpable aggressors. Certainly, removing the Causal Connection Requirement will make more people potentially liable. But precisely this increase in the number of potentially liable people will also mean that preventive harms can be more evenly divided.

Suppose that in *Faulty Detonator* there was another option which allowed you to save the victim by splitting the defensive harm between Imminent and Apparent. This third option, it seems to me, would then be the option you should choose. It is unclear, therefore, whether the overall impact of removing the Causal Connection Requirement would be worse for culpable aggressors. Certainly, it would be better for victims, who could be saved in cases in which they would otherwise be killed.

1.2 Causal Uncertainty Cases

In considering *Faulty Detonator*, I argued that if we do not know which villain has the working detonator, we do not need to find out before we proceed with saving the victim. From this, I concluded that having the working detonator is not a necessary condition of liability to be harmed in defense of the victim. Perhaps this was too hasty. It might be objected that in cases of uncertainty we can proceed without knowing which aggressor is causally responsible, even though causal responsibility remains essential to liability.

This objection is suggested by cases of causal indeterminacy in tort law. These are cases in which it can be proved that two or more people culpably risked causing some harm to a victim, and it can be proved that at least one of these people culpably caused harm to the victim, but which of the culpable people actually caused the harm to the victim is not known. In many jurisdictions, compensation can be awarded in causal indeterminacy cases even though causal responsibility cannot be proved. What changes in these cases is the burden of proof. The victim does not need to prove which of the defendants is causally responsible in order to receive compensation from some or all of them. For the sake of this objection I will assume a close analogy between cases of preventive harming and cases of compensation. I certainly believe that there is some sort of connection between these cases but it may not be as close as I will here assume.

I argue that it is not possible to accept a reversal in the burden of proof in causal indeterminacy cases without rejecting the Causal Connection Requirement. Even those who only share my intuitions about harming apparent threats in cases of causal indeterminacy should agree that the Causal Connection Requirement is false. We can see this by considering the possible justifications for reversing the burden of proof, and by considering what reversing the burden of proof amounts to.

1.2.1 The Prevented Claim Theory

Suppose that two people A and B both culpably risked causing a harm to a victim, but that we do not know which of these two people actually caused the harm that resulted. According to the Prevented Claim Theory each of A and B can be liable to compensate the victim.⁷⁶ This is because each person either culpably caused the harm to the victim, or culpably caused it to be the case that the victim cannot prove which of them caused the harm. Suppose that A caused the real harm. B

⁷⁶ I take my account of this theory from Sandy Steel, *Proof of Causation in Tort Law* (Cambridge UP, 2011) Chapter 4.

could still be liable, according to this justification, because if B had not culpably risked causing the same harm, victim would easily be able to prove that A ought to pay compensation.

In *Faulty Detonator*, this argument might mean that Apparent can be causally responsible for the threat of harm to Victim, even if Apparent does not have a working detonator. Perhaps, if Apparent had not culpably attempted to kill Victim, you would not be unsure whether Apparent or Victim poses the real threat. If the Causal Connection Requirement is correct, then by introducing this uncertainty Apparent has prevented you from saving Victim. In this way, it might be argued, Apparent is a cause of the threat to Victim and can thus be permissibly sacrificed without violating the Causal Connection Requirement.

The problem is that the link between causing the uncertainty and causing the threat to Victim depends on what it is permissible for the intervening person to do. If it is permissible for the intervening person to kill Apparent in cases of causal indeterminacy, then Apparent has not actually contributed to the threat to the Victim by bringing about a case of causal indeterminacy. In this way, at least as an attempt to preserve the Causal Connection Requirement, the Prevented Claim Theory undermines itself.⁷⁷

It is not possible to consistently hold both a) that it is permissible to harm people like Apparent in cases of causal indeterminacy, and b) that people like Apparent causally contribute to the threat to the victim by creating a situation of causal indeterminacy. If (a) is true, then creating causal indeterminacy does not harm the victim so (b) is false. If (b) is true, then it must be because creating causal indeterminacy means that the victim cannot be saved. This will only be the case if (a) is false.

Those who believe that the Prevented Claim Theory justifies reversing the burden of proof should accept that the Causal Connection Requirement is false. This is because if it is permissible to

⁷⁷ Steel anticipates this and acknowledges that it is a problem, *Proof of Causation in Tort Law*, 178-9.

reverse the burden of proof in cases of causal indeterminacy, then it must sometimes be permissible to impose defensive harms on (or demand compensation from) culpable apparent threateners. If it is permissible to impose defensive harms on apparent threateners, then either the Causal Connection Requirement is false, or creating a situation of causal indeterminacy, by posing an apparent threat, counts as causing a threat to the victim. Creating a situation of indeterminacy only counts as causing a threat to the victim if it is not permissible to harm apparent threats. Thus, if we accept that it is permissible to reverse the burden of proof in cases of causal uncertainty, then we should accept that the Causal Connection Requirement is false.

1.2.2 The Reliance on Wrongdoing Principle

A second related justification for reversing the burden of proof in cases of causal indeterminacy is the Reliance on Wrongdoing Principle.⁷⁸ According to this principle a person cannot evade tort liability by relying on the wrongdoing of another. Steel argues that this situation is somewhat analogous to that of causal indeterminacy. It is analogous because both situations are covered by the principle that “a defendant is not permitted to rely upon another's wrongful conduct in order to avoid a liability to pay damages, which would otherwise arise against the defendant.”⁷⁹

If it were not for the wrongdoing of the mere attempter, the real threatener would be clearly liable. This shows why the real threatener, Immanent in our example, should be liable. However, the merely apparent threatener, Apparent in our example, does not need to rely on anyone else's wrongdoing in order to evade liability. In fact, if the merely apparent threatener is liable to be harmed in *Faulty Detonator* it is only because of the real threat of harm caused by the wrongdoing of

⁷⁸ Again, I take my account of this principle from Steel, *Proof of Causation in Tort Law*, Chapter 4.

⁷⁹ Steel, *Proof of Causation in Tort Law*, 191.

someone else, that must be prevented. Thus, the Reliance on Wrongdoing Principle is applicable to the case of Immanent but not to that of Apparent.

In a case of causal indeterminacy, we do not know whether the person we can harm to save the victim is more like Apparent or Immanent. If the Causal Connection Requirement is correct then Apparent is not liable to be harmed. The Reliance on Wrongdoing Principle is not applicable to the case of Apparent. It certainly does not provide us with any reasons for supposing that Apparent is causally responsible for a threat to Victim. Thus, if it is permissible to sometimes harm Apparent in these cases it must be the case that the Causal Connection Requirement is false.

1.2.3 The Burden of Proof and Liability

It might be objected that altering the burden of proof is not the same as making a judgment about liability. Perhaps we can permissibly harm people such as Apparent, without establishing whether they are causally responsible, despite the fact that causal responsibility remains a necessary condition of liability. In this way, we could accept a change in the burden of proof, without rejecting the causal connection requirement.

However, reversing the burden of proof for culpable people in cases of causal indeterminacy is tantamount to claiming that these people have a liability. I am assuming that the burden of proof must be justified by comparing the badness of allowing a merely apparent threatener to be harmed to the badness of allowing the victim to be harmed, or go uncompensated. I am also assuming that it is no worse for a victim to be harmed or go uncompensated in a case of causal indeterminacy than it would be in any other case. Thus, reversing the burden of proof for culpable people in cases of causal indeterminacy is only justified if it is less bad for such people to suffer defensive harms than it would be to impose these harms on entirely innocent people.

This can be illustrated with an example. Suppose that three companies have each recklessly manufactured equally defective versions of the same product, and that a victim was injured using one of these defective products. It is clear that one of the three companies is causally responsible for the harm to victim, but there is no way to prove which of the three companies this is. A court might decide that the victim need not prove which of the companies manufactured the product that injured her. Instead, the court might hold that each company can only avoid liability by proving that they did not manufacture the product that caused the injury.

We must consider how this reversal of the burden of proof could be justified. In particular, we should consider what it is that distinguishes these companies from any other set of people or companies that might pay compensation to the victim. We can assume that in ordinary cases, and for other people and companies, the standard burden of proof is justified. We can also assume that it is no worse for a deserving victim to go uncompensated because she cannot meet the standard burden of proof due to causal uncertainty than it would be for a deserving victim to go uncompensated because she cannot meet the standard burden of proof for some other reason. Thus, there must be something special about the three companies involved in this case that reduces their claim to be protected by the ordinary burden of proof. In other words, there must be something special about these companies that makes it less bad for two of them to pay compensation even if they are not causally responsible for the injury.

I have defined a liability as a reduction in the badness of defensive harms applied to the liable person. Thus, if we reverse the burden of proof in cases of causal indeterminacy we must accept that the culpable people in these cases have a liability. There is something about these people that makes it easier to justify imposing defensive harms on them, compared to bystanders and entirely innocent people. Admittedly, reversing the burden of proof does not require us to suppose

that apparent threateners are as liable as real threateners. However, it does require us to suppose that apparent threateners are somewhat liable to defensive harms.

The obvious candidate for the basis of this liability is culpability. After all, there is an important difference between cases causal indeterminacy and cases of culpability indeterminacy. Suppose that two people each causally contributed to some harm suffered by a victim. One of these people negligently caused harm to the victim in a way that would normally require compensation, while the other person acted innocently and causally contributed to the harm in a way that does not require compensation. Yet, we do not know which person is negligent and which is culpable.

I take it that there are no grounds for reversing the burden of proof in such a case. Making these people prove their innocence in this case would not so much reverse the burden of proof as abandon it all together. Reversing the burden of proof in cases of causal indeterminacy is only plausible because the people in these cases are already known to be culpable. Those who accept a reversal of the burden of proof in these cases should acknowledge that the Causal Connection Requirement is false.

1.3 Necessity Cases

There are some cases in which rejecting the Causal Connection Requirement the would benefit victims without making any culpable aggressors substantially worse off. Consider,

Robin Hood: Robin Hood and Little John are rival villainous outlaws. Robin Hood shoots one arrow at a farmer and Little John shoots two arrows at the same farmer. Being hit by arrows is painful but not fatal. Being hit by two arrows is twice as painful as being hit with one arrow. You can either a) deflect the arrow shot by Robin Hood back towards Robin Hood, or b) deflect both arrows shot by Little John, but in such a way that one of them will hit Robin Hood.

In this case, it seems to me, we should clearly choose option (b). Choosing (a) would make the Victim worse off and make nobody better off. Yet according to the Causal Connection Requirement

option (a) is the one we must select. This gives us a further reason to reject the Causal Connection Requirement.

The strangest thing about the Causal Connection Requirement in this case is that it prevents you from saving the farmer from both arrows on grounds that doing so would be unfair to Robin Hood. This is because according to the Causal Connection Requirement, and the moral responsibility view of liability, Robin Hood can only be liable to prevent those threats that he caused. The effect of the Causal Connection Requirement is especially strange in this case, because it means that inflicting a certain magnitude of harm on Robin Hood is permissible if it does less good, but inflicting that same magnitude of harm would be impermissible if it did more good. In this way, the Causal Connection Requirement can require us to violate what I understand to be constraint of necessity: the requirement that we only harm people in a way that brings about the best outcome overall. Option (b) appears worse, and under the circumstances impermissible, because it sacrifices Robin Hood to block one arrow, when it would be possible to sacrifice Robin Hood to block two arrows.

We can accept that Robin Hood might be made better off if he were prevented from harming the farmer, but this does not give us a strong enough reason to justify allowing a greater harm to befall the victim. Any plausible principle for the distribution of defensive harms must balance the interests of victims against the interests of culpable aggressors. As we have seen, a recurrent problem with the Causal Connection Requirement is that it appears to give far too much weight to the interests of culpable aggressors.

This same point can also be made using cases of compensation. Consider,

Robin Hood (Compensation Variation): Robin Hood and Little John are rival outlaws who have been firing arrows at and stealing from the local farmers. You, the local Sheriff, have been charged with extracting compensation for the victims. One day you manage to confiscate Robin Hood's horse and plan to sell it to raise money to compensate the victims.

Suppose in this case that the neediest victims are those of Little John. Robin Hood tends to steal from wealthier people and tends to harm them less. We can also assume that Robin Hood inherited the horse, and did not steal it from anyone. It seems, then, that you ought to compensate the neediest victims first, or at least prioritize the distribution of compensation based on the degree of injury. Yet if the Causal Connection Requirement is correct then this option is unavailable.

What makes this implication of the Causal Connection Requirement so strange, is that it is permissible to take Robin Hood's horse in the first place. Taking the horse only becomes impermissible, according to the Causal Connection Requirement, if the compensation is then given to Little John's victims. Yet it is hard to see how this subsequent allocation of compensation could wrong Robin Hood, or even treat him unfairly. As long as the treatment of Robin Hood is proportionate and necessary in light of his culpability, he should have no legitimate complaints.

1.3.1 The Conceptual Limits of Compensation

It might be objected that this last argument goes too far. Compensation, it might be argued, is inherently a matter of reversing harms or wrongs. It thus requires two parties causally connected by a past wrong. We can only permissibly reverse the wrongs done by Robin Hood if we take from him and give to those he has wrongfully taken from. Taking from Robin Hood to repair the wrongs done by Little John could not be justified as a form of compensation. It is simply a redistribution of resources from Robin Hood to the victims of Little John.

I suspect that this line of thought lies behind some of the opposition to rejecting the Causal Connection Requirement. Despite the differences between them, it is tempting to suppose that the ethics of preventive harming and the ethics of compensatory harming should be founded on the same principles. If compensation conceptually requires reversing a transaction, it is tempting to build this requirement into the ethics of preventive harming as well. This might explain why some are

tempted to retain a Causal Connection Requirement, despite its unintuitive implications. It might also explain why causal connections remain essential to tort law, despite the exception made in cases of causal indeterminacy.⁸⁰

I accept that it should be possible to explain the ethics of compensatory and preventive harming with a single set of principles. Nearly all acts of compensation are also acts of prevention, since the compensation should help to prevent any ongoing suffering caused by the original harm. I also accept that the concept of compensation implies reversing or undoing harms. However, I believe we should take a broader view of what counts as undoing a harm. When someone suffers a loss we can compensate them for this loss by restoring them to the state they would have been in had they not suffered the loss. It does not matter whether we restore them to precisely the same state. We cannot undo the passage of time and often we cannot return exactly what has been lost. What matters is that we restore the person to a state that is as close as possible *in terms of the things that matter* to the state they would have been in had they not suffered the loss. What matters are things like the welfare of the person and the quality of their life. How this should be done is itself a complex question, but this is the basic picture.

Compensating a person involves remedying a harm that she suffered. To compensate we only need to reverse a part of the transaction that caused this harm in the first place. Specifically, we need to reverse the part – the harm – that requires compensation. However, it is conceptually possible for compensation to come from a person who is not at all causally responsible for the harm that is compensated for. Consider,

Beneficence. A person heroically sacrifices his leg to save a drowning child. A wealthy business owner sees the man's story on the news and willingly pays him a large amount of money so that he can afford to rebuild his life and end up as well off as he was before. The business owner is satisfied that she was able to compensate the man for his loss.

⁸⁰ This point is made by Steel in *Proof of Causation in Tort Law*, Chapter 3.

In this cases, I do not think that the business owner is making a mistake a conceptual mistake in supposing that she has performed an act of compensation. This is the case even though she is in no way responsible for the loss and did not benefit from it.

We should think of compensation in the way that we think of self-defense. Ducking behind a boulder is an act of self-defense if it prevents a harm, even though the ducking does not impose a harm on anyone. Deflecting a harm towards an innocent bystander is an act of self-defense, even though the bystander is in no way responsible for the threat. In the same way, we compensate a person when we give them something to remedy a loss. The concept of compensation does not tell us where that remedy must come from.

As in the case of self-defense, it is a separate question whether taking from someone to provide this remedy is justified. This is the question that we have been considering. We must consider what sorts of objections different people could reasonably make to different redistributions of harms. Some of the best justifications for redistributing harms in cases of self-defense do not depend on causal connections. There is nothing about the concept of compensation that should prevent us from applying these justifications to cases of compensation.

In fact, to the extent that the ethics of compensation work the same way as the ethics of self-defense, the same sorts of considerations about proportionality should apply. Sometimes it may be proportionate to extract compensation from a culpable aggressor. Other times, however, it may be proportionate to extract the resources need for compensation from innocent bystanders, perhaps in the form of taxes or mandatory contributions. As in self-defense cases, liabilities are key to determining how far we can permissibly go.

I will not go any further towards a theory of compensation, and I concede that there may be important disanalogies between preventive and compensatory harming which I have neglect. All I

wish to establish is that the concept of causation does not give us any reasons to retain the Causal Connection Requirement as part of the ethics of preventive harming. It is reasonable to suppose that the same principles can be used to explain self-defense and compensation without including a Causal Connection Requirement.

1.4 Responsibility Without Causation

Many theorists who appear to accept the Causal Connection Requirement do so because causal responsibility is thought to be a necessary condition of moral responsibility. I made this assumption when discussing moral responsibility in Chapter Two. Yet there are some cases in which it is at least unclear whether moral responsibility follows causal responsibility. These cases present a problem for the moral responsibility theorists, who must tell us how moral responsibility works and is restricted in these cases. They also give us further reason to doubt the Causal Connection Requirement. Even those hold that a person must be morally responsible for a threat in order to be liable to be harmed in the prevention of that threat, should perhaps admit that the Causal Connection Requirement is false.

We can see this by considering cases in which bringing about the best outcome requires us to perform more than one defensive action. Intuitively, these cases are not more difficult than standard cases. Yet they can become surprisingly difficult if we attempt to follow the Causal Connection Requirement. Consider,

Missile: Suppose that Villain culpably launches a missile that will kill a thousand victims. You can save the victims but only by pushing the missile's self-destruct button. The only way to reach the button requires you to knock an innocent Bystander off a bridge. Falling off the bridge will kill Bystander unless you then throw Villain (who is more aerodynamic) to cushion his fall. If you do this, then Villain will die and Bystander will be saved.

In this case, let us suppose, it is widely proportionate to kill Bystander in order to save the thousand victims. Clearly, you ought to knock over Bystander and push the self-destruct. It is also clear that you should then throw Villain off the bridge to save Bystander. Doing this will allow you to save the victims in a better way, by hurting Villain rather than Bystander.

The problem is that the Causal Connection Requirement may get in the way of this second action. Suppose you have knocked over Bystander and pushed the self-destruct button. You now just need to throw Villain to save Bystander. Yet Villain may not be liable to this further harm, according to the Causal Connection Requirement, since Villain is not obviously causally or morally responsible for the threat to Bystander.

To explain these cases using moral responsibility we must either accept a broader notion of causal responsibility, or reject the Causal Connection Requirement. It may be true that had Villain not attacked the victims, you would not have knocked Bystander. In some sense, therefore, the Villain's action is a cause of the threat to Bystander. However, your intervening agency, and your decision to knock Bystander, may also be the sort of thing that would otherwise block an attribution of causal responsibility.

In *Faulty Detonator* it might be the case that Apparent innocently gave Imminent the directions that allowed Imminent to reach the victims' house. In some sense, then, Apparent is a cause of the actual threat to the victims, but I do not think that this sort of causation is what proponents of the Causal Connection Requirement have in mind. Returning to *Missile*, we should recognize that it is most clearly the intervening person, not the Villain, who has caused the threat to Bystander.

Instead of accepting a broader notion of causal responsibility, moral responsibility theorists might propose that people such as Villain, can be considered to be morally responsible for any defensive action that is necessary to prevent a threat of unjust harm for which that person is morally

responsible. He may be morally responsible for such an action, even if it is questionable whether he is causally responsible for it. This is a promising approach for the moral responsibility theorist to take. However, it requires accepting that the Causal Connection Requirement is false.

While promising, this approach still presents some problems for the moral responsibility view. In some cases, it will be impossible to show that a villain is responsible for a necessary defensive action, and therefore liable, without first assuming that the villain is liable. This is because whether a defensive action is necessary can depend on the Villain's liabilities. Consider,

Missile (Nearstander Variation): The same as *Missile* except that there are two different routes to the button, each blocked by a different innocent bystander. You have a choice between knocking over Bystander and knocking over Nearstander. If you knock Nearstander he will be paralyzed but not killed. However, you will not have time to throw Villain to break his fall.

In this case it is still intuitively clear that you ought to knock over Bystander and then throw Villain to save him. It is much better for Villain to be killed than for Bystander to be killed or for Nearstander to be paralyzed. However, even if we revise our view of moral responsibility, it is not clear how the moral responsibility view can explain why Villain can be sacrificed to save Nearstander in this case.

If it is not permissible to push Villain to save Bystander, then you ought to knock over Nearstander on the way to save the victims instead. This is because paralyzing Nearstander is better than killing Bystander. For this reason, the defender of the moral responsibility view once again needs to assume that pushing Villain to save Bystander is permissible in order to establish that Villain is morally responsible for the threat to Bystander.

This is because whether Villain is morally responsible for the threat to Bystander depends on whether pushing Villain to save Bystander is permissible. Yet, according to the moral responsibility view, we need to know whether Villain is morally responsible for the threat to Bystander in order to know whether pushing Villain to save Bystander is permissible. Once again, we are faced with a

muddle. The moral responsibility view can only explain why it is permissible to push Villain if they first assume that it is permissible to push Villain.

It may be possible to further to revise our account of moral responsibility, or causal responsibility, to account for these cases. It is significant, however, that the moral responsibility view and the Causal Connection Requirement are susceptible to these sorts of muddles in the first place. After all, *Missile* is not really a hard case. In none of the variations is our intuition that we should push Villain to save Bystander ever in doubt.

We can see this by recognizing that it would be clearly permissible for you to perform one action that would harmlessly push Bystander and sacrifice Villain to save the thousand victims. Making this a case where multiple actions are required creates inordinate trouble for the moral responsibility view and the Causal Connection Requirement without for a moment ruffling our intuitions about what we ought to do. Like the other cases I have considered, this suggests that our intuitions depend on culpability and not causation.

2. Liability, Responsibility, and the Future

In the previous section I offered a series of arguments to show that the Causal Connection Requirement is false. In this section, I continue this argument by focusing on future threats. Self-defense, and defensive harming more generally, is inherently forward looking. It concerns what it is permissible to do to prevent some event from occurring. This complicates things for views that make causal or moral responsibility for some threat a necessary condition of liability to be killed in prevention of that threat. To see this, consider,

Preemption. Villain culpably intends to murder Victim by pushing him off a cliff. Victim is standing on the edge of the cliff. You can save Victim but only by killing Villain as he walks over to push Victim.

In this case it should be permissible to kill Villain to save Victim. Yet, it is difficult to see how Villain could have already caused a threat to Victim. So far Victim has simply begun walking towards Victim. Walking towards Victim does not in itself pose any threat to Victim, indeed it is something Villain is permitted to do. The only way in which walking towards Victim poses a threat to Victim is because of Villain's intention. Thus, if Villain is liable in this case it is surely because of what he culpably intends to do, not because of anything he has caused.

It might be suggested that the permissibility of killing Villain to save Victim does not need to depend on what Villain has already caused, but could also depend on what Villain will cause in the future. This way of thinking about Villain's culpability is fairly common, but it has an obvious problem. If Villain's liability to be killed in defense of Victim depends on the fact that Villain will cause some threat to Victim in the future, then successful attempts to prevent Villain from creating this threat will be morally self-defeating. If you know that your defensive action will succeed, then you also know that Villain's attempt to pose a threat to Victim will not be causally effective. If you know this, and causal responsibility is a necessary condition of liability, then you have lost your justification for harming Villain.

Perhaps Villain's liability could depend on the fact that unless you do something Villain will wrongfully push Victim. This can only provide a partial solution. Suppose there is another person nearby who is ready to save Victim if you do nothing. However, this person can only save Victim in a way that would also impose a substantial cost on an innocent bystander. Clearly, it would be preferable if you could simply kill Villain to save Victim rather than leaving someone else to kill Villain and harm a bystander to save Victim. Yet if the permissibility of harming Villain depends on what he will cause unless you do something, then the presence of the other potential intervener makes your intervention impermissible.

In reply to this example, it might be suggested that Villain's liability could depend on the culpable action he will perform 'unless we together do something' or 'unless somebody intervenes.'⁸¹ This would help in the case I just described, but it will not resolve the wider problem. Suppose that between Villain and Victim there is a landmine, which you know Villain will step on when he gets closer to Victim. Stepping on the landmine will kill Villain, thus saving Victim, but it will also kill an innocent bystander. Once again, it seems like you ought to be able to kill Villain before he reaches the landmine, if this is the only way you have to intervene.

Yet, as in the previous case, there is no possibility that Villain will cause a threat to Victim. Stepping on the landmine could not itself be culpable since Villain could not be expected to realize that it is there. Moreover, the action of walking over to Victim, which will cause the landmine to explode, is something Villain is permitted to do. It seems particularly difficult for causal views of liability to account of these important cases of preemptive liability.

It might be suggested that Villain could be killed to prevent him from stepping on the landmine, because he would otherwise be responsible for the deaths of the bystanders. This suggestion relies on the thought that people can be responsible even for unforeseen consequences of their wrongful actions. However, we can reasonably doubt whether this premise should be accepted. Consider,

Two Gunshots: Two men, Hunter and Murderer, each fire a rifle in the woods during the hunting season. Hunter fires at a deer, while Murderer fires at a person he wants to kill. In each case the sound of the gunshot causes a nearby pedestrian to unforeseeably suffer a heart attack and die.

We can assume that Hunter's action was permissible and Murderer's was not. Yet, at least to my intuitions, Murderer is no more responsible for the ensuing heart attack than Hunter is.

⁸¹ For a discussion of these concepts see Derek Parfit, *Reasons and Persons* (OUP, 1984) 71.

The noise of the gunshot, and the fact that it caused a heart attack, is not a part of what makes Murderer's action wrongful. It would have been perfectly permissible for Murderer to fire his gun at exactly that moment in a different direction. The same is true in the case of *Preemption*. The villain is perfectly permitted to walk over to the cliff edge to admire the view, and doing so would set off the land mine in the same way. Moreover, even if this premise is accepted, it is not clear that walking towards the Victim is a wrongful action. The thing that might make it wrongful, pushing the Victim, has not yet occurred and will never occur.

2.2 Responsibility for Harms that will Never Occur

This raises a related problem that responsibility based views face in these cases. This is to explain why non-actual differences between people should make a crucial difference for liability. Consider, for example, the moral responsibility account of liability. On this account a person is liable to be harmed in the prevention of some threat if and only if that person is morally responsible for that threat. It is not entirely clear what responsibility for a threat involves. Yet, to be plausible this view must allow some people to be harmed before they have caused a substantive independent threat. We should not have to wait until a trigger has been pulled or a bullet is on its way. Now consider,

Preemptive Responsibility: Two gun carrying villains, Lucky and Unlucky, each have the same culpable intention to shoot Victim. When the villains have pointed their guns and are about to fire, you can prevent them both from firing, and save Victim, if you shoot either of the villains. In addition, you somehow know that if you do not intervene a dove will fly into the path of Unlucky's bullet and prevent him from killing Victim.

On the moral responsibility view of liability, it seems that only Lucky should be liable to be killed in defense of Victim. This is because only Lucky does, or might, pose a real threat of harm to Victim.

This is odd, however, because at the present, when the defensive action is necessary, Lucky and Unlucky are identical. The only difference between them is the fact that a dove will happen to fly in front of Unlucky's bullet. Yet if you save Victim, by shooting one of the villains, this will never occur. Thus, on the assumption that you are going to save Victim, neither of these villains is morally responsible for a risk of actual harm to Victim, nor will either of them get a chance to be causally responsible for a substantive threat of real harm to Victim. Why, then, should only Lucky be liable to defensive harming?

To this it might be pointed out that before you intervene only Lucky poses a real threat of harm, because Unlucky's bullet will be stopped by the dove. Yet if you do plan to intervene then you already know that Unlucky's bullet will not be stopped by the dove. It will be stopped by you. Thus, as soon as you plan to intervene, you know that neither Lucky nor Unlucky poses a causal threat to the victim, and that they will each be equally unsuccessful in their attempt to cause harm to Victim for the same reason. The only difference between Lucky and Unlucky is an event that has not occurred, and will not occur if you intervene.

To find a moral difference between Lucky and Unlucky it seems that we must consider other possible worlds. There is some nearby possible world in which Lucky will kill Victim and Unlucky will not. It seems to be this world which, on a moral responsibility view of liability, is highly significant. Yet, there is presumably also a nearby world in which the dove flies a slightly different course, and Unlucky kills Victim and Lucky does not. There are innumerable possible worlds in which things might go differently and Lucky, or Unlucky, might cause or not cause all sorts of harm. Yet in the actual world, once you have decided to intervene, you know that neither of these people will cause any harm. Once you decide to intervene, the risk that they each pose to Victim is the same.

I do not mean to claim that the moral responsibility view is impossible, or contradictory. It does seem, however, that the moral responsibility theorist ought to explain why morally significant differences in one possible world should be crucially important for deciding liability. Surely if one person is liable and another is not it should be because of facts that are true about these people in the actual world. Lucky and Unlucky only pose a threat to Victim, because of what they intend to do in the moments before you interfere. It seems to me that these actual culpable intentions, rather than some non-actual difference in causal responsibility, should justify their liability.

3. The Scope Objection

Despite all these arguments, some may still be tempted to retain the Causal Connection Requirement. It is sometimes suggested that despite its inherent implausibility, this requirement should be preserved because it is the best available means of restricting liability. Discussions of the Causal Connection Requirement often end with the conclusion that it should be maintained because there is no clear alternative to take its place. Without the Causal Connection Requirement, or some replacement, it is feared that there will be no way of limiting the 'scope' of liability. There are, I think, two worries here, both of which are important.

The first and simplest is a practical objection to abandoning the Causal Connection Requirement in favour of a culpability centered approach to liability. Such a theory, if widely adopted, might lead to dangerous chaos by seemingly authorizing vigilante actions against apparently culpable people. It is important to distinguish between the theoretical claim that a given culpable aggressor is liable to a defensive harm, and the practical claim that a given person is permitted to harm that culpable aggressor on the basis of the aggressor's liability.

Judgments of culpability are very difficult to make without the easy omniscience of thought experiments. These limitations will provide a major restraint against acting on the basis of liability

justifications in realistic cases.⁸² Put another way, there is reason to suppose that evidence relative liability will be far more restricted than fact relative liability. We should make laws and legal procedures which reflect this. The existence of such laws and legal procedures should, in turn, provide independent fact relative restrictions on the permissibility of defensive harming.

Then there is the more substantive theoretical objection. Removing the Causal Connection Requirement opens up many more culpable people to preventive harming. As I argued above, this does not necessarily mean that the overall quantity of preventive harming will increase. There is a trade off between the number of potentially liable people and the amount of harm that must be inflicted on each. But this does mean that more people will be available to be preventively harmed in more cases. People unrelated to a threat can be preventively harmed simply because they are culpable.

For many, even for those who question the Causal Connection Requirement, this is an unpalatable implication. Some of these worries can be assuaged. Of course, part of the point of rejecting the Causal Connection Requirement is to make more people potentially liable. But this is not as bad as it might seem. Culpability can play a significant role in limiting the scope of liability. We must be sensitive to the degree of a person's culpability, to the nature of culpability, and to how culpability functions over time.

The scope objection is often explained using negligent or otherwise minorly culpable people. McMahan writes,

It would be absurd to suppose that my right of self-defense could be triggered by [a] driver's culpably lying to his wife over his cell phone, so that the only issue would be one of proportionality. Culpability for lying to his wife cannot make the driver liable to defensive action by me.⁸³

⁸² Christie makes this point, *Killing One to Save Another*, 157.

⁸³ Quoted in Christie, *Killing One to Save Another*, 132.

There are two points here that should be considered. The first is whether the lying husband's culpability is serious enough to justify harming him. This might depend on what he is lying about, but it is intuitive to suppose that it is not. In this way the constraint of proportionality can play an important role in limiting the scope of liability.

This constraint can also explain why it seems strange to suppose that culpability for one type of wrong, like lying, can make one liable to an unrelated harm such as being killed. Harms of different kinds are usually of different magnitudes. If we imagine that instead of lying to his wife, he was using the cellphone to order someone to be killed, our intuitions begin to change. We might limit the scope of liability by supposing that the threat a person is harmed prevent must be of the same kind as a threat they are culpable for attempting to create, or that a person can only be liable to harms that are of the same kind as harms that they are culpable for attempting. Yet these constraints, once separated from the question of proportionality, become much less intuitive. They also violate the Swapping Principle, and are implausible in variations of many of the cases discussed above. Still, despite the implausibility of the different options, there remains a strong impulse to use some such constraint to limit the scope of liability.

Other writers express a similar concern. Thomson worries about a negligent driver who nearly runs over a pedestrian.⁸⁴ Can that driver be permissibly killed later that day to save the pedestrian, or even an unrelated victim? If a person negligently risks harming someone, but nothing happens, surely that should be the end of it. Removing the Causal Connection Requirement risks creating a whole underclass of onetime culpable people who may permissibly be harmed whenever doing so would benefit somebody else.

There is an undeniable force to these objections. It is important that a theory of preventive harming be sensitive to the rights and suffering of culpable people. I share this sensitivity. It is the

⁸⁴ Judith Thomson, "Remarks on Causation and Liability," *Philosophy and Public Affairs* 13, 2 (Spring 1984) 104.

reason I have argued against the claim that people can deserve to be harmed, and against the claim that harms applied to culpable aggressors do not aggregate in judgments of proportionality. However, I question whether a Causal Connection Requirement is the best way to protect culpable people from overly permissive preventive harming. It does limit the number of people who are potentially liable to be harmed in the prevention of a given threat, but it often does so in an implausible and morally arbitrary way. As cases like *Faulty Detonator* show, it seems better for a preventive harm to befall a highly culpable attempted murder, rather than a merely negligent person who poses a real threat. What is needed, therefore, is a better and more sensitive way to limit liability. I propose that we can find this by looking more closely at culpability. In the remains of this chapter I want argue for the basic plausibility of using culpability to limit liability.

4. Liability, Culpability, and the Past

The most important way in which culpability restricts the scope of liability is the way in which culpability changes over time. The role of time is, I think, much more important than the role of space. Of course, in almost all practical cases space will be a constraint. We cannot normally impose preventive harms on someone who is a long way away. If we could, however, I do not see how sheer distance could make a difference. There is something special about time that makes it play a bigger role in our intuitions about liabilities. Many of the worries raised by the ‘scope’ objection reflect a concern about the way culpability functions over time.

Consider the objection raised by Thomson that removing the Causal Connection Requirement would mean that a person who behaved negligently earlier in the day could later be sacrificed to save an innocent person from an entirely unrelated threat. This certainly seems implausible, and it is this sort of case that tempts many to retain the Causal Connection

Requirement. We are left wondering: what about the next day, or in a year? Without the Causal Connection Requirement how will culpability be limited over time?

This is an important concern. Certainly, if a person was once negligent, or even highly culpable, this should not give them an indefinite liability. To resolve this concern, however, we do not need to reach for a Causal Connection Requirement. Instead, I propose looking more closely at culpability itself.

There is an important intuitive difference between my culpability for what I am currently doing or plotting, and my culpability for culpable actions or attempts I have performed in the past. This is a common sense distinction. It is the same as the difference between reading a book and having read a book. There is a difference between saying that someone is reading *Lord Jim*, and saying that the person has read *Lord Jim*. This is true even though the only reason the person has read *Lord Jim* is that it was once true that he was reading *Lord Jim*. The case of culpability is similar, though slightly more complicated.

To make sense of liability over time we need to distinguish three ways of speaking about culpability over time. First, we have the culpability of what I am currently doing or plotting, which we can call primary culpability. We often distinguish this sort of culpability by using the adverb ‘culpably.’ If someone is doing something ‘culpably’ then they have primary culpability. This is different from what we can call residual culpability. This describes my present responsibility for the primary culpability I had in the past. If I am walking home after murdering someone it would be right to say that I am culpable, but it would not be right to say that I am walking home culpably. Residual culpability still carries some moral weight. It implies that I am still answerable for my past culpability and can be treated accordingly.

To make the distinction clearer, we need to know a little more about the structure of culpability. My own view, which I will elaborate in the coming chapters, is that a judgment of

primary culpability is an assessment of the moral quality of a person's present intentions as a response to the person's circumstances. A person becomes culpable when he currently holds a subjectively wrongful intention. This is an intention that manifests insufficient concern for morally significant considerations, given his current beliefs and available evidence.

This judgment of primary culpability depends on facts about the person's mental states and facts about the person's circumstances. As these things change it may stop being the case that the person's current intentions manifest an insufficient concern for morally significant considerations given the person's current beliefs and available evidence. This does not mean that the person is no longer culpable. But it does mean that he is no longer primarily culpable.

We often treat people as culpable not because of what they are currently doing or plotting but because of what they have done. People change as time goes on. There may be little reason, other than identity, for supposing that it is reasonable to attribute an intention a person had many years ago to that same person today. One task for a theory of culpability is to identify the factors that matter for making attributions of subjectively wrongful intentions. As Parfit argues, identity may not be one of those factors.⁸⁵ A judgment of residual culpability is an assessment of how reasonable it is to attribute a past intention to a present person.

The reasonableness of attributing a past wrongful intention can depend on a variety of factors, such as the person's current attitude towards that past intention, and the changes the person has undergone during the intervening time. It is an assessment of how reasonable it is to use the past intention as a basis for our current treatment of the person.

This distinguishes it from the third way of talking about culpability. This is what we could call historical culpability. Historical culpability is simply the assertion that someone once had primary or residual culpability in the past. When it comes to historical culpability we can reasonably say that a

⁸⁵ Derek Parfit, *Reasons and Persons* (OUP, 1986), Part 3.

person is no longer culpable, even though he was culpable earlier. This is not true if the person still has residual culpability. If the person still has residual culpability, then we say that they are culpable even though they are not behaving culpably. These three categories are a part of the way we think and speak about culpability independent of any claims about liability. However, these distinctions are very helpful in understanding the way culpability justifies liability over time.

4.1 Why These Distinctions Matter for Liability

In most cases, the kind of culpability we care about in self-defense cases is primary culpability. This is the kind of culpability at issue in cases like *Faulty Detonator*. When Apparent and Imminent are currently attempting to kill an innocent victim they both have primary culpability. This kind of culpability generates the most powerful liabilities. It makes it permissible to kill multiple culpable villains in order to save the life of a single innocent person. It also makes it permissible to kill culpable villains, such as Apparent, even when they luckily are not causally responsible for an actual threat. Primary culpability depends on the current culpability of the person. It is a description of the moral quality of what they are currently doing or plotting. Consequently, this kind of culpability wears off very quickly once the attack is over.

In replying to Thomson's case, therefore, we could agree that there is an important moral difference between a person who is currently negligently risking harm to others, and a person who was once negligent in the past. This difference need not depend on causation, but rather is internal to culpability itself. The former person manifests a degree of primary culpability, while the latter has only some residual culpability.

I think it is intuitive that primary culpability is simply much more serious than residual culpability. One reason for this is that a person who has primary culpability will always also have residual culpability. If it is presently true that I have a subjectively wrongful intention, then it is also

true that a subjectively wrongful intention can reasonably be attributed to me at this time. We can think of primary culpability as residual culpability at its maximum strength. If it is ever reasonable to base our treatment of a person at a time on some subjectively wrongful intention, it is surely reasonable to do so when the person still has that intention.

As time passes the moral reasons we have for caring about an instance of past culpability seem to fade away. The moral significance of the past in the present is an important question in its own right. Yet it often receives little attention even in those fields, such as the ethics of compensation, where it is clearly central. Circumstances and other people move on, as does the life of the formerly culpable person. All of these factors might help explain why past culpability is generally less morally significant than present culpability.

Past culpability may also be less morally significant, for the person in question. When a person has primary culpability, there are things about that person that make him culpable. When a person has residual culpability, there are things that make it reasonable to attribute past culpable intentions to the present person. As time passes, the things that made the person culpable may no longer hold. The person may even lose those beliefs and tendencies that led him to become culpable. As these changes put more distance between the present person, and their past culpable intention, it becomes less reasonable to attribute those past intentions to the present person. This is another reason why residual culpability is less serious than primary culpability, and may itself decrease over time.

We can find these differences in our intuitions about self-defense and punishment. We feel that it is permissible to save an innocent victim by killing a person who is culpably attempting to murder someone. Yet it is much less intuitive to suppose that a convicted murderer can be permissibly killed to save an innocent life. The convicted murderer is still culpable, and liable because of that culpability. This liability can help explain why it is permissible to imprison him as a

means to deter crime. Yet the murderer no longer has primary culpability. His liability is therefore weaker, and may not be strong enough to make killing him permissible.

To unify our theory, we should see how this distinction fits into our theoretical justification for liabilities. According to the significance of choice argument, people who culpably attempt to harm others cannot reasonably object to being treated as liable because the morally weighted costs of avoiding becoming culpable are so low. The morally weighted costs of avoiding becoming culpable are low because avoiding culpability is something these people ought to do. People cannot reasonably complain when costs are attached to acts that these people ought to avoid.

This unreasonableness is a matter of degrees. The stronger the moral reasons a person has to do something, the less that person can reasonably object to costs being attached to doing that thing. It is this difference that explains proportionality. It is more unreasonable to claim that I should be able to murder without liability than it is to claim that I should be able to be negligent without liability. Thus, we can justify imposing greater costs on murderers than we can impose on negligent people.

The distinction between primary and residual culpability matters when assessing how unreasonable certain objections and appeals to costs can be. A primarily culpable person ought to immediately stop whatever it is he is culpably doing. This is why primarily culpable people can make very little reasonable objection to a moral rule that attaches costs to continuing this culpable act. After all, a primarily culpable person can always stop being primarily culpable simply by complying with the demands of morality.

By contrast, a person who has done something wrong, but is not currently doing anything wrong, is in a much stronger position to object to the imposition of costs. Such a person can point out that there is nothing they can now do to undo the fact that they were once culpable. They can also reasonably demand that we base our treatment of them on how they are now and not merely on

how they were in the past. It is in general more reasonable to claim that you ought to be able to have acted culpably in the past without ongoing consequences than it is to claim that you ought to be able to continue acting culpably in the present without consequences. It is for this reason that residual culpability justifies weaker liabilities than primary culpability.

Our intuitions tell us that residual culpability is not constant, but rather declines as time goes on. Moments after committing a culpable action, a person is more culpable than that same person a week after the action. Residual culpability declines until it disappears and is replaced with mere historical culpability: the claim that the person was once culpable. We can also see this decline in our intuitions about punishment. When a wrongful action is in the distant past, it seems less reasonable to impose harsh punishments for it.

This fits with our theoretical justification for liability. When justifying liabilities, we have been considering the various perspectives from which these moral rules could be objected to. Consider a single culpable person cut into time slices: the person at t^1 , and the person at t^2 , and so on. Each of these slices is a perspective from which the person, at that time, could reasonably object to a moral rule. Suppose we propose that the person should be liable from t^1 to t^5 . There are then five perspectives from which this proposal might be objected to. Alternatively, if we propose that the person should be liable from t^1 to t^{10} then there are ten standpoints from which this proposal might be objected to. This captures the fact that a liability has a greater impact, and is therefore harder to justify, the further into the future it extends.

From any one of these time segmented standpoints the force with which the person can reasonably object to being treated as liable can vary. Some things that might matter include changes to the person's personality and values, remorse or pride in the intention, and attempts to remedy the consequences of the intention. These things might matter for liability because they can make it more or less reasonable for the person to object to being treated according to their past intention. For

example, if the person has reformed, then it is more reasonable for him to object that there is nothing he can now do to change the fact that he was once culpable. By contrast, it is less reasonable for an unrepentant person to make this objection. These things also matter for residual culpability because they affect how reasonable it is to attribute the past wrongful intention to the present person.

4.2 The Problem of Long Threats

It might be objected that while this approach does a good job of restricting liability in most cases it does too good a job of restricting liability in cases of long threats. These are cases in which a long time passes between the moment when a person culpably launches a threat and the moment when the threat will potentially cause harm. These are tricky cases for any theory of liability, and my intuitions in these cases are not always clear. In this last section I discuss some of the problems raised by these cases. Doing so provides an opportunity to revisit the relationship between duties, culpability, and liabilities.

We can begin with an example. Consider,

Long Threat. Nearly forty years ago Older planted a bomb with a time delayed fuse. We can prevent this bomb from going off, and save a victim from being killed, but doing so requires us to sacrifice someone's life.

This is a difficult case. If there were no time delay, or only a short one, it would be clear that Older can be permissibly sacrificed to save the victim. The intervening time makes things less clear.

It is possible that Older is still primarily culpable. This might be the case for two reasons. First, if Older has been patiently waiting for the bomb to go off, it is possible that Older is still intending to kill the victim. This present intention can make Older primarily culpable. Second, even if he no longer intends to kill the victim, Older may be primarily culpable for his present failure to rescue the victim. This can be true even if Older has forgotten that he planted the bomb. Older

ought to remember and realize that he must save the victim, a failure to do this can be primarily culpable.

To the extent that Older's liability in this case is generated by culpability, the swapping principle still applies. Suppose that Older had a rival bomb maker forty years ago who also planted a bomb with a time delayed fuse. If both these people are still culpable, then it should not matter which of them happened to plant the bomb that currently threatens the victim. In fact, if the other bomb maker has not reformed at all, while Older is now much changed and deeply repentant, then it may be better to sacrifice the more culpable bomb maker even if he is not causally responsible for the current threat.

If Older is no longer culpable, then he may still be liable if his liability depends on something else. I am unsure about this, but I will not be able to pursue the issue further here. My focus in this thesis is on culpability and its role in generating liabilities. I have left open the possibility that other things generate liabilities, and discussed few plausible examples, but I do not claim to know what those things are or how those liabilities work.

5. Conclusion

In the preceding chapters I have drawn attention to some reasons for doubting the significance of causal responsibility in the justification for liabilities. This chapter builds on those doubts with a sustained argument. In Chapters Two and Three I argued that culpability can justify liability even without causal responsibility. In this chapter I have argued that causal responsibility for a threat is not a plausible necessary condition of liability. This is a fairly limited claim. It leaves open the possibility of many other roles for causation and causal responsibility in the ethics of defensive harming. Yet this claim is significant because it undermines widely held assumptions about the justification for defensive harming. It is also significant because it suggests that culpability alone is

sufficient to justify liability to defensive harming. In the final two chapters of this thesis I discuss the nature of culpability in more detail. In particular, I discuss the kind of primary culpability that justifies liability.

Chapter Five:

Culpability in Theory

Introduction

In “Moral Luck,” Nagel complains that Williams neglects some of the most important roles that culpability plays in our lives.⁸⁶ Williams had focused on the nature of regret, and the way we assess the success of our lives and projects. This is only one aspect, Nagel argues, not even a very important aspect, of culpability. The core of culpability lies elsewhere, in the ultimate “moral assessment.”⁸⁷ The same critique might be offered to Nagel. In focusing on the conditions for ultimate moral assessment, Nagel neglects the more practical role that culpability plays in judgments of proportionality and permissibility. This pattern can be found across the literature on culpability. We disagree as much about what culpability is and does, as about how it applies in specific cases.

Consider just three such axes of disagreement. The first concerns the object of judgments of culpability. Some argue that culpability is fundamentally an assessment of the quality of our choices and actions, while others argue that culpability depends on the quality of our characters and dispositions. Within each of these camps there is further disagreement about what in particular should be assessed. Choice theorists, sometimes called volitionists, disagree over whether acting, willing, trying, or intending, is the most fundamental.⁸⁸ Character theorists, sometimes called attributivists, disagree over which of our beliefs, desires, inclinations, or instincts can be assessed by culpability, and over whether culpability must assess the whole, or only part, of our characters.⁸⁹

⁸⁶ Bernard Williams and Thomas Nagel, “Moral Luck” *Proceedings of the Aristotelian Society, Supplementary Volumes* 50 (1976).

⁸⁷ Nagel, “Moral Luck,” 138.

⁸⁸ See for example, Michael Zimmerman, *Living with Uncertainty* (Cambridge UP, 2008) and “Moral Responsibility and Ignorance,” *Ethics* (1997), 410-426. Also Douglas Husak, *Ignorance of Law* (OUP, 2016); Neil Levy, “The Good the Bad and the Blameworthy;” and in a more skeptical vein Gideon Rosen, “Skepticism about Moral Responsibility,” *Philosophical Perspectives* (2004) 295-313, and “Culpability and Ignorance,” *Proceedings of the Aristotelian Society* (2003).

⁸⁹ See for example, Holly Smith, “Culpable Ignorance,” *The Philosophical Review* (1983) 543-571; and “Non-Tracing Cases of Culpable Ignorance,” *Criminal Law and Philosophy* (2011) 115-146. See also my discussion of this distinction in Chapter Six 1.3 and 1.4.

The second axis of disagreement concerns the background conditions that are taken into account when assessing whatever is the object of culpability. Volitionists tend to have a narrow focus and look only at the person's beliefs and intentions at a given time. Attributivists, by contrast, are comfortable taking a wider view and assessing the person's whole situation, and even their past, when assessing culpability.

Finally, the third axis of disagreement concerns the subjectivity of culpability judgments. Subjectivists claim that culpability depends only on what is internal to the person, such as her will, her intentions, or her character. Objectivists disagree and assert that culpability can depend on the consequences a person causes to occur in the world. For example, they might claim that a successful murderer is more culpable than an attempter, simply because he causes a death. These are just a few of the central issues over which culpability theorists profoundly disagree.

How are we to make progress in the face of such disagreement? One promising approach is to recognize that there may be many kinds, or parts, of culpability. This approach was advanced by Garry Watson, who first distinguished volitionism and attributivism as two different approaches to culpability, each applicable to a different context.⁹⁰ Unfortunately, these terms have come to be used to distinguish two rival views of culpability. This need not be the case. We should recognize that culpability plays many very different roles in morality, and that each different role may bring us to a different place on these axes of disagreement. In many cases, debates over what culpability is about obscure more fundamental differences over what culpability is for.

These different kinds or parts of culpability are united by the fact that they are all different ways of talking about fault. In any context, fault is a description of an insufficient response or sensitivity to the reasons that apply in that context. There are many different types of fault because there are many different types of reasons. Yet just as there are different types of reasons, there are

⁹⁰ Garry Watson, "Two Faces of Responsibility," *Philosophical Topics* (1996) 227–248.

also many different ways of measuring our responsiveness to these reasons. Thus, in every context there will be many different ways of talking about fault that must be distinguished.

We can distinguish different kinds of culpability, by looking at the different ways that we use culpability. For example, consequences might matter for the kind of culpability involved in judgments of regret, or of how well a life has gone. Yet, as I will argue, a subjectivist approach is needed for the kind of culpability that we use to make judgments of proportionality. This need not be a contradiction. We are simply appealing to different kinds, or parts, of the multifaceted concept of culpability. There is no single measure of culpability, because there is no single thing that culpability is for.

I am not proposing a functionalist definition of culpability, where culpability just is whatever plays a certain role in morality. Rather I am proposing a teleological approach to distinguishing different kinds of culpability. We still begin with a basic understanding of culpability as a judgment of moral fault, but we recognize that there are different kinds of moral fault. To really understand culpability, we need to know the kind of culpability we are interested in.

This is why the discussion of culpability has been placed at the end of the thesis. I believe that understanding the role of culpability in cases of self-defense and liability is very useful for understanding the nature of culpability itself. This somewhat unusual approach is justified because it is so difficult to make progress understanding culpability in any other way. Rather than deciding what culpability is in order to know what it does, we must first understand what culpability does so that we can work backwards to what it is.

In this chapter, I articulate a moral reason based account of the kind of culpability that matters for judgments of proportionality and liability in cases of defensive and preventive harming. I do not claim that this is the most important kind of culpability, but it is certainly among the most important. Culpability is centrally important to the ethics of self-defense, punishment, liability, and

compensation. To understand these pressing practical questions, we must understand the nature of the relevant sort of culpability.

In the first section, I sketch the basics of my view of culpability and connect it with my account of the justification for liabilities. In section two I argue that intentions should be the main focus of this kind of culpability. In section three I consider the relationship between culpability and permissibility and argue that the two are closely linked. In section four I consider a special set of excuses that do not deny culpability, but instead make a claim about the person's identity and personality. Finally, in section five, I return to the problem of moral luck and discuss how a culpability based justification for liability is compatible with moral luck.

1. Moral Reasons and Liability

We can understand fault, in any context, as a failure to respond appropriately to the reasons we can be expected to respond to in that context. Thus, moral culpability can be understood as a failure to respond appropriately to moral reasons. Yet we need to know what kind of moral reasons, and what kinds of responses, matter. This is where different theories of culpability come apart. On my view, for the kind of culpability that matters for liability, culpability consists in an insufficient sensitivity to action-guiding moral reasons. To justify this approach to culpability, we must return to the justifications for liability.

1.1 The Significance of Choice

In Chapter Two, I elaborated culpability based justifications for liability, based on an updated version of Scanlon's significance of choice argument. This argument begins with the premise that, for several reasons, a person's options matter when deciding how we should respond

to that person's choices. In particular, the less reason a person has to do something, the more reasonable it is to attach consequences, such as liability, to doing that thing.

If a person has strong moral reasons not to do something, and particularly if doing it would be impermissible, then the person can make very little reasonable complaint to costs being attached to doing that thing. This is because, to avoid such choices, the person would only have to do what they are already obligated to do. Another way of putting this is that the quality of the opportunity a person has to avoid certain choices matters for how we should respond to those choices.

As I showed, it does not matter whether the person could have acted otherwise, or what would have happened if she did. Rather, what matters is the quality of opportunity the person had to act otherwise, given her available evidence. The quality of this opportunity depends, in large part, on the person's subjective action-guiding moral reasons. When a person does something culpable, she acts contrary to these reasons. Consequences such as liability can be justifiably attached to such choices because, and to the extent that, these are choices that the person's subjective moral reasons gave her an excellent opportunity to avoid.

Following this argument, the kind of culpability that justifies liability is fundamentally based on choices. Indeed, this is just what Watson suggested. An updated version of Scanlon's argument can show why, as Watson noticed, the kind of culpability that matters for liability is a choice based theory of liability. The central premises of this argument all concern the costs of avoiding certain choices, and the permissibility of adding further costs to such choices. People who act impermissibly, and make choices that they have subjective moral reason to avoid, can make very little reasonable complaint if they are treated as liable as a result. Culpability justifies liability, in this argument, because culpability involves an insensitivity to moral reasons, and people cannot reasonably complain when costs are attached to options that they have strong moral reasons to avoid.

1.2 Action-guiding Reasons

We can summarize this point by distinguishing between different sorts of moral reasons. The kind of culpability that justifies liability is a failure to respond sufficiently to subjective action-guiding moral reasons. These reasons are action guiding because they tell us what to do, not what to believe, to think, or to desire. They are the reasons relevant to moral judgments of permissibility, impermissibility, obligation, and other similar categories. They are subjective because they are the action-guiding moral reasons that we have given our available evidence.

Action-guiding reasons are directly relevant to what we ought to do. They differ from other sorts of reasons, such as epistemic reasons, which concern the fit between some mental state, such as a belief, and something about the world. We can distinguish action-guiding reasons from other reasons by considering a special set of cases known as wrong kind of reason cases.⁹¹ Consider,

Perverse Desire: In the future, chips can be inserted into our brains that allow machines to respond to our desires without us even intending to fulfill them. Chip, a person who has such a chip, is captured by a villain and confronted with a villainous machine, a puppy, and a kitten. Villain explains that the machine will give the kitten a painful shock unless Chip desires that the puppy is shocked, in which case both the puppy and the kitten will be unharmed. Should Chip desire that the puppy be shocked?

The most obvious answer to this question is yes. If he can, Chip should desire that the puppy is shocked. However, this does not mean that the puppy being shocked is morally desirable. Even from Chip's perspective, the puppy being shocked is not desirable. There is thus a less obvious sense in which Chip should not desire that the puppy be shocked.

⁹¹ The classic wrong kind of reason case was developed by Roger Crisp in "Review of Jon Kupperman, *Value...and What Follows*," *Philosophy* 75 (2000) 458-462. I am using this type of case for a different purpose. For an overview of the original context see Pamela Hieronymi, "The Wrong Kind of Reason," *The Journal of Philosophy* 102, 9 (2005) 437-457.

We can distinguish these two senses by considering the types of reasons that apply. From the perspective of action-guiding moral reasons, Chip ought to desire that the puppy be shocked. This is because having this desire will bring about the best outcome. However, this does not mean that Chip has reason to desire that the puppy is shocked. The shock would be painful for the puppy and is for this reason is not desirable. Thus, although Chip has reason to try to desire to shock the puppy, and reason to desire that he desire to shock the puppy, he does not have reason to desire to shock the puppy. From the perspective of desirability, desiring to shock the puppy would be unreasonable.

These kinds of cases reveal an important difference between action-guiding reasons and other reasons. Our reasons to do things, including our reasons to try to desire things, are sensitive to perverse consequences in a way that our reasons to desire things are not. The villain's machine can give Chip a reason to try to desire to shock the puppy, but it cannot give him a reason to desire to shock the puppy.

Once we have separated reasons to desire things from action-guiding moral reasons, we can see that it is only action-guiding moral reasons that matter for the kind of culpability that justifies liability. We can see this by considering,

Perverse Desire (Success Variation): The same as perverse desire, except that Chip now tries to save the kitten by desiring that the puppy be shocked. He forces himself to focus on feelings of anger and cruelty. He reminds himself of the time he was bitten by a dog as a child, and of Descartes's argument that animals work by a kind of clockwork. He thinks about how funny the puppy would look jumping back from the shock. Through an incredible mental effort, Chip manages to sincerely desire that the puppy be shocked. The machine registers this desire and releases the kitten and puppy unharmed.

It seems to me that Chip is not at all liable in this case. Chip has responded admirably to his action-guiding moral reasons and has done the right thing. This is despite the fact that Chip has failed to respond appropriately to his reasons to desire things, because he has desired something that is

morally undesirable. Thus, it seems that it is action-guiding moral reasons, not reasons to desire, that matter for the kind of culpability we are investigating.

These cases show how different kinds or reasons, and thus different kinds of fault, can come apart. There is still some sense in which Chip is at fault in this case. His desires are morally faulty because they include a desire for something that is morally undesirable. Yet, intuitively, this is not the kind of moral fault that matters for liability. That kind of moral fault is a failure to respond appropriately to subjective action-guiding moral reasons. Immoral beliefs and desires can be part of the background assessment of what makes an intention culpable, but it is fundamentally our intentions and our intention formation that matter for the kind of culpability that makes us liable.

We can see the role that beliefs and desires can play as background conditions by considering a further case:

Perverse Desire (Failure Variation): The same as *Success Variation*, except that Chip's attempt to desire that the puppy be shocked does not work. Chip finds himself unable to form this perverse desire. As a result, the machine shocks the kitten.

In this case, although Chip has failed to conform to his action-guiding moral reasons he is not culpable. Chip's beliefs and desires can help explain why this is the case. Chip has a strong and justified desire that the puppy not be shocked, and a similar belief that it would be bad for the puppy to be shocked.

We can assume that the strength of this desire is part of the reason Chip was unable to sincerely desire that the puppy was shocked. Since this desire is fully justified, given Chip's evidence and from the perspective of desirability, it can form the basis of an excuse. Chip's failure to sincerely desire that the puppy be shocked is excused because he had such a strong and justified desire that the puppy not be shocked. This desire made it possible for Chip to do enough in trying to desire that the puppy be shocked and still not succeed. Beliefs and desires form important background conditions for assessing culpability. They can both make it the case that a person can be sufficiently

sensitive to their action-guiding moral reasons, despite failing to do what these reasons give him reason to do.

1.3 Subjective Culpability

By looking at the justifications for liability, we can see that the kind of culpability that justifies liability is an insufficient response to evidence relative action-guiding moral reasons. It is concerned with choice, rather than character, because it is through our choices that we respond to action-guiding moral reasons. Using this same reasoning, we can also see that this kind of culpability requires an internal, or subjective, assessment of the person's sensitivity to action-guiding moral reasons. It is a measure of the person's sensitivity to, rather than conformity with, action-guiding moral reasons. In my view, consequences do not matter for the kind of culpability that justifies liability.

This is not to say that consequences, and conformity with action-guiding reasons, do not matter for any sort of culpability. It may be appropriate to respond differently to our own bad choices, and those of others, when those choices result in serious harms. This is the point that Williams defends in "Moral Luck." We understand our own action, and that of others, in large part by looking at consequences. As Anthony Duff points out, "our understanding of human action is not subjectivist."⁹² We should not expect our understanding of culpability to be purely subjectivist either.

My approach to culpability embraces this point as a part of the multifaceted nature of culpability. We can accept without contradiction that the concept of permissibility includes both fact relative and evidence relative permissibility. It should be no more difficult to accept that the concept

⁹² Anthony Duff, *Criminal Attempts* (OUP 1997), 347.

of culpability can include both subjective and objective culpability. The question we must investigate is not which of these sorts of culpability exists, or is the authentic sort of culpability. I am not even sure how we could answer such questions. Rather, the question we must investigate is what these different sorts of culpability are useful for. At least for the kind of culpability that justifies liability, we have good reason to suppose that a strictly subjectivist approach is needed.

We can see this by looking at the arguments I have elaborated in the preceding chapters. The theoretical justifications I have given for liability depend on the person's sensitivity to moral reasons in their decision making, not the consequences of the person's actions. Indeed, the significance of costs argument suggests that it would be unfair to treat successful attempters as any more liable than unsuccessful ones. This is because the costs of avoiding making a successful attempt are in no way greater than the costs of avoiding making an unsuccessful one.

In Chapter Four I offered a further series of arguments to show that causal consequences of our culpable choices are not a necessary condition for liability. I defended my culpability based view of liability against a moral responsibility based view of liability, according to which causal consequences are important. In making these arguments I appealed to a subjectivist account of culpability. If these arguments are correct, then we have further reason to suppose that subjective culpability is sufficient to justify liability. I considered a few ways in which consequences could matter for liability, but these arguments appealed moral responsibility rather than an objectivist sort of culpability.

The centrality of culpability to liability judgments is most clearly manifested in the relation of proportionality. This is the relationship between the harm that can be permissibly imposed on a person, the harm thereby prevented, and the degree of that person's culpability. As the person's culpability increases so does the harm that can permissibly be imposed relative to the same harm prevented. I believe that this relationship is not simply a constraint on liability, it is the essence of

liability. Even if my other arguments are mistaken, proportionality still describes a crucial link between culpability and liability.

The kind of culpability that matters for judgments of proportionality must be subjectivist. It must not depend on the causal consequences of the person's choices or intentions. This is because of the role it plays in judgments of permissible defensive harming. We need to know what it is proportionate to do, given the person's culpability, before we can decide whether it is permissible to prevent the person from causing some harm. This calculation would produce absurd results if the fact of the person causing harm would make the person more culpable.

For example, if we used an objectivist sort of culpability to determine proportionality then the less likely a given defensive action was to succeed, the more harm could justifiably be inflicted as part of that defensive action. This is because as the likelihood of preventing a wrongful attack decreases, the more culpable we can assume the attacker is likely to be from an objectivist perspective. The more culpable we can assume the attacker to be, the more harm we can proportionately inflict. Thus, absurdly, a large amount of defensive harm might be proportionately imposed on an attacker only if it were unlikely to successfully prevent the attacker from causing harm, while the same amount of defensive harm would be disproportionate if it were more likely to succeed.

Using an objectivist sort of culpability for judgments of proportionality would yield perverse consequences and set up a direct and undesirable conflict between the constraints of proportionality and necessity. Thus, without denying the centrality of consequences to our lives and to a variety of moral appraisals, we should accept that a subjectivist measure of culpability is needed for judgments of proportionality and liability. We should be open to the possibility that subjectivist culpability is useful and important for morality, even if it is less useful for evaluating the moral quality of our lives.

2. The Object of Culpability

We can understand culpability, or at least the kind or part of culpability that justifies liabilities, as a measure of the person's internal response to subjective action-guiding moral reasons. A good way to describe this is as a measure of the person's sensitivity to action-guiding moral reasons. This makes it clear that we are interested in the impact these reasons have in the person's decision making, rather than the impact the person's decisions have on the world. A person responds appropriately to moral reasons from a subjectivist perspective, when the person is sufficiently sensitive to those moral reasons. This section investigates the kind of sensitivity that matters.

I have focused my account of culpability on moral reasons. However, this does not mean that people have to be aware that they are responding to moral reasons. Talk of moral reasons just describes the role that different morally significant considerations ought to have in our deliberations about what to do. The fact that kicking someone will cause them pain gives us a moral reason not to kick them. This does not mean that we must decide not to kick them because we have a moral reason not to kick them. That adds an unnecessary step. We should decide not to kick them, because kicking them will cause them pain. In reasoning this way, we can be sensitive to moral reasons, without thinking about moral reasons at all.

There are many ways we can be sensitive to action-guiding moral reasons, and not all of them are directly relevant to culpability. For example, in order to have justified moral beliefs we ought to be sensitive to action-guiding moral reasons in deciding what to believe. This kind of sensitivity is not what we have in mind when evaluating the quality of our sensitivity to action-guiding moral reasons. We have epistemic reasons to be sensitive to our action-guiding moral reasons when forming beliefs about action-guiding moral reasons. The relevance of sensitivity to action-guiding moral reasons to belief formation is indirect, because it depends on our epistemic

reasons. Belief formation is not what action-guiding moral reasons are about, and it is not the part of us to which they apply.

In forming an immoral belief, we are not directly failing to respond to moral reasons. If we are at fault for forming such a belief, it is because we are directly failing to respond to epistemic reasons. By contrast, in forming an immoral intention we are directly failing to respond to the moral reasons. We can see this contrast by considering the sense in which a belief and an intention can be wrong. If an intention is wrong, it is because it goes against our moral reasons. If a belief is wrong, it is because it is untrue or against our epistemic reasons. The sorts of things that can make an intention wrong are different from the sorts of things that can make a belief wrong.

What matters for culpability is our sensitivity to action-guiding moral reasons in that part of ourselves to which these reasons directly apply. In other words, it is our sensitivity to action-guiding moral reasons in that part of ourselves in virtue of which we can be said to have action-guiding moral reasons. Loosely speaking, this is our capacity to make decisions, form intentions, and act in response to reasons. We can be said to have action-guiding moral reasons to do things, because we have this capacity. Other things, like falling boulders, do not have moral reasons to do things because they lack this capacity.

We can call this capacity agency. So we can say that culpability, of the type we are investigating, consists in an insufficient sensitivity to action-guiding moral reasons in the exercise of agency. A complete account of how this capacity works would be an account of agency. I do not have such an account to offer, so I will have to make do with a rough intuitive account of what agency involves. I am assuming that it involves guiding our actions through intentions and decisions that are formed in response to our perception of reasons for action.

Choice theorists, or volitionists, about culpability sometimes disagree over which precise exercise of agency should be the object of culpability. It might be our decisions, or our intentions, or

what we will, or what we try to do, that is the ultimate basis for culpability judgments. My account of liability will not help us make much progress on this question. Indeed, I do not think it is not necessary for us to choose a single type of mental state that is the ultimate source of our agency.

The phrase ‘ultimate source of our agency’ may seem unacceptably vague and metaphorical, so I will try to be clear what I mean. Most things that we can be said to control we control by virtue of a causal relationship with something else. I control the direction of my boat by moving the rudder, and I control the movement of the rudder by moving my arm, and I control the movement of my arm through my willing or intending to move it. Yet while I can be said to control my will, or my intentions, I do not control these things in virtue of anything else. Rather, I only control other things insofar as they are responsive to my intentions or my will. These mental states are the ‘source’ of my agency because they are causally as far upstream as my control goes.

Of course, it is controversial whether mental states, such as intentions, can on their own cause bodily movements. It might be better to say that action is the consequence of either intentions or of the brain states that correspond to, instantiate, or are supervened on by, intentions. Which of these descriptions is best depends on questions in the philosophy of mind that I will put to one side. For simplicity, I will continue to describe willing and intending as causes, but this can be understood as a shorthand for the longer statement given above.

I do not wish to offer a theory of the source or nature of our agency. So I will not insist that any particular mental state is the ultimate source of our agency. Indeed, it strikes me as plausible that our agency may have multiple sources. If we fail to will something that we earlier intended to do, this act of will can be a source of agency and something that we do not control in virtue of controlling anything else. This can be true even if at other times we can control what we will through our intentions and decisions. These are difficult questions and interesting questions, but I

will not pursue them further here. For simplicity, from here on I will generally refer to our choices and intentions as the exercises of our agency evaluated by judgments of culpability.

2.1 Actions and Intentions

There are many different things that might count as an exercise of agency. In theory it is plausible to suppose that actions, intentions, tryings, willings, and perhaps even reflexes can be considered exercises of agency and a basis for culpability. However, I have also argued that the kind of culpability that justifies liability is a subjectivist sort of culpability, one that depends on a person's sensitivity to action-guiding moral reasons, not on the consequences the person causes to occur in the world. We must be particularly careful when assessing the culpability of actions, because any sort of action, even simple bodily movements, can function as a consequence from the perspective of a subjectivist theory of culpability.

For this reason, it may be better to focus on intentions and other such mental states, even though actions may also be considered exercise of agency. To see this, consider the following pair of cases:

Two Buttons. Mistaken has two buttons she could push. She reasonably believes that if she pushes a button labeled TREAT, a kitten will receive a delicious treat, and if she pushes button labeled SHOCK, the kitten will receive a painful shock. Unbeknownst to her, a villain has rewired the buttons so that TREAT shocks and SHOCK gives the treat.

This is exactly the sort of case that demonstrates why consequences do not matter for culpability. If Mistaken pushes TREAT, and shocks the kitten, this horrible consequence does not make Mistaken culpable. Now consider a variation of the same case:

Two Buttons (Spinal Variation). Patient has two buttons she could push. She correctly believes that if she pushes a button labeled TREAT, a kitten will receive a delicious treat, and if she pushes button labeled SHOCK, the kitten will receive a painful shock. Unbeknownst to her, a villainous doctor has tampered with her

spinal column so that impulses sent to her right hand actually go to her left, and impulses sent to the left actually go to the right.

Suppose that Patient has her left hand resting on TREAT and her right hand resting on SHOCK. When Patient decides to give the kitten a treat by pressing TREAT with a finger on her left hand, the impulse is actually sent to Patient's right hand, so that Patient presses SHOCK instead.

The simplest way to analyze this variation is to offer the same analysis that worked for the first case. Indeed, the two cases have a nearly identical structure. Patient, like Mistaken, is not culpable because she could not have known that trying to push TREAT would actually send a shock. Yet if we focus on acts, even just on bodily movements, then this analysis is unavailable.

This is because Patient's bodily movement, pressing the SHOCK button with her right hand, was actually subjectively wrongful. This movement was subjectively wrongful because Patient knew that pressing SHOCK would harm the kitten. Patient's exculpatory ignorance in this case is ignorance about the consequences of *trying* to press each button, it is not ignorance about the consequences of moving her fingers to press either button. In order for this ignorance to be exculpatory, it must be the *trying*, not the actual movement, that is the object of culpability.

By focusing instead on intentions, or other such mental states, we can explain both variations of *Two Buttons* with the same simple analysis. Both Mistaken and Patient intended to give the kitten a treat by pressing TREAT. This intention was not subjectively wrongful in either case, so neither Mistaken or Patient is culpable. Because of unforeseeable villainous tampering, both Mistaken and Patient caused the kitten to be shocked, but this bad consequence does not matter for their culpability. This analysis is satisfying because it treats both cases the same way. It also avoids the bizarre consequence of supposing that Patient might be culpable for moving a finger she was not trying to move.

To this it might be objected that the separation between intentions and actions is not as clear as we like to suppose. For example, Duff makes a compelling case that intentions cannot even be said to cause actions because it is not possible to define intentions except in terms of actions.⁹³ On this view, intentions or willings just are the mental part of actions. This may be the case, but it does not undermine my claim that we should focus on mental states rather than actions. Even if intentions can only be understood as the mental part of actions, it is that part of actions that we are interested in. This is because it is that part of action that is sensitive to action-guiding moral reasons and insulated from causal consequences in the way required for a subjectivist account of culpability. For simplicity, in the remainder of the thesis I will generally refer to intentions as the basis of culpability.

2.2 The Fallibility of Intentions

Cases like *Two Buttons* can lead us to suppose that our intentions are infallible in a way that our actions are not. In *Two Buttons* the villain can change the moral quality of the person's action, but the villain cannot alter the person's intention. However, we only have to change the case a little bit and the villain could just as easily interfere with Patient's intentions. Consider,

Intentional Interference: Unless Victim presses a button, a kitten will receive a painful shock. Victim is well motivated and would like to save the kitten. However, whenever Victim is about to form an intention to press the button Villain plays a loud tone that temporarily distracts Victim. For this reason, Victim is unable to intend to save the kitten.

This is only one of many ways in which a person's intention formation can be interfered with. We can imagine intentions being blocked, as in this case, or implanted, as in a case of hypnosis. While these cases are far-fetched, they are no harder to imagine than cases like *Two Buttons*.

⁹³ Anthony Duff, *Criminal Attempts*, Chapter 11.

These are the sorts of cases that might lead us to suppose that we should prefer a character based theory of culpability. A character theory seems preferable because the victim in *Intentional Interference* is clearly not culpable for her failure to intend to save the kitten. This might show that culpability measures the quality of our characters, not our choices. Victim is not culpable in this case because she still has a good character and would save the kitten if not for the interference. Yet, while there may be contexts in which an assessment of the moral quality of our character is useful, we have good reason to suppose that the kind of culpability that justifies liability depends on choices not character.

We can resolve this problem by showing how a choice based theory view of culpability can respond to cases like *Intentional Interference*. A person is culpable, in the way that matters for liability, to the extent that she is insufficiently sensitive to subjective action-guiding reasons in the exercise of her agency. If Victim was not interfered with, and still refused to try to save the kitten, then she would be culpable because of her failure to respond to the kitten's suffering. However, in *Interference* the Victim's failure to try to save the kitten is not a result of insufficient concern for action-guiding moral reasons. It is for this reason that the Victim is not culpable.

We have good reason to take a volitionists, or choice based, approach to the kind of culpability that justifies liability. However, this does not mean that we must take a volitionist approach to the background conditions that matter when assessing the quality of a person's choice. To evaluate the sensitivity of a person to action-guiding moral reasons we need to look at more than the content of the person's beliefs and intentions. We need to look at the person's whole situation and consider the quality of the person's choices as a response to that situation. Culpability measures the person's sensitivity to her action-guiding moral reasons in the exercise of her agency. This requires us to look beyond what the person wills or intends and consider how the person came to

form that will or intention. It is not just the quality of the will that matters, but the quality of the will as a response to evidence relative action-guiding moral reasons.

3. Culpability and Permissibility

People can be more or less sensitive to their moral reasons. Unless we accept that everyone is always slightly culpable, there must be a threshold for culpability. People can respond imperfectly to their moral reasons without being culpable, so long as they respond appropriately enough. I will not attempt to describe where this threshold is. However, to understand culpability we must consider how this threshold relates to another important threshold.

For many, the moral world is divided by the standard of permissibility. Impermissible actions are those that morality forbids, permissible actions are those that morality tolerates, requires, or recommends. If some action is morally required, then doing something else is impermissible. These basic deontic categories structure the way we understand morality and our obligations to one another. These categories can be applied in a fact relative sense, or in an evidence relative sense. An act is evidence relative impermissible for some person if it would be fact relative impermissible if the world were as the person's evidence should lead him to suppose.

We can describe these categories further in terms of moral reasons. Actions can go against the balance of moral reasons and still be permissible. This is the case when I choose to do something that is permissible instead of something that is supererogatory. However, if an action goes against a sufficient balance of moral reasons it is impermissible. Similarly, an action can be supported by the balance of moral reasons without being required. Yet, if an action is supported by a sufficient balance of moral reasons then it is required. In such a case doing otherwise would be impermissible. Thus, there is a threshold of sufficient response to moral reasons built into the threshold of permissibility.

In the case of evidence relative permissibility, it seems natural that this threshold should be the same as the threshold for culpability. This is for two reasons. First, both thresholds describe the same thing. They are both measures of the sufficiency of a person's subjective response to moral reasons. Specifically, they both measure whether the person has done enough in her subjective response to these reasons.

Second, if permissibility is not tied to culpability, then it is less clear what claims of permissibility are about. It makes sense to suppose that an action is morally required just when it is required to avoid culpability, or that an action is permissible just when doing it would not be culpable. Without linking to a standard such as culpability, I am not sure what it could mean to say that an action is simply required. There may be other possible standards, but culpability is the most natural choice. In this way, the standards of culpability and permissibility can support and give meaning to each other.

There is one possible exception to this, as I will discuss in the next chapter. Some ignorant people may perform actions that are evidence-relative impermissible, despite not being culpable. Thus, subjective impermissibility can only be a necessary, rather than a sufficient condition of culpability. Strictly speaking, in these cases, morally required actions are required in order to avoid satisfying a necessary condition of culpability. Nevertheless, there remains a close link between culpability and permissibility. Both these concepts are made clearer when they are linked to each other.

The kind of culpability that is tied to permissibility should be the kind of culpability that justifies liability: a choice based subjectivist kind of culpability. It is clear that the kind of culpability linked to permissibility must be choice based, since permissibility is a judgment that applies to choices not characters. In addition, the kind of culpability linked to evidence relative permissibility must be subjectivist, it cannot depend on causal consequences. This is because evidence relative

permissibility evaluates actions from the perspective of the person before he acts. What actually happens, or would actually happen, is not relevant to evidence relative permissibility. For this reason, it cannot be relevant to the kind of culpability that is linked to the standard of evidence relative permissibility.

It might be objected that claims about permissibility are normally applied to actions as well as decisions and intentions. Yet, I have argued that culpability is more an assessment of our intentions than it is of our actions. This does not stop us from linking the threshold for permissibility with the threshold for culpability. We can suppose that actions are impermissible when they result from and are guided by culpable and impermissible choices.

This is a tidy and appealing way of relating permissibility with culpability. Moreover, it allows us to more easily and simply explain what claims about permissibility really mean. Yet as obvious as this link might appear, it requires us to reevaluate a few widely held beliefs about permissibility and excuses.

It is often assumed that the standard of permissibility should apply equally to all people. If two people have the same beliefs and evidence, then it is tempting to assume that the same actions should be permissible and impermissible for each of them. In contrast, judgments of culpability are tailored to each person and take into account diverse aspects of the person's psychology. For this reason, it might be objected that even if the threshold for culpability and permissibility coincide in standard cases, they are not fundamentally the same.

However, by thinking about specific cases we can see that this is not correct. We can retain both the assumption that permissibility applies equally to all people, and the thought that culpability judgments are tailored to individuals. Consider,

Artist: Artist loves his work and has a passionate desire to devote himself to it. This desire is neither moral nor immoral, it is just part of the way that Artist is. When Artist promises to do something for a friend he often makes the painful decision that he must break his promise so that he can devote himself to his work.

An assessment of Artist's culpability for breaking his promises would take into account his intense desire to keep working. This is an example of how culpability judgments are tailored to individuals. Given the strength of this desire it is plausible that Artist could give his promises sufficient weight in his decision making and still decide to break his promise. Thus, it is possible for Artist to break his promises without being culpable in circumstances where a similar person without Artist's desire to keep working would be culpable.

It seems to me that we ought to reach the same conclusion about the permissibility of Artist's promise breaking. The fact that keeping his promise would mean giving up on his desire is the sort of consequence that should be part of an assessment of permissibility. This factor is only relevant to Artist because of his psychology and would not apply to another person. In this way, while the same standard of reason responsiveness applies to all people, deciding whether a person has met that standard requires us to look at the person's psychology. This is true for both culpability and permissibility. It gives us a further reason to suppose that these two thresholds are the same.

Making this link also has consequences for the way we think about some excuses. It is tempting to suppose that a person who acts under a threat of severe harm is not culpable even though what she does is impermissible given the available evidence. In order to link the threshold of culpability with the threshold of permissibility, we must reject this description. We can still observe that the action would be impermissible if it were not performed under duress. However, if the threat of harm is sufficient to make doing the action non-culpable, then it ought to be sufficient to make the action permissible, under the circumstances.

We can illustrate this with an example. Consider,

Duress: Curator is threatened by a villainous art thief. Unless Curator disarms the alarm, and allows villain to steal a painting, villain will break Curator's fingers. Curator allows the painting to be stolen.

In this case we can assume that things would have gone better overall if Curator had heroically refused to allow the painting to be stolen. Villain's threat gives Curator an excuse, not a justification. In addition, we can assume that the duress is not sufficient to render Curator unable to make decisions and respond to reasons, so Curator is still acting as a morally responsible person when she allows the painting to be stolen.

In this case the threat of harm removes Curator's culpability because it makes it possible for her to give enough weight to her moral reasons to prevent the theft while still deciding to act against them. When we say that Curator is excused because of her duress we mean that, under the circumstances, she responded appropriately enough to her moral reasons.

This same judgment should lead us to suppose that what Curator did was, under the circumstances, permissible. For a person under duress to do more to respond to her moral reasons would be supererogatory, because it would go beyond what we can reasonably expect. If doing more would be supererogatory, then it cannot also be morally required. Thus, people who act non-culpably under duress must also act in a way that is evidence relative permissible. Such actions would not be permissible if it were not for the duress, but they are permissible under the circumstances.

Above, I argued that permissibility should be tied to culpability, because it allows us to easily explain what judgments of permissibility and moral obligation are about. We can easily understand what it means for an action to be morally required, if this means required to avoid culpability. Without this link, it is unclear what is meant by the claim that an action is required. Some actions may be morally better than others, but where does this bright deontic line come from, if not from culpability?

To this it might be objected that I have neglected another possible standard that could be linked to permissibility. This is the standard of respecting the rights of others. It might be thought that permissibility is fundamentally about rights and duties, rather than culpability. If this were the case, then there might be cases where judgments of culpability and permissibility diverge.

Suppose, for example, that Artist's friends have a right to Artist keeping his promise, and that Artist has a duty to keep his promise which is a correlate of that right. If permissibility is tied to rights and duties, then we could say that Artist acts impermissibly by breaking his promise, even though he is not culpable because of his intense desire. Similarly, we might suppose that the painting's owner has a right to have his painting protected by the curator, and that curator has a duty to protect the painting which is a correlate of that right. We could say that Curator's action was impermissible because it went against her duties, even though it was excused from the perspective of culpability.

To better understand this proposal, we need to know more about rights and duties. Specifically, we need to know how far they extend when there are other factors in a case that make it justifiable or excusable to act against them. Consider the Curator's duty to protect the painting from theft. Does the Curator always have this duty, even in cases where it would not be reasonable for her to act against it, or does this duty only extend as far as it can reasonably be applied? However we answer this question, we will see that permissibility ought to be tied to culpability rather than rights and duties.

First, suppose that the Curator always has a duty to protect the painting. In some circumstances, of course, there will be strong reasons for the Curator not to protect the painting. We can assume that if Curator had to choose between saving the painting or saving a life, she ought to save the life. Doing so would clearly be permissible, even if the life Curator could save was her own.

Yet, in doing so Curator would act against her duty to protect the painting. We should conclude that Curator can act against her duty to save the painting without acting impermissibly.

Indeed, there is an important distinction in rights theory that applies to exactly this kind of situation. This is the distinction between infringing and violating a right.⁹⁴ A right is infringed but not violated when a person acts against the right while still acting permissibly. This distinction makes it clear that permissibility is not tied to simply acting against rights. Rather, permissibility is tied to being responsive enough to rights, given the circumstances. The sufficiency of a response to moral reasons is exactly what culpability and permissibly are both about.

Alternatively, we might suppose that our rights and duties only extend as far as they can reasonably be applied, so that the painting owner only has a right to have his painting protected in circumstances where it is reasonable to expect it to be protected. My own view of rights and duties is a compromise between these two. As I discussed in Chapter Three, I prefer to think that the Curator always has a general duty to protect the painting, but that this general duty only puts her under a specific duty to perform acts that she can reasonably be expected to perform.

On either of these views, we need to know what can reasonably be expected of Curator in order to know how far her duties, or her specific duties, extend. For example, we need to know both what it is possible for Curator to do and what it is permissible for her to do. Thus, on this view the threshold of permissibility is linked to rights and duties but only because duties are limited by the threshold of permissibility and obligation. This is the threshold of what we can reasonably expect from people like Curator. This same threshold, whether a person has done what we can reasonably expect from her in response to her moral reasons, is exactly what is measured by culpability.

This link between permissibility and the kind of culpability that justifies liability is important for my argument in two ways. First, it supports my view of the justification of liability because

⁹⁴ For this distinction see Judith Thomson, *The Realm of Rights* (Harvard, 1990) 122.

permissibly plays an important role in the significance of costs argument. It is particularly unreasonable for a person to object to costs being attached to impermissible choices, since these are choices that the person is already obliged to avoid. People who make evidence relative impermissible choices are culpable, and this link makes it even clearer why that kind of culpability can justify liability.

Second, this link with permissibility helps to show why the kind of culpability that justifies liability is so important. It is not only important because of the consequences it has for liability. It is also important because it is tied to, and helps to elucidate, claims about permissibility and impermissibility. Even without liability, these categories are central to our everyday experience of morality, and to the way many of us structure our moral reasoning. In this way, Culpability can make judgments of permissibility more substantive, while permissibility can make judgments of culpability more central to our everyday understanding of morality. There are many ways of measuring moral fault, and so there may be other kinds of culpability. However, the kind of culpability we are investigating is particularly useful and must be among the most important kinds of culpability.

4. Excuses that Do Not Deny Culpability

We have seen how excuses work to reduce or remove culpability. An excuse, such as ignorance or duress, shows that the person did respond appropriately to his evidence relative moral reasons even if what he did was fact relative impermissible, or would have been fact relative impermissible if it were not for his duress. In the next chapter we will discuss how this works in more detail, particularly when it comes to excuses of ignorance. However, there is an important set of excuses that does not work in this way. There are some successful excuses that do not even attempt to deny or mitigate culpability. Rather these excuses work by denying the attribution of an

admittedly culpable intention to a given person. These excuses are as much about identity and liability as they are about culpability.

To see how this works, we need to consider a set of cases that is particularly problematic for any theory of culpability. These are cases in which some sort of intervention or disruption in a person's brain directly alters that person's sensitivity to moral reasons and leads her to form subjectively wrongful intentions. For example, consider,

Super Soldier. A morally normal person, Soldier, is drafted into the army of a technologically advanced state. As part of his training Soldier undergoes a procedure that temporarily alters his brain in a way that makes him far more violent and less sensitive to the suffering of others. When Soldier has finished his tour of duty the procedure will be reversed. However, while on duty Soldier wrongfully kills innocent civilians.

In this case we can assume that Soldier would not have killed the civilians if it were not for the procedure. My intuition in this case is that Soldier should not be considered culpable for what he has done. However, this case is problematic because Soldier clearly forms a subjectively wrongful intention and fails to be responsive enough to his subjective action-guiding moral reasons. Soldier's intention to kill the civilians is thus clearly culpable according to my theory of culpability.

Things would be easier if the army simply took over Soldier's brain and used him like a puppet. In that case he would not be culpable because he would not form wrongful intentions. It would also be easier if the army simply implanted the wrongful intentions in Soldier's brain, because we could then say that the intentions do not manifest insufficient concern for morally significant considerations on the part of Soldier. What makes this case difficult is that the army interferes directly with Soldier's capacity to make good decisions, without removing that capacity entirely. By reducing Soldier's sensitivity to moral reasons, they turn him into the kind of person who is culpable.

One solution that has been proposed for these cases is to claim that Soldier is not culpable because his wrongful intentions do not reflect a sufficiently large portion of his personality or moral character. Holly Smith considers the case of a person who has their capacity to respond to morally significant considerations suppressed through hypnosis. She argues that this person is not culpable because his intention resulted from only a part of his moral character.⁹⁵

There are two problems with this solution. The first is that it will only work if a part of the person's character is being suppressed. There are other possible interventions that would still be problematic. For example, the army might add something to Soldier's moral character, rather than taking something away. Moreover, if the army simply removed or altered a part of Soldier's brain, then Soldier's wrongful intention would be reflective of the entirety of his moral character at that moment.

The second problem is with the premise that a person can only be culpable if his wrongful intention reflects all or enough of his moral character. People sometimes act impulsively and do immoral things that are out of character. A usually mild mannered man with a violent temper cannot claim that he is not culpable because his violent actions do not reflect the mild mannered part of his character. There is good and bad in every person, and sometimes people are culpable because they are motivated by the bad part of their character.

It might also be suggested that Soldier is excused because he has been deliberately manipulated, or because his decision making has been altered by external forces. Unfortunately, this solution is both too narrow and too wide. It is too wide because we are all constantly influenced, and occasionally manipulated, by external forces. Our environments shape our moral character and

⁹⁵ Holly Smith, "Dual-Process Theory and Moral Responsibility," *The Nature of Moral Responsibility: New Essays* ed. by Clarke, McKenna, and Smith (OUP, 2015) 189-190.

decision making and sometimes this influence is exerted intentionally. If all these influences removed culpability, then no one could be culpable for their actions.

This proposed solution is too narrow because it ignores cases that are similar in structure to Soldier, yet do not involve deliberate or obviously external interference. These are some of the hardest cases of culpability. Consider,

Tumor. A morally normal person, Malignant, begins to act increasingly violent and cruel. One day, he intentionally and wrongfully kills an innocent victim. Later, doctors discover a large tumor pressing on Malignant's brain that has brought about this violent disposition and wrongful intention.

This is a perplexing case. My immediate intuition is that Malignant is not culpable for a similar reason that Soldier and Citizen are not culpable. If the tumor were removed, then it would not be right to hold Malignant culpable for the intentions formed in his brain while under the influence of the tumor.

Yet how can we distinguish Malignant from any ordinarily immoral person who happens to develop or be born a violent and cruel disposition? Consider,

Tumor (Brother Variation): Malignant has a younger brother, Malicious, who was unluckily born with a genetic predisposition towards violence and cruelty. Even while healthy, Malicious has the same violent and cruel disposition that Malignant develops as a result of the tumor. One day, Malicious intentionally and wrongfully kills an innocent victim.

We can suppose in this case that Malicious is not a psychopath. Rather, he is what we might call an ordinary immoral person. The explanation for Malicious's immoral disposition is the same sort of explanation we could give for anyone's disposition: namely that it can be explained by some combination of his genetics, circumstances, and life experiences.

The contrast between Malignant and Malicious poses a dilemma. To treat them both as culpable, I would have to give up on my intuition that Malignant should not be treated as culpable for his actions. However, treating them both as non-culpable is even less appealing. If we did this, it

would mean that even ordinary explanations of the sources of intentions could be excuses. This would undermine all our practices of holding each other to be responsible. It would mean rejecting all of the intuitions that support these practices. The only way out of this dilemma is to find a principled reason for distinguishing between the two cases.

Those who are skeptical about culpability will not be surprised by this dilemma. This kind of dilemma, they would point out, is exactly why we should be skeptical about moral responsibility. It is certainly hard to argue against skepticism, since the contention that our current theories are inadequate is almost always correct. Yet, I think the practices of holding ourselves and others to be responsible, and of treating others according to the quality of their intentions, are important if not integral to the kind of lives we have most reason to value. For this reason, while I agree with the skeptics that there is much more work to be done, I believe that this work is worth doing.

We can distinguish between two different moral questions that the case of Malignant and Malicious brings up. The first is the skeptical question of why anyone should be considered culpable given the possibility of cases like this. The second is the non-skeptical question of how we can best explain the difference between Malignant and Malicious within a non-skeptical theory of culpability. I will propose a way of answering the second of these questions, but I will not directly address the first.

I propose that we explain the cases of Soldier and Malignant not by denying culpability but by denying the fairness of treating the person according to that culpability. The intentions formed by Soldier and Malignant are clearly culpable because they are clearly morally at fault and fail to respond appropriate to their moral reasons. We cannot deny the culpability of their intentions without denying the culpability of people like Malicious as well.

What makes Soldier and Malignant special, and motivates us to exonerate them, is that there is a clear sense in which their culpable intentions do not result from their normal personality. We

can describe a person's personality as that set of mental states that makes them who they are and makes them the same person over time. There is a sense in which both people are not themselves while under the influence of the army's procedure, or the tumor.

We can put distance between the person as they are before and after the interference, and the person as they are while under the influence of the interference. It is this distance that leads us to suppose that the people should not be treated as culpable. The same is not true of people like Malicious, whose current immoral intentions are just a product of his normal personality. There is no other version of Malicious we can appeal to who does not have the same immoral personality.

In this way, the excuse I am proposing is as much about personality and identity as it is about culpability. Of course, identity is a binary notion. Two things are either the same thing or they are not. However, the morally significant part of personal identity, what I have called personality, can come in degrees. It can be more or less appropriate to consider some part of the character currently animating a person's body to be a part or product of their personality.

In "Moral Luck" Nagel recognizes this intimate link between agency, identity and culpability. In fact, he speculates that this link may be one reason the problem of moral luck is so intractable. We cannot give up on our judgments of culpability because we cannot give up on our internal view of ourselves as people. This remains the case even when we realize that all actions, even our own, are simply events in the world. He writes,

We are unable to view ourselves simply as portions of the world, and from the inside we have a rough idea of the boundary between what is us and what is not, what we do and what happens to us, what is our personality and what is an accidental handicap. We apply the same essentially internal conception of the self to others.⁹⁶

⁹⁶ Nagel, "Moral Luck," 149.

We find the limits of a person's culpability at the limits of her agency and identity. For this reason, hard questions of culpability may turn out to be hard questions of agency and identity.

We can see something like this in the real case of Alien Hand Syndrome. This is a condition in which one of a person's hands performs complex and seemingly goal-oriented movements without the person being able to control the hand or having access to its intentions. The hand is clearly guided by some sort of intention because it tries to do, things like grasping doorknobs or stopping the other hand. Although the intentions that guide the alien hand are clearly a product of the person's brain, they would not be a fair basis for treating the person as culpable. This is because these alien intentions are clearly not a part of that person's personality. This shows that a person's personality need not include all the intentions that happen to occur in their brain.

To see how this thought might help to exonerate Soldier and Malignant we can consider a more analogous, if less realistic, case. This is a variation on the story of Dr. Jekyll and Mr. Hyde.⁹⁷

Consider,

Jekyll and Hyde. Dr. Jekyll innocently drinks a potion that temporarily transforms him into a very different and villainous person, Mr. Hyde. These two people share one body, but their personalities are clearly distinct. Mr. Hyde is not (as in the original story) a nearly wild beast, he is just an ordinarily immoral person who acts on some subjectively wrongful intentions.

Consider what the potion has done to Jekyll. It has not made Hyde innocent. Hyde is fully culpable for his immoral intentions. Rather the potion has disrupted the personality that normally controls Jekyll's body so dramatically that while his body is under the influence of the potion we treat it like a completely different person, who we call Hyde. This case is clear because the separation between

⁹⁷ Robert Louis Stephenson, *The Strange Case of Dr. Jekyll and Mr. Hyde* (Scribner's Sons: 1886). Husak briefly mentions the case of Jekyll and Hyde in "Intoxication and Culpability," *Criminal Law and Philosophy* 6, 3 (2012). Rosen also briefly alludes to Jekyll to explain the nature of culpability in "Skepticism about Moral Responsibility," *Philosophical Perspectives* 18, 1 (2004).

Jekyll and Hyde is so complete, and because we have a different name for Jekyll's body when it is under the influence of the potion.

Although the intentions formed by Mr. Hyde are straightforwardly culpable, we ought not to treat him as culpable. At least, we ought not to treat him as culpable if this means treating Jekyll as culpable as well. Whatever we do to Hyde we also do to Jekyll, and Jekyll has a prior and more permanent claim to be the owner of the body he shares with Hyde. Moreover, when the potion has worn off, it would not be fair to consider Jekyll to be culpable for what Hyde did. Thus, Jekyll has an excuse that allows him to deny that the intentions his body formed while under the influence of the potion can be attributed to him, without denying that those intentions were culpable. It is an excuse based on personality, not culpability.

What happens to Malignant and Soldier is simply a less extreme case of what happens to Jekyll. What these people do while under the influence of the army procedure or the tumor is clearly culpable, just as what Mr. Hyde does is clearly culpable. Yet, there is also a clear contrast between Malignant before and after the tumor, and Malignant while under the influence of the tumor. It is unfair to take the intentions formed by Malignant's brain under the influence of the tumor to be a basis for holding Malignant the person to be culpable. In the same way, it is unfair to take the intentions formed in Jekyll's brain while he is Hyde to be a basis for holding Jekyll to be culpable. Although Hyde does not have an excuse, Jekyll does. In the same way, though perhaps to a lesser extent, Malignant and Soldier as they are before and after the interference have an excuse, despite the fact that what their bodies do while under the influence of the interference is culpable.

The obvious disanalogy between these cases is that Jekyll does not remember what Hyde does with his body. Memory is an important component of what matters for personal identity. This may be why we are inclined to use different names for Jekyll and Hyde but not for Malignant and Soldier while they are interfered with. Indeed, Malignant and Soldier may share all sorts of mental

states with their interfered with selves, while Jekyll and Hyde do not. This difference explains why the Jekyll and Hyde case is so much clearer, but I am not sure how much it matters for attributing culpability.

We should consider not just how similar different versions of Malignant and Soldier are, but how similar they are in terms of the things that matter for attributing culpability. It is not unreasonable to suppose that parts of a person's personality that matter for attributing culpability are those parts, such as sensitivity to action-guiding moral reasons, that are assessed by judgments of culpability. There may be many respects in which Malignant is the same while under the influence of the tumor. The one respect in which he is clearly not the same is that part of his personality that makes him culpable. That part has been radically altered by the tumor and is no longer a part or product Malignant's personality. It is for this reason, I believe, that we are disinclined to hold Malignant to be culpable while he is under the influence of the tumor.

There are three respects in which Malignant differs from his brother Malicious, and which make it reasonable to treat Malignant, but not Malicious, in the same way that we treat Jekyll. First, Malignant, like Jekyll but unlike Malicious, has suffered a radical change in his moral character. This change is sufficiently great to mark a clear break between Malignant's character before and after the change. This is important because it makes it possible and worthwhile to note the distinction.

Second, for Malignant, as with Jekyll, this change has been brought about by a source that is clearly distinct from the source of the rest of his personality. Malignant's normal personality evolved gradually from the interaction of his genes, his environment, and his experiences. The tumor brings about a change of a different sort. This is even more obvious in the case of Soldier. The medical procedure performed by the army is clearly different from the way the rest of Soldier's personality evolved. This condition is important because it makes it possible to isolate a part of what happens in the person's brain from the rest of the person's personality.

Third, for Malignant, as with Jekyll, the change that has been brought about is reversible. If the tumor is removed, or the procedure is reversed, the person will go back to more or less the way they were before. At least, the change that the intervention brought about will be undone. In the same way, when the potion wears off Hyde will be replaced by Jekyll. This condition is crucially important, because it provides the main grounds for supposing that these people should not be treated as culpable. It is only because the interventions in these cases are reversible that we must be careful how we treat the people while they are under the influence of the intervention.

We can see this by supposing that the potion Jekyll takes has a permanent effect. In that case we cannot say that Jekyll's body has been temporarily taken over by Hyde. Instead, it is more accurate to say that Jekyll has died and been replaced by Hyde. I have assumed that Hyde is not a wild animal, or even a psychopath, but simply an ordinarily immoral person. The sort of person whose lack of sensitivity to morally significant considerations leads him form culpable intentions. Thus, once Jekyll has been removed there is no reason not to treat Hyde as a culpable. Hyde might complain that he did not choose to be predisposed towards culpability, but the same can arguably be said of any ordinarily immoral person such as Malicious.

It is only while the return of Jekyll is possible that we must be careful in the way that we treat Hyde. This is because Jekyll is the true owner of the body Hyde is using, and whatever we do to Hyde we also do to Jekyll. We can imagine that Jekyll is a sort of human shield that Hyde carries about with him wherever he goes. The same is true, though less obviously so, in the case of Malignant. If the change the tumor has brought about is permanent, then we have lost our reason for not treating Malignant as culpable. It is true that this would make Malignant very unlucky, but he would be unlucky in the way that his brother Malicious is unlucky. They would both suffer from having dispositions that predispose them towards culpability. This is bad luck, but it is not the sort of bad luck that can excuse, without dismantling the whole edifice of moral responsibility.

5. Moral Luck and the Uses of Culpability

This last point brings us back to where we started this chapter, with Nagel's account of moral luck. The problem of moral luck arises when we consider the many ways in which people are not in control of whether they become culpable. Even if, in a narrow sense, we control our intentions, we do not control the situations in which we find ourselves, or the character with which we are born, or the circumstances in which we receive our moral education. The everyday requirement of control seems to undermine the very possibility of culpability. We are left with the skeptical worry that none of our culpability judgments can ever be fair or justified.

A classic solution to this problem is to redefine the requirement of control. We could argue that the control necessary for culpability is not a complete freedom from the influence of character and circumstance. Rather, it is just control of the kind necessary for something to count as an exercise of agency, to be voluntary rather than involuntary. There is some truth in this classic solution. After all, we should not give too much weight to the simple intuition that control is necessary for culpability until we have clarified the kind of control that we have in mind. Nevertheless, there remains something cheap about this solution. Redefining control does not rid us of the intuitive strangeness and unfairness of holding people to be culpable for what they control while acknowledging that people lack control over the conditions that lead them to become culpable.

Nagel rejects this classic solution and argues that the problem of moral luck is unsolvable. The problem of moral luck does not arise because of a misunderstanding of the control condition. It is simply a paradox that arises when we investigate this condition. He makes this point by drawing an analogy between what he sees as two intractable philosophical problems: the problem of moral luck in morality and the problem of radical skepticism in epistemology. He writes,

Our beliefs are always, ultimately, due to factors outside our control, and the impossibility of encompassing these factors without being at the mercy of others leads us to doubt whether we know anything ... Moral luck is like this because while there are various respects in which the natural objects of moral assessment are out of our control or influenced by what is out of our control, we cannot reflect on these factors without losing our grip on the judgments.⁹⁸

The epistemological problem is that all our efforts to verify our beliefs are subject to the same doubts that lead us to try to verify our beliefs. There is no external standard we can appeal to for justification. Moral luck presents a similar problem because there is no part of a person's moral life that is not subject to some kind of luck. We can restrict our judgments to actions, or intentions, but we cannot find a way to assess a person without introducing some element of luck.

Yet there is also a substantial disanalogy. This is that in the case of culpability, unlike the case of knowledge, there is an external standard we can appeal to for justification. While the standards for knowledge are the ultimate epistemological standards, the standards for culpability are not the ultimate moral standards. Culpability is simply a useful part of morality, but it is not all there is. Our judgments of culpability can be morally justified by principles that are external to culpability. These are principles about goodness or rightness, and about what we ought to do.

This is another reason why a teleological approach to culpability is valuable. By focusing on what culpability is for, we have new resources for showing why our judgments and practices of culpability are justified. We can concede that culpability judgments cannot be entirely isolated from luck. However, this does not mean that these judgments cannot be justified. We are justified in making culpability judgments because these judgments are useful and important elsewhere in morality.

The kind of culpability I have been describing is useful because it describes the kinds of choices that justify liability and separates the permissible from the impermissible. Culpability may

⁹⁸ Nagel, "Moral Luck," 140.

also be essential to the way we understand our identity and agency. These uses for culpability give us reason to suppose that our culpability judgments are morally justified. They can also provide us with principled reasons for distinguishing the kinds of control that are essential for culpability from the kinds of control that are not.

We can make the same point about the fairness of culpability judgments. Since these judgments cannot be completely isolated from luck, there is a *prima facie* sense in which culpability is unfair. Yet, culpability is also very useful and valuable. We can assume that some system, such as culpability, is fair if reasonable people would agree to that system if they did not know whether they would be among the best off or worst off under that system. We have seen a few ways in which the concept of culpability is very valuable both within morality and for our understanding of ourselves as people. We therefore have reason to believe that we would want to be treated as subject to culpability judgments, even if we did not know in advance whether we would be morally lucky or unlucky. It is for this reason that some judgments of culpability can be fair, despite first appearances.

These arguments ought to be explored further, but I cannot do so here. My point is simply to show an additional advantage of my teleological approach to understanding culpability. Perhaps there are some kinds of culpability judgment that cannot be fair or justified, given the problem of moral luck, because they are not useful. This may be the case for those broad and vague judgments of ultimate moral worth that are sometimes thought to be the core of culpability. Yet the kind of culpability that justifies liability can be both fair and justified because it is a necessary part of a very important set of moral practices. The usefulness of culpability is the key to its justification.

Another way to make this point is to compare it with the classic solution to the problem of moral luck. The classic solution asks us to accept a narrower and less demanding understanding of control. I am proposing that we also accept a narrower and less demanding understanding of culpability. The kind of culpability I have in mind is simply a useful moral concept that plays a key

role in our broader moral theories. It describes a certain kind of moral fault that is relevant for judgments of liability and permissibility. It is certainly less grand than that ultimate Godlike moral assessment that Nagel seem to have in mind. Given the problem of moral luck, perhaps this reduction in grandeur is justified. We can save the part of culpability that we need, by jettisoning the part that is unworkable.

6. Conclusion

I began this chapter with the suggestion that there are many kinds or parts of culpability. I am interested in the kind of culpability that justifies liability and plays a central role in our judgments of proportionality and necessity. Just as an account of liability requires an account of culpability, a full account of culpability may require an account of liability. An account of liability can help us to identify the important kind of culpability that matters for liability. In this chapter described this kind of culpability in terms of moral reasons. I concluded that we are culpable in the way that matters for liability when we manifest an insufficient sensitivity to action-guiding moral reasons in the exercise of our agency.

In the second half of this chapter I discussed some theoretical questions that confront a theory of culpability. I considered the relationship between culpability and permissibility and the possibility of excuses that do not deny culpability. In the next chapter I turn to more practical questions and discuss more precisely what it means to manifest insufficient sensitivity to action-guiding moral reasons.

Chapter Six:

Culpability in Action

Introduction

This chapter continues the discussion of culpability by adding more detail to the basic view sketched in Chapter Five. In the first section, I give some more substance to my suggestion that an intention can be a basis for culpability if it manifests an insufficient sensitivity to action-guiding moral reasons. I argue that this depends both on the content of the intention and on how the intention comes about. This discussion provides an opportunity to discuss the place of my proposed approach to culpability within the current debate between volitionist and attributivist approaches to culpability. In the second section, I show how my account can describe the culpability of people who have highly culpable intentions but have yet to commit highly culpable actions. I conclude, in the final section, by discussing a series of hard cases for culpability posed by morally ignorant people.

1. A Rough Guide to Measuring Sufficient Sensitivity

A person is culpable, in the sense relevant to liability, to the extent that she manifests an insufficient sensitivity to action-guiding moral reasons in the exercise of her agency. How are we to measure this insufficient sensitivity? To know what is sufficient and what is insufficient, we will ultimately have to rely on our intuitions about what counts as doing and caring enough for others. Yet, we can still wonder how sensitivity should be understood, and how we can go about making these judgments. These are difficult questions. In this section I offer a few considerations for how these judgments should be made.

To measure the sensitivity to action-guiding moral reasons manifested in a given intention, or other such exercise of agency, we need to look at two things. First, we need to look at the content of the intention, what the person intends to do. This content is evaluated for its fit with action-guiding moral reasons given the person's beliefs and available evidence. To be culpable, the intended action must be wrongful from either the belief or evidence-relative perspective. However, we also need to look at how that intention came about. To be culpable, an intention must also be the result of insufficient moral motivation on the part of the person. In this section, I will elaborate briefly on this two-part test, and discuss why each of the two parts is necessary.

1.1 Sensitivity in the Content of the Intention

The simplest way to evaluate the moral quality of an intention, meaning the sensitivity it manifests to action-guiding moral reasons, is to look at the content of the intention and the beliefs of the person. According to this belief-relative standard, the sensitivity to moral reasons manifested by an intention depends on how wrongful the intended action would be, if the world was as the person believed it to be. As I argued in Chapter Five, this belief-relative standard can only be a part of our evaluation. We also need to consider the available evidence.

Cases where people form their intentions while ignorant of pertinent moral or non-moral facts are often divided into tracing cases and non-tracing cases.⁹⁹ In tracing cases, the person's ignorance is the result of some previous culpable intention, such as deliberately not reading a safety manual. In non-tracing cases, by contrast, the person's ignorance cannot be traced to some prior culpable intention or omission on the part of the person. In non-tracing cases a person is typically ignorant of some pertinent fact about their situation either because there was no way they could

⁹⁹ For this distinction see Holly Smith, "Culpable Ignorance," *The Philosophical Review* (1983) 543-571; and "Non-Tracing Cases of Culpable Ignorance," *Criminal Law and Philosophy* (2011) 115-146.

reasonably be expected to know it, or because of a failure to remember or realize it on the part of the person.

Those in this latter type of non-tracing case often appear culpable, and potentially liable, yet from the belief-relative perspective they are blameless. For example, consider,

Ignorance: Worker is in charge of monitoring pressure in large vat. She notices that the pressure has gotten too high, and so presses a button to release it. Worker forgets, or fails to realize, that doing so will seriously injure some other workers who are repairing the discharge pipe.

From the belief-relative perspective, Worker's intention – pressing the button to clear the vat – is not at all culpable. This is because Worker believes, due to her failure to realize or remember, that releasing pressure from the vat will not put anyone in danger. Nevertheless, in at least some non-tracing cases we want to say that people such as Worker are negligent and culpable for their decisions.

Suppose that Worker has been repeatedly reminded about the dangers of venting the vat but simply does not care very much about safety. If any normally moral and competent person would have realized that discharging the pressure was highly dangerous, then there is clearly something subjectively wrongful, and potentially liable, about Worker's intention to do so.¹⁰⁰

Thus, Worker's culpability must also depend on whether the false belief she formed is reasonable based on the available evidence. In these cases, we should also adopt an evidence-relative standard of subjective wrongfulness. I will follow Parfit in considering an intention to be wrongful in the evidence-relative sense to the extent that the intended action would be objectively wrongful if the person "believed what the available evidence gives [her] decisive reasons to believe, and these beliefs were true."¹⁰¹

¹⁰⁰ Not everyone shares this view. For an argument against the culpability of negligent people see Larry Alexander and Kimberly Ferzan, *Crime and Culpability* (Cambridge UP, 2009) 70.

¹⁰¹ Parfit, *On What Matters: Vol. 1*, 152. What evidence counts as 'available' is a persistent problem I will not attempt to solve.

Both the belief-relative and evidence-relative standard of subjective wrongfulness can be relevant for evaluating the subjective wrongfulness of an intention. A person who genuinely believes his intended actions to be harmful, even if that belief is unsupported by the evidence, may still be considered culpable. At the same time, people like Worker who have enough evidence to realize that their intended action is harmful, should may sometimes be culpable.

Other things being equal, the most culpable people will be those who satisfy both conditions. An intention manifests the greatest disregard for morally significant considerations when the intended action is wrongful from both the belief-relative and the evidence-relative standard and manifests this disregard to a lesser extent if it is wrongful according to only one of these standards. However, to get a full picture of a person's culpability we need to look beyond the content of her intentions. We also need to investigate how those intentions come about.

1.2 Sensitivity in the Formation of the Intention

There are two ways in which a person's internal, or subjective, response to action-guiding moral reasons can be sufficient or insufficient. As we have seen, a person's internal response to action-guiding moral reasons can be measured by the moral quality of what the person intends to do, given her beliefs and the available evidence. This standard alone will not always separate the culpable from the non-culpable. We must also consider the role that action-guiding moral reasons play in forming the person's intention.¹⁰² This is particularly clear in the case of negligent people. To see this, consider,

Ignorance (Careful Variation): Two workers, Careless and Careful, are each responsible for realizing the pressure in separate vats. Unfortunately, because of the layout of the control rooms, determining whether it is safe to discharge the vats is not always easy. It often requires making a difficult judgment on the basis of readings on several different dials. Careless and Careful are presented with the

¹⁰² Holly Smith has defended a view of blameworthiness that depends on the extent to which intentions result from immoral evaluative attitudes. See Smith, "Non-Tracing Cases."

same training and the same evidence showing discharging the vats is unsafe. Both fail to realize this, vent their vats, and injure their fellow workers.

According to a belief or evidence-relative standard both these people should be equally culpable.

However, this will not always be the case.

Suppose that Careless would have realized that venting was dangerous, and not vented his vat, if he cared more about the safety of his fellow workers. However, because Careless does not care much about safety, he did not notice when the dials were showing that venting was dangerous. Careful, on the other hand, cares greatly about safety but is simply confused by the multiple dial system. Despite trying hard to be diligent about venting the vat, Careful fails to realize that venting is dangerous because he misreads the message the multiple dials are giving him. In this case, it seems to me, Careless is clearly somewhat culpable while Careful is not. Even though they were both presented with the same evidence, and may have had the same beliefs, Careless's omission manifested a lack of sensitivity to action-guiding moral reasons while Careful's did not.¹⁰³

We can now see why, as I suggested in the last chapter, negligent people are an exception to the otherwise close connection between culpability and subjective permissibility. It is true that an intended action must be subjectively impermissible in order for that intention to be culpable. In this way there remains a close connection between culpability and permissibility. However, in cases like *Ignorance* subjective impermissibility is not sufficient for culpability. These cases show that to evaluate a person's sensitivity to moral reasons we need to look beyond the content of their intentions.

I propose that an intention, or other such exercise of agency, is insufficiently sensitive to action-guiding moral reasons to the extent that two conditions are met. First, the intended action is wrongful according to either the belief-relative, or evidence-relative standard. Second, the intention results from an insufficient moral motivation, on the part of the person. Moral motivation is simply

¹⁰³ For an excellent account of moral concern see Nomy Arpaly, *Unprincipled Virtue* (OUP, 2003) 84.

the tendency or disposition to be motivated, in the right way, by those things in the situation that are morally significant. Things that are morally significant are those things that give us action-guiding moral reasons.

In evaluating a person's moral motivation, we might ask how much the person cares about those things that are morally significant. As Nomy Arpaly points out, being motivated by what is morally significant, "is not a matter of *de-dicto* caring about morality, but *de-re* caring about what is in fact morally significant."¹⁰⁴ We should consider how sensitive the person's decision making was to the things in the situation that actually give us moral reasons, like the fact that some action would cause pain or that another action could prevent it. We cannot evaluate culpability without making some judgment of the role that sensitivity to morally significant considerations played in the formation of an intention.

Measuring how much a person is motivated by something requires us to look at that person's desires and moral beliefs.¹⁰⁵ This is one of the ways in which beliefs and desires, which are not themselves the object of culpability judgments, can form part of the background conditions for assessing culpability. We must look at the whole of the person's motivations, and the whole of the person's circumstances, and ask whether the person was sufficiently motivated by the morally significant considerations. This is another way of asking whether the person cared enough. These judgments will not be easy, and they are worthy of further study in their own right. All I can do is offer a few observations before moving on to consider how judgments of caring enough about morally significant considerations fit into my account of culpability.

My first observation is that caring enough is a matter of being motivated in the right direction. Consider a sadistic person who acts out of a desire to cause pain. This person might care a

¹⁰⁴ This quote comes from Elizabeth Harman, "Does Moral Ignorance Exculpate?" *Ratio* (2011) 460. Harman is describing the views of Arpaly in *Unprincipled Virtue*.

¹⁰⁵ Particularly if we assume weak motivational internalism.

great deal about pain, but they do not care about it in the right direction. People who have immoral desires, who desire things that are morally undesirable, do not count as caring enough about these morally significant considerations.

My second observation is that judgments of caring enough must be comparative. To say that a person does not care enough about something is to say that they do not care enough about it relative to the strength of their other motivations. Plausibly, one way in which Careless might care too little about the safety of his fellow workers is by caring too much about a game of online Poker. A person can be insufficiently motivated if they do not care enough about what is morally significant, if they care too much about things that are not morally significant, or if they have immoral desires and care about things in the wrong way.

This comparative element of the evaluation of motivations raises a host of difficult questions. For example, if Careless's concern for online Poker is overpowering, if it is something he prioritizes even over his own safety, then it is plausible that Careless does care enough about the safety of his fellow workers. He might care enough about the safety of his fellow workers when compared to his concern for his own safety, yet not care enough about the safety of his fellow workers when compared to his concern for online Poker. It is difficult to know how we should judge Careless in this case.

One possibility is to borrow a technique that economists use when comparing the purchasing power of different currencies. This is to use a common basket of goods and ask how much this basket of goods would cost in each currency. To know whether a person cares enough about something, we might make a comparison with a core basket of concerns: things like concern for his own safety or his own future wellbeing. There is presumably an amount of concern for his fellow workers that Careless ought to have compared to this core basket of concerns. If he cares enough by this standard then I am tempted to suppose that he is excused, even if he also cares far

too much about online Poker. This is only a suggestion; for the rest of this thesis I will simply rely on intuitive judgments about what counts as caring enough about morally significant considerations.

If a person forms an intention that is wrongful from the evidence or belief-relative perspective, we should then ask, would the person still have formed this intention in the right way if they were motivated enough by morally significant considerations. This question can help us to separate the culpable Careless from the innocent Careful. Both people behaved wrongfully from the evidence-relative perspective. However, Careful did so despite caring enough about the safety of his fellow workers. By contrast, Careless did not care enough about safety. We condemn Careless because we suppose that if he had cared enough about safety, he would have realized that venting the vat was dangerous and would not have done so. We can never know such a thing for sure. Yet, I think a judgment of this kind lies behind our assessments of culpability.

It can explain why we care about the non-moral capacities of people, when assessing culpability. It is relevant that Careless was better at reading the dials in the control room than Careful was. This matters because it suggests that Careless would have noticed the danger he cared enough about morally significant considerations. It also matters because it can explain how Careless could fail to notice the buildup of pressure despite caring enough.

Perhaps despite initial appearances, this point about culpability is analogous to the point that Kant makes about virtue in the *Groundwork*.¹⁰⁶ Kant argues that virtue is not simply a matter of the content of one's intentions, of intending what morality requires. A virtuous action must also be done for the right reasons. It must be done in response to those aspects of the situation that really are morally significant and give us moral reasons. A person's motivation can have a contingent overlap with morality, if the person is lucky enough to be in a situation where what morality requires is

¹⁰⁶ Immanuel Kant, *Groundwork of the Metaphysics of Morals*, trans. by Mary Gregor (Cambridge UP, 2012) 59.

something that the person is motivated to do for non-moral reasons. In such a situation a person might do the right thing without being virtuous at all. Thus, Kant proposes a counterfactual test. The truly virtuous person is the one who would still be motivated to do what morality required, even if he did not have any non-moral reason to do it.

This is analogous to my understanding of culpability. To be culpable a person must be insufficiently motivated by those things that actually are morally significant. Yet sometimes a person can simply be unlucky and find themselves in a situation where their intentions have a contingent overlap with what is morally forbidden. This is true of people like Careful, who do what is evidence-relative impermissible, but do not do it because of an insufficient moral motivation. To capture these cases, we need a rough counterfactual test. A truly culpable person is one who would not make the same mistake, if they cared enough about those things that actually are morally significant. In both these cases the point of the counterfactual is to test the extent to which the person's action resulted from a genuine sensitivity, or insensitivity, to those things that are morally significant, rather than some contingent facts about the situation that make it easier or harder to do what is morally required.

It is a mistake, or at least uncharitable, to suppose that Kant considers desires to be irrelevant for assessing virtue. Kant is classically understood to suppose that willpower, or strength of will, is what matters for virtue. Of course, this could not be sheer willpower, in the pursuit of any goal. Rather, it is strength of will in doing what morality requires. Virtue, for Kant, is a counterfactually robust disposition to be motivated to do those actions that are required because of action-guiding moral reasons, rather than a contingent inclination to do what is required when doing so is pleasurable or convenient.

Desires are simply one of the ways in which we describe what people are disposed to be motivated by. Our desires can lead us to act morally if we desire things that happen to overlap with

what is morally required, or they can lead us to act morally when we desire those things that really are morally significant. Thus, Kant's point is not that desires do not matter for virtue, rather it is that we must separate the genuinely virtuous desires from the contingently virtuous ones.¹⁰⁷ For Kant, a person who keeps his promises because he has a strong desire to act rationally in accordance with the categorical imperative is truly virtuous, while a person who keeps his promises because he has a strong desire to be popular is not. Where Kant's view may differ from ours is not over the significance of desires, so much as it is over the sorts of considerations that give us moral reasons. However, at this point we have moved from discussing the necessary conditions for virtue and culpability, to a discussion of the justifications for morality itself.

To summarize, judging how sensitive a person was to action-guiding moral reasons in the exercise of her agency requires a two-part test. First, we look at the content of the intention. We then ask, how wrongful was what the person intended to do from the belief and evidence-relative perspectives. Second, we look at the process of intention formation. We then ask whether we suspect that the person would still have formed this wrongful intention if she was sufficiently motivated by, or cared enough about, the relevant morally significant considerations. Many, if not most, excuses will be relevant for both of these tests.

For this reason, it might be objected that having both tests is unnecessary. The case of Careful and Careless shows why we need to look beyond the content of intentions, but why do we need to look at the content of the intention at all? The reason for this is that not all intentions formed out of an insufficient sensitivity to morally significant considerations will be equally culpable. Some might not even be culpable at all.

Consider the case of a student who decides to throw an eraser at his classmate. Suppose that the student is filled with a jealous hatred for this classmate and has absolutely no regard for the

¹⁰⁷ This point has been made elsewhere.

welfare of this classmate or the moral importance of not harming her. This student's action resulted from a serious disregard for morally significant considerations.

This, however, only tells part of the story about the moral quality of the student's intention. While the student's intention did result from a serious disregard for morally significant considerations, the intention itself is not especially culpable because throwing an eraser is not an especially objectively wrongful act. If the student threw a pair of scissors, instead of the eraser, this would be a more culpable intention even if this action resulted from the exact same disregard for morally significant considerations that led the student to throw the eraser. Both parts of my test are necessary in order to capture this complexity.

Of course, neither of these tests can be deployed with any precision. My purpose here is simply to give some further structure and substance to my suggestion that the culpability that justifies liability is a measure of a person's sensitivity to action-guiding moral reasons in the exercise of her agency. The tests I have suggested give us a way of understanding what that sensitivity amounts to, and a very rough idea of how it can be measured. Ultimately both the tests rely on intuitive judgments of what counts as doing and caring enough for others. Moreover, both depend on hard to determine facts about the person's psychology and the content of her intentions. Both these problems are inescapable for a subjectivist theory of culpability. Much more work must be done before we can approach a fuller understanding of culpability.

1.3 Between Choice and Character

The two-part test I have proposed sits in between the volitionist and attributivist approaches which currently dominate discussions of culpability and blameworthiness.¹⁰⁸ Like any such labels,

¹⁰⁸ For a good overview of this distinction, and a discussion of whether these positions necessarily conflict see Elinor Mason, "Moral Ignorance and Blameworthiness," *Philosophical Studies* (2015) 3037-3057; and Neil Levy "The Good, the Bad, and the Blameworthy," *Journal of Ethics and Social Philosophy* (2005) 1-16.

attributivism and volitionism can only be described in general terms and with due recognition of the fact that no two views in either camp are the same. While these theories attempt to describe blameworthiness in the abstract, rather than the specific kind of culpability that matters for liability, it is worth considering how the account I have proposed fits into this larger debate.

Attributivists approach culpability as primarily a matter of describing the quality of a person's moral character. Actions and attempts are culpable, on this approach, because of what they reveal about the moral character of the person.¹⁰⁹ A culpable intention is one which reveals immoral character traits on the part of the person. Given that what matters, on this view, is the underlying moral character of the person, attributivists do not limit culpability to voluntary actions. People can be culpable, on this view, for their non-voluntary responses as long as those responses are caused by immoral character traits on the part of the person. In particular, people can be culpable for failures to realize or remember something if this failure can be traced to an immoral character trait or disposition.

Volitionists, by contrast, reject the attributivists' focus on character traits and non-voluntary reactions.¹¹⁰ What matters for culpability, they argue, is not the overall moral character of the person, but rather the moral quality of the person's choices. A person is most clearly culpable, on this view, when that person chooses to do something she knows to be wrong. For volitionists, people can only be held culpable for things that are within their control. Thus, whatever a non-voluntary reaction might reveal about the moral quality of a person's character, such reactions are not appropriate

¹⁰⁹ The name 'attributivism' comes from Garry Watson, "Two Faces of Responsibility," *Philosophical Topics* (1996) 227–248. My understanding of attributivism is primarily based on Holly Smith, "Non-Tracing Cases." Other recent examples include Harman, "Does Moral Ignorance Exculpate?" and Angela Smith, "Responsibility for Attitudes: Activity and Passivity in Mental Life" *Ethics* (2005) 236-271.

¹¹⁰ Clear recent examples include Michael Zimmerman, *Living with Uncertainty* (Cambridge UP, 2008) and "Moral Responsibility and Ignorance," *Ethics* (1997), 410-426. Also, Douglas Husak, *Ignorance of Law* (OUP, 2016); Neil Levy, "The Good the Bad and the Blameworthy;" and in a more skeptical vein Gideon Rosen, "Skepticism about Moral Responsibility," *Philosophical Perspectives* (2004) 295-313 and "Culpability and Ignorance," *Proceedings of the Aristotelian Society* (2003) 61-84.

objects of culpability because they are not something the person chooses or controls. In the same way, volitionists often claim that failures to realize and remember can only be culpable if they were themselves caused by some prior culpable choice on the part of the person: such as a failure to inquire further or take precautions against forgetting.¹¹¹

According to the account of culpability I have sketched, there is an element of truth to each view. To see how this works, and in order to better understand the dispute between attributivists and volitionists, it is useful to distinguish two different questions over which attributivists and volitionists disagree. The first question asks what people can be culpable for. Volitionists maintain that people can only be culpable for those things that I have called exercises of agency, such as actions or intentions. As I argued in the last chapter, for the kind of culpability that justifies liability this volitionists answer is correct. It is our culpability for our choices that can make us liable to defensive harming.

The second question is how the thing we are culpable for should be assessed. On this front, I have argued, we should accept the broader attributivist position. We should consider not just the person's beliefs and the content of her intentions, but also all the other facts about her circumstances. We must also look at her available evidence, as well as her decision making process, her desires, and how much she cares about the relevant morally significant considerations. Only by taking these background factors into consideration can we get a full picture of the moral quality of a given choice.

Some volitionists might object that my appeal to the evidence-relative standard is unwarranted, since culpability is meant to be an internal or subjective appraisal of the moral quality of the person's choices.¹¹² This internal perspective, they might argue, is exactly what distinguishes

¹¹¹ For a defense of this claim see Rosen, "Skepticism about Moral Responsibility," 301.

¹¹² This objection is suggested by remarks by Husak in *Ignorance of Law*, 129.

culpability judgments from external assessments of the evidence-relative or fact-relative wrongfulness of an intention.¹¹³ While I recognize the appeal of this line of reasoning, I believe that it misunderstands the nature and purpose of culpability judgments.

When we assess the subjective wrongfulness of a choice or intention, our fundamental interest is in what it tells us about the culpability of the person who formed that intention. We do not particularly care about the moral quality of the intention taken in isolation. Rather, we care about the moral quality of the intention as a response to the world on the part of the person. To make this judgment we cannot simply ignore the person's circumstances. We need to know not only what the person was thinking as she formed the intention, but also what the person could have been, or ought to have been, considering based on the available evidence.

1.4 The Culpable Ignorance Premise

It might be objected that my proposed compromise between volitionism and attributivism should be impossible. The reason for this, it could be argued, is that because I take the attributivist line on non-tracing cases of ignorance and assert that people such as Worker and Careless can be culpable, I implicitly accept that unintentional failures to remember or realize can themselves be culpable. This, in turn, contradicts my volitionist assertion that only intentions, or the failure to form intentions, can be the proper object of culpability.

This objection relies on the premise, common to both attributivists and volitionists, that an intention formed in ignorance of a pertinent fact can only be culpable if that ignorance is itself culpable. Let us call this the culpable ignorance premise. In this section, I intend to argue that this premise should be rejected. Rejecting the culpable ignorance premise not only makes my proposed

¹¹³ The term 'fact-relative' also comes from Parfit, *On What Matters: Vol. 1*, 152.

compromise between volitionism and attributivism possible, it also, as I will argue below, helps solves persistent problems surrounding the culpability of morally ignorant people.

The culpable ignorance premise is shared by both volitionists and attributivists.¹¹⁴ Perhaps for this reason, it often goes without scrutiny and arguments in its defense are hard to find. Michael Zimmerman, in his book, *Living with Uncertainty*, defends the culpable ignorance premise by arguing that “this premise is also deeply embedded in our everyday practice of blaming people for their ignorant behavior ... we often blame people for performing actions that they *didn't* know where wrong on the grounds that they *should have* known they were wrong.”¹¹⁵ He argues that because we only hold ignorant people to be culpable when they ‘*should have* known’ otherwise, we are tacitly accepting the culpable ignorance premise. The ‘*should have*’ suggests that the person does something wrong, and thus may be culpable, simply by being ignorant.

I agree that there is a sense in which ignorant yet culpable people should have known better. However, the normativity of this ‘should have’ does not depend on action-guiding reasons. After all, beliefs are not the sorts of things to which these reasons apply. Consider the case of Careless, who is culpable for unsafely venting the vat. According my first test, a person who forms an intention that is not wrongful from the belief-relative perspective can only be culpable if that intention is wrongful from the evidence-relative perspective. Thus, only in cases where the person’s ignorance is unreasonable, given the available evidence, will the person be culpable. In this way, the normativity of the ‘*should have* known’ claim can be understood as epistemic rather than moral.

According to the second test, Careless is culpable, in part, because his failure to vent the vat resulted from the fact that he did not care very much about the safety of his fellow workers. It is natural to say that Careless should have realized that venting the vat was dangerous, because we

¹¹⁴ For example, see Harman, “Does Moral Ignorance Exculpate?” and Smith “Non-tracing Cases,” as well as Rosen “Culpability and Ignorance,” and Zimmerman, *Living with Uncertainty*.

¹¹⁵ Emphasis in the original. Zimmerman, *Living with Uncertainty*, 177-8.

assume that if he cared enough about safety he would have realized that venting the vat was dangerous. From this perspective, Careless should have realized that venting the vat was dangerous because he should have cared more about safety. Caring about safety is not simply a matter of responding to action-guiding moral reasons, it is also a matter of desiring things that are desirable. Careless should have cared more about safety, and this should have claim may be part of what we mean when we say that he should have realized that venting the vat was dangerous. This does not imply that not caring about safety, or not realizing the danger of venting the vat, is the kind of thing that can be culpable in the way that justifies liability.

Finally, it is natural to say that a person 'should have' done something when doing that thing is a part of something that should have been done. Suppose that Villain points a loaded gun at Victim and pulls the trigger. It is natural to say that Villain should not have pulled the trigger. This does not mean that pulling the trigger is itself the problem. Rather, the wrongness of pulling the trigger depends on the fact that pulling the trigger is a part of killing Villain. In the same way, it might be the case that Careless should have realized that venting was dangerous, because realizing this would have caused him to not vent the vat. This is compatible with it being the decision to vent the vat that is ultimately the source of Careless's culpability.

Overall, we ought to reject the culpable ignorance premise, at least when it comes to the kind of culpability that justifies liability. This is compatible with there being several senses in which people like Careless should have had better beliefs. Indeed, I accept that Careless cannot be culpable unless it is true in at least some sense that he should have realized that venting the vat was dangerous.

Clearly, however, none of these arguments undermine the claim that the kind of culpability that justifies liability is culpability for choices and exercises of agency. Beliefs cannot be directly culpable in this sense because they are just not the sorts of things to which action-guiding reasons

apply. Perhaps there are other types of culpability for which the culpable ignorance premise is true. This too is compatible with all I have said. Yet since I am not sure how these other sorts of culpability judgments work, I will not pursue the question further.

2. Liability for Future Wrongs

In Chapter Four, I discussed the way that culpability changes in the time after a culpable act. These changes, I argued, are very important for liability. Time is also very important for liability when it runs in the other direction. People can be liable to be harmed before they have completed any culpable actions. Recall,

Preemption. Villain culpably intends to murder Victim by pushing him off a cliff. Victim is standing on the edge of the cliff. You can save Victim but only by killing Villain as he walks over to push Victim.

This kind of preemptive liability is essential to the very possibility of defensive harming in most cases.¹¹⁶ As I have argued, this suggests that culpable intentions, without causal responsibility, can be a basis of liability. In this section I will discuss how the kind of culpability that justifies liability should be applied to these cases.

If we accept that culpable intentions are what matter, then a better understanding of intentions will help us understand liability. For reasons of space, and because I want my view to remain compatible with different theories of action and intention, I cannot delve too deeply into debates in the philosophy of action. Nevertheless, I think it is useful to make two simple observations. First, like beliefs, intentions can be either occurrent or latent. Generally, an occurrent intention or belief is one the person is currently contemplating while a latent intention or belief is

¹¹⁶ Jeff McMahan, "Preventive War and the Killing of the Innocent," *The Ethics of War* ed. by David Rodin and Richard Sorabji (Ashgate Publishing, 2005) 172.

one the person is not currently contemplating but nevertheless holds.¹¹⁷ How exactly this distinction is to be drawn is controversial, although it does not matter much for my purposes. I consider both occurrent and latent intentions to count towards culpability since both can be understood as an exercise of agency.

Second, as I mentioned in Chapter Five, my focus on intentions is a shorthand for the diverse exercises of agency – willing, trying, acting, intending – that can be the object of culpability. It includes both wide-scope intentions for the future, such as my intention to write this chapter, and the fine-grain intentions, sometimes called proximal intentions, that guide and control action: such as the intentions I form to type certain words which guide the movements of my fingers.¹¹⁸

This flexible approach to intentions is useful for two reasons. First, it accommodates the not unlikely possibility that there is no firm line to be drawn between these wide-scope and fine-grain intentions. Second, it allows us to make judgments of liability in cases like Villain's, where a culpable wide-scope intention has been made (the decision to kill Victim by pushing him off the cliff) but culpable fine-grain intentions (such as the intention that guides Villain's hand and arm to push Victim) have not yet been formed.

In general, wide-scope plans to commit wrongful actions in the distant future are thought to be less culpable than the fine-grain intentions that immediately precede and guide wrongful actions. Deciding to kill someone by shooting them the next time you see them is less culpable, it is often thought, than deciding to pull the trigger when they are in your sights. Correspondingly, a person who forms a wide-scope intention to kill is less liable, relative to the same harm prevented, than a

¹¹⁷ For a discussion of this distinction in the context of belief see Rick Peels, "What Is Ignorance?" *Philosophia* (2010) 57-67; and Husak, *Ignorance of Law*, 163.

¹¹⁸ While my distinction between wide-scope and fine-grain intentions is closely related to the distinction between proximal and distal intentions my distinction depends more on the content of the intention than on the time at which it initiates action. To avoid confusion, I will use my simple distinction in what follows. For the distal/proximal distinction see Mele, *Springs of Action*, 141-6.

person who is about to kill by pulling the trigger. I share these intuitions, although I do not think they are as easy to explain as is sometimes supposed.

The common sense explanation for the increase in liability as the action gets closer to the present depends on the harm prevented side of the proportionality relation. The reason it is clearly permissible to shoot someone a few seconds before they shoot you, on this account, is that at that point you know with near certainty that your life is in danger. Weeks earlier, while they are planning their attack, it is much less certain that a defensive action will actually prevent a substantial harm.

There is a good deal of truth in this evidence relative account. Certainly, the quality of your evidence that Victim will otherwise succeed in killing Victim is crucial to determining whether it is permissible for you to try and kill Villain preemptively. Nevertheless, as I will now argue, this account fails because it relies on epistemic failings on the part of the potential intervener.

This means that it does not truly address the question we are asking. Except where I have stated otherwise, my investigation into the permissibility of harming on the basis of liability is an investigation into the *objective* permissibility of harming on the basis of liability. I want to know whether it is objectively permissible to harm a given person from what Parfit calls the fact-relative perspective: given all the true facts of the situation.¹¹⁹ By contrast, the evidence relative explanation is best understood as an account of the *subjective* permissibility of harming someone on the basis of their liability; it tells you whether it is permissible to harm someone defensively, given the evidence available to you.

In any of the cases we have been considering, the subjective permissibility of harming the aggressor can change depending on the evidence you have at your disposal. This is not unique to cases of wrongdoing that will occur in the distant future. Suppose you see Villain go to press the button that will demolish Victim's house, but you are unsure whether Victim is home. This would

¹¹⁹ Parfit, *On What Matters: Vol. 1*, 152.

certainly change the subjective permissibility of killing Villain to save Victim. Yet what we really want to know whether it is objectively permissible to kill Villain.

After all, it may not be possible to determine the subjective permissibility of killing under uncertainty, unless we first decide the objective permissibility of killing given different possible outcomes. In order to decide whether it is subjectively permissible to kill Villain to save Victim, when you are unsure whether Victim is home, you first have to know whether it would be permissible to kill Villain if Victim certainly was home.

The evidence-relative account attempts to explain liability for future wrongdoing by appealing to the uncertainty that surrounds the future. Unfortunately, because it focuses on the epistemic limitations of the would be intervener, it only tells us about subjective, not objective, permissibility. Moreover, when these limitations are removed, the predictability view can no longer distinguish between the culpability of wide-scope and fine-grain intentions. If you somehow know with certainty that Villain will follow through on her wide-scope intention to kill Victim, then the common sense view cannot explain why we still consider Villain to be less culpable during the months she is planning the attack, compared to moments before she presses the button.

A more promising strategy, I propose, is to focus not on the epistemic limitations of the would-be intervener, but rather on the limitations of the culpable person herself. This change of focus raises two distinct but related reasons to explain why wide-scope intentions regarding the distant future are generally less culpable, and hence a weaker basis for judgments of liability, when compared to fine-grain intentions for the near future. The first reason concerns the firmness with which the person holds the intention. The second reason concerns the extent to which wide-scope intentions are often conditional on future events. I will discuss these reasons one at a time.

The first reason why wrongful plans for the distant future are generally less culpable than wrongful intentions for the near future depends on the plausible assumption that people can be

more or less committed to their plans for the future. I will describe this level of commitment as the firmness with which a wide-scope intention is held. Firmly held wide-scope intentions are those that the person considers to be a settled plan of action and is highly committed to seeing through. Less firmly held wide-scope intentions, by contrast, are those that the person has given less thought to and is only partially committed to. The person may proceed on the basis of these intentions at the present, while consciously reserving or considering the possibility of changing her mind later on.

The second reason I propose to explain the increase in culpability as the time of action approaches depends on the epistemic limitations of the person in question. To see how this works, it helps to take a step back, and first consider another difficulty raised by the culpability of wide-scope intentions. Many of our plans for the future, be they nefarious or perfectly innocent, depend on eventualities about which we are uncertain. Let us call such plans conditional intentions. For example, consider Candidate, who decides that if she is ever runs for office she will stuff the ballot box to make sure she is elected. What are we to make of Candidate's culpability for forming such an intention?

One thing this might depend on, in my view, is how likely Candidate believes it to be that she will be chosen by her party.¹²⁰ If Candidate is the clear frontrunner and reasonably believes that she will be chosen by the party in a matter of days, then Candidate's conditional intention to stuff the ballot box appears to be a fairly culpable intention to commit a serious fraud. By contrast, if Candidate is just a young politician plotting what she would do if she ever gets nominated, then her conditional intention appears substantially less culpable. Of course, if Candidate later becomes the nominee, and follows through on her plans, then her intentions to stuff the ballot box at that later

¹²⁰ This possibility is raised and rejected by Alexander and Ferzan in *Crime and Culpability* as part of an objection to making intentions the object of criminal culpability. The point of their argument is to show that an appeal to intentions would be too impractical for the criminal law. Since my interest is in a more theoretical question, I find the conditionality of intentions to be a useful and not cumbersome feature. My thoughts on this developed independently, out of my investigation of liability to defensive harming. See *Crime and Culpability*, 204-206.

time will be just as culpable in either case. Nevertheless, at the present, while Candidate is still planning her action, the chance Candidate believes she has of becoming the nominee makes a difference to the culpability of her intention. This holds true, I contend, even if Candidate is equally firm in her intention in each case.

To see why this is the case, recall the claim I defended above that a person's beliefs evidence about the likely outcome of her intention matter for determining how subjectively wrongful, and hence culpable, that intention is. I pointed out that, other things being equal, the decision to throw an eraser at a fellow student is much less culpable than the decision to throw scissors at a fellow student, even if both of these intentions were motivated by an equally immoral disregard for the welfare of other students. Culpability, in my view, should be based in part on the amount of harm the person believes, or should reasonably believe, that their intention will cause.

If this point is agreed, then we need a way to determine culpability in cases where people are uncertain what the result of their intention will be. This is a complex question which raises difficult questions about the comparative culpability of risk imposition and partial ignorance. For now, however, it is sufficient to point out that, other things being equal, people who choose to impose a small risk of harm on others are much less culpable than people who choose to impose a large risk of harm.

This has implications for the culpability of conditional intentions, because the formation of a wrongful conditional intention is akin to a decision to impose a risk on others.¹²¹ The more likely the person believes the condition is to eventuate, the greater the risk the intention imposes. To see this consider,

Sneak and Kill: Two rival villains, Sneak and Kill, plan to rob jewelry stores in the near future. Both intend to kill any guards that get in their way. However, Sneak

¹²¹ Again, Alexander and Ferzan briefly consider this possibility as part of their rejection of intentions as the object of criminal culpability in *Crime and Culpability*, 204-206.

plans to rob a very poorly guarded store while Kill plans to rob one that is heavily guarded.

While both Sneak and Kill form the same conditional intention to kill guards, Kill's conditional intention is much more culpable.

The reason for this is that the condition in Kill's conditional intention is much more likely to come about and thus imposes a much greater risk of harm on others. Kill's conditional intention is tantamount to the decision to kill a guard while Sneak's conditional intention is more like the decision to impose a small risk of death on a guard. Of course, if either Sneak or Kill is actually confronted with a guard, then the fine-grain intentions they form to shoot that guard would be equally culpable. However, when it comes to the culpability of their wide-scope conditional intentions in the present, what they believe, or ought to believe, about the likelihood of the condition eventuating matters for the culpability of their intentions.

The same is true for the case of Candidate. If Candidate believes it to be very likely that she will be chosen by her party, then deciding to stuff the ballot box if she is elected is tantamount to deciding to stuff the ballot box. At the time she forms this intention, Candidate is accepting a serious chance that she will defraud others. By contrast, if Candidate is just a young politician who resolves to stuff the ballot box if she is ever nominated, then at the time she forms the intention she is merely accepting a small chance that she will defraud others. This is why, in my view, unlikely conditional wrongful intentions should be considered less culpable than likely conditional wrongful intentions.

This can also help explain why wide-scope long term plans for the future are generally much less culpable than intentions for the near future. Plans for the future are always to some extent conditional, whether this is explicitly considered by the person or not. Of course, I might perform some action now that ensures a certain result will occur in the future. This however, is not the kind

of plan for the future I am considering here. Rather, I am considering the culpability, and hence potential liability, of people who presently intend to perform some wrongful action at some point in the future.

As a general rule, such plans will tend to be less certain, and more conditional, the further they run into the future. If I plan to murder someone next year, many things could occur between now and the time of action that would prevent me from carrying out my intention. All I have so far accepted is that I will murder next year on the condition that I still believe it to be a good idea. Any number of things could go wrong, and I retain the power to change my mind regardless of how firmly I hold my intention at the time.¹²²

These uncertainties will tend to decrease, however, as the moment of action approaches. The closer my plan comes to fruition without me backing out or something going wrong, the more likely I am entitled to believe it is that my wide-scope intention to murder will actually lead to a completed murder. This, in turn, makes that wide-scope intention increasingly culpable. When, at the time of action, I form the fine-grain intention to pull the trigger, I believe with near certainty that this intention will kill another human being. For this reason, the fine-grain intention I form at the time of action is much more culpable, and hence makes me much more liable, than the wide-scope intention made me when I formed it a year ago.

In this way, my proposal incorporates much that is correct about the common sense evidence-relative account for the increase in culpability and liability as the time of action approaches. However, because my proposal does not depend on the epistemic limitations of the would-be intervener, it describes the objective liability of wide-scope intentions. Even if you somehow knew with certainty that Villain would follow through on her plan to murder in a year's time, this does not

¹²² Alexander and Ferzan make these points about intentions in order to argue that intentions alone should not be criminally culpable in *Crime and Culpability*, 204-206.

alter the fact that Villain does not yet know this with certainty. This uncertainty is what makes wide-scope intentions for the distant future generally less culpable than fine-grain intentions and wide-scope intentions for the near future.

3. The Problems of Moral Ignorance

Cases of moral ignorance raise a host of intricate problems for theories of culpability and liability. In most of these cases it is not the ignorance itself that is problematic, but rather some feature of the person's circumstances that leads to or results from the ignorance. The range of questions that cluster under the heading of moral ignorance is thus much larger than it might appear. While I will propose answers to some of these questions, for others I can only point to the direction in which I believe the answer might lie.

We can begin with a simple case of moral ignorance:

Moral Ignorance: Villain sets out to kill his co-worker, Victim, in order to secure a promotion. Villain knows that killing Victim will harm Victim. He also knows that the Victim's family and friends will suffer as a result. Nevertheless, Villain falsely concludes that killing another human being is permissible if it is the only way for you to make it in your career.

Intuitively, Villain is clearly, perhaps even paradigmatically, culpable for his intentions. The fact that Villain believes killing his co-worker to be permissible counts towards and not against Villain's culpability. I consider Villain's moral mistake to be a central part of what makes his conduct culpable and hence potentially liable.¹²³

This is reflected, as I discussed in Chapter Two, in my treatment of culpability. The fact that Villain's intention results from his repugnant moral beliefs clearly satisfies the condition that culpable intentions result from an insufficient sensitivity to morally significant considerations.

¹²³ In this way I agree with Scanlon, *What we Owe to Each Other* (Harvard UP, 1998) 284.

Moreover, since Villain is aware of the harm that killing Victim will cause, as well as the widespread disapproval of murder, he has more than sufficient evidence to realize that his intended action would be wrongful. In this way, my account supports the intuitive conclusion that morally ignorant people are culpable for their intentions.

However, many advocates of a volitionist, or choice based, account of culpability argue that people like Villain are not culpable.¹²⁴ Let us call adherents to this view moral ignorance exculpators. It is of course possible, though I think unlikely, that these moral ignorance exculpators are describing a different kind blameworthiness from the culpability that grounds liability. If this is the case, then there need be no direct conflict between their view and the one I am advancing. Nevertheless, it is worth considering the reasoning that lies behind exculpating to consider whether it undermines my conclusions. If morally ignorant people are not culpable, this would seriously limit the range of people that could be liable to defensive harming on the basis of their culpability.

The temptation to exculpate morally ignorant people arises because of the combination of two key premises.¹²⁵ The first premise is the claim that a person can only be culpable for something if the person was in control of that thing. The second premise, is what I have called the culpable ignorance premise, the claim that a person who makes a decision in ignorance of morally relevant facts about her situation can only be culpable for that decision if she was culpable for the ignorance in which she made her decision. These premises combine because, as the moral ignorance exculpators point out, people are not generally in control, at least not in direct control, of their moral beliefs. This means, by virtue of the first premise, that moral ignorance cannot be culpable in itself.

¹²⁴ See for example, Gideon Rosen, "Culpability and Ignorance," "Skepticism about Moral Responsibility," and "Kleinbart the Oblivious and other Tales of Ignorance and Responsibility," *The Journal of Philosophy* (2008) 591-610. Also, Zimmerman, "Moral Responsibility and Ignorance," and *Living with Uncertainty*.

¹²⁵ I have simplified this argument considerably to reveal its key points. This argument is most clearly laid out by Zimmerman in *Living with Uncertainty*, 175. It is also expressed by Rosen in "Culpability and Ignorance."

Thus, by virtue of the second premise, decisions made on the basis of moral ignorance cannot be culpable.

The only exception to this, allowed on some views, is that moral ignorance can be derivatively culpable if it resulted from a prior culpable failure to take steps to prevent that ignorance. Of course, this prior culpable failure cannot itself be done from moral ignorance, or the whole argument repeats itself once again. Thus, the moral ignorance exculpators conclude, intentions can only be culpable if they are instances of *akrasia* – actions done contrary to what the person genuinely believes she ought to do – or are done from ignorance that resulted from a previous instance of *akrasia*.¹²⁶

The problem with this argument, besides its unintuitive conclusions, is its reliance on the culpable ignorance premise. As I argued above, at least for the kind of culpability that justifies liability, it is not the case that an action done in ignorance can only be culpable if the ignorance it resulted from was culpable. Beliefs are simply not the sort of thing to which action-guiding reasons apply, so they cannot be the object of evaluation for a choice based theory of culpability. To the extent that culpability is an evaluation of the moral quality of our choices, beliefs are in themselves neither culpable nor un-culpable.

This is compatible with there being some other sense in which negligent and morally ignorant people should have known better. Yet if we substitute this weaker claim, then the exculpator's argument no longer works. This is because a lack of control over something does not mean that it should not have been otherwise. It can be the case that a person should have known better, even though people do not directly control their beliefs in the way they control their actions.

¹²⁶ This way of using the word '*akrasia*' comes from Rosen, see "Skepticism about Moral Responsibility," 309. I will not consider whether it is a genuine instance of *akrasia* in the traditional sense.

The culpable ignorance premise seems plausible because it equivocates between different kinds of fault. It may be true that an action done in ignorance can only be culpable, in the sense that justifies liability, if the ignorance was the result of some kind of moral and epistemic failing. This latter kind of failure need only be a part of, but not the same as, the kind of moral fault measured by the kind of culpability that justifies liability.

3.1 Partially Exculpatory Moral Ignorance

There remain several hard cases of moral ignorance where exculpation is tempting. Epistemic limitations that result from the person's cultural or historical circumstances may warrant a slight reduction in culpability. In all these hard cases, I contend, it is not the person's moral ignorance itself that is potentially exculpating, but rather features of the person's circumstances that contribute to that ignorance.

Consider Gideon Rosen's case of the ancient slave-owner who lived at a time when no one questioned the practice of slavery.¹²⁷ Rosen writes,

No one denied that it was bad to be a slave, just as it is bad to be sick or deformed. The evidence suggests, however, that until quite late in antiquity it never occurred to anyone to object to slavery on grounds of moral or religious principle ... Unlike race slavery in the Americas, [ancient] slavery was not supported by myths about the biological or psychological inferiority of the slave. One became a slave through bad luck or imprudence; in principle the status could befall almost anyone.¹²⁸

If we make this assumption, then there is certainly something compelling in the thought that the ancient slave-owner is less culpable than he would otherwise be.

¹²⁷ My understanding of this case comes from Rosen, although it was previously discussed by Michele Moody-Adams, "Culture Responsibility, and Affected Ignorance," *Ethics* (1994) 291–309.

¹²⁸ Rosen, "Culpability and Ignorance," 64.

This case is not unique. Susan Wolf's case of JoJo, the son of a vindictive dictator whose only moral education came from watching his father's misdeeds, has many of the same features.¹²⁹ In all these cases we are torn between our intuition that these people are culpable, and potentially liable, when they form wrongful intentions, and our temptation to exculpate morally ignorant people whose cultural or historical circumstances made it more difficult for them to avoid moral ignorance.¹³⁰

Consider, for example, the ancient slave-owner. It seems clear that this person ought to be liable to defensive harming. If, for example, the only way to free his slaves involved killing the slave-owner this would be permissible. It is, of course, possible that the ancient slave-owner's liability depends on something other than culpability. Yet, I can find no evidence of this in my intuitions. Causal connections, for example, do not suddenly appear more relevant when I learn that this slave-owner lived at a time when slavery went unquestioned. It appears permissible to harm the slave-owner if this is the only way to free somebody else's slaves, or to prevent some substantial harm from befalling them. Presumably, therefore, it is the slave-owner's culpability for keeping slaves that makes him liable to this defensive action. If my other arguments have been correct, this strongly supports my intuition that the ancient slave-owner is culpable for his intentions.

At the same time, however, the ancient slave-owner is clearly less culpable than a person who keeps slaves in the 20th century and is aware of moral arguments against slavery. On the assumption that culpability is what grounds the liability of these slave-owners, if we had a choice between freeing some slaves either by killing the ancient slave-owner or by killing a person who believes slavery to be permissible despite living at a time when moral arguments against slavery are widespread, I contend that we should prefer the latter option. The 20th century slave-owner, even if

¹²⁹Susan Wolf, "Sanity and the Metaphysics of Responsibility," *Responsibility, Character, and the Emotions*, ed. by F. D. Schoeman (Cambridge UP, 1987).

¹³⁰ For a good summary of these issues see Elinor Mason "Moral Ignorance and Blameworthiness."

he too is morally ignorant, appears more culpable, and hence more liable relative to the same harm prevented, when compared to the ancient slave-owner.

This suggests two things. First, it suggests that moral ignorance on its own is not exculpatory. Second, it suggests that in at least some hard cases of moral ignorance, the person's circumstances give us reason to treat them as slightly less culpable, and thus slightly less potentially liable, than they would otherwise be.

Consider the evidence-relative wrongfulness of slavery in each of these cases. The 20th century slave-owner, we can safely conclude, has much better evidence that slavery is wrongful than is available to the ancient slave-owner. This, I think, is the key reason for holding, at an intuitive level, that while both slave-owners are culpable, the ancient slave-owner is slightly less culpable than the 20th century slave-owner. Making this case at the level of theory, however, is more difficult than it might appear.

An intention is wrongful, according to the evidence-relative standard, to the extent that it would be objectively wrongful if the facts of the situation were as the person has decisive reason to believe based on the available evidence. To this, I will add the additional claim that an intention is more wrongful, according to the evidence-relative standard, the clearer and more obvious the evidence that the intended action is objectively wrongful. This is a plausible addition which captures our intuitions in a range of cases. Recall the case of Careless, who puts his co-workers at risk. If the evidence Careless overlooks is very clear, this plausibly makes Careless more culpable for putting his co-workers at risk than if the evidence he overlooks was not as obvious.

This has implications for the case of moral ignorance. Without getting enmeshed in meta-ethical and epistemological debates, it may be helpful here to draw a simple distinction between two kinds of evidence for the wrongfulness of an intended action. The most important kind of evidence, what we might – for the sake of discussion – call primary moral evidence, is evidence regarding what

I have called morally significant considerations. These are the non-moral facts on which the wrongfulness of an intended action supervenes. In the case of Careless, this is evidence about the danger facing his co-workers. In the case of slavery this includes things such as the suffering inflicted on slaves.

This kind of evidence alone should be sufficient to make it the case that a given intention is wrongful from the evidence-relative perspective. It is very reasonable for a person confronted with evidence of the suffering that slavery inflicts on its victims to conclude that slavery is wrongful. Of course, the word reasonable can be used to describe beliefs and intentions in many different ways. I do not insist, for example, on the further claim that there is something these slave-owners could reasonably be expected to do to avoid their ignorance.¹³¹ Rather, I am making a point about the quality of the evidence available to them. The ancient and 20th century slave-owners can be assumed to have clear primary moral evidence for the objective wrongfulness of slavery. This supports the conclusion that slavery is wrongful from the evidence-relative perspectives of both people.

There is a different sort of evidence, however, that can make the objective wrongfulness of an intended action even clearer. What we might call secondary moral evidence, is evidence other than primary moral evidence that can make the objective wrongfulness of a given intention more or less obvious. This can include facts about other people's moral beliefs, and arguments for or against different moral conclusions.

In cases such as the two slave-owners, thinking about this kind of evidence can help explain why we consider the ancient slave-owner to be slightly less culpable. The fact that the ancient slave-owner has never encountered the idea that slavery is wrong, and is unaware of the moral arguments against slavery that are commonplace today, makes the wrongfulness of slavery less obvious from

¹³¹ This further claim may be defensible but it is more controversial. See Neil Levy "Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick," *Ethics* (2009) 729-741.

his perspective. When we take these different kinds of evidence into account, we reach the intuitive conclusion that while both slave-owners are culpable, and hence potentially liable, the ancient-slave owner's historical and cultural limitations make him slightly less culpable than he would otherwise be.

These conclusions transfer to other cases where the person's historical or cultural circumstances make it more difficult for the person to avoid moral ignorance. Thus, on my view, people such as JoJo may be slightly less culpable, and thus less liable relative to the same harm prevented, when compared to equivalently ignorant people in better epistemic situations. In this way, my proposed account gets intuitive results in these hard cases of moral ignorance.

3.2 Moral Ignorance and the Nature of Psychopathy

Not all hard cases of moral ignorance, however, are cases where the ignorance resulted from cultural or historical epistemic limitations. Some are tempted by the suggestion that people who are predisposed to be morally ignorant and to be unmoved by moral considerations deserve exculpation because of this condition.¹³² Consider, for example, Vicious who due to a genetic condition is much less motivated by morally significant considerations and is consequently pre-disposed to form false moral beliefs. Depending on the best definition of psychopathy, we might choose to call Vicious a psychopath. If this condition is severe then it is tempting to conclude that it counts against her culpability when she forms a wrongful intention.¹³³

Nevertheless, on my account a person such as Vicious is highly culpable for her wrongful intentions. In my view, Vicious's serious lack of concern for morally significant considerations counts towards and not against her culpability. It seems I ought to provide a reason, therefore, for

¹³² Garry Watson, "The Trouble with Psychopaths," *Reasons and Recognition* ed. by R. J. Wallace, R. Kumar, and S. R. Freeman (OUP, 2011) 308-324.

¹³³ Watson, "The Trouble with Psychopaths."

resisting the inclination to exculpate in hard cases such as Vicious's but not in cases such as the ancient slave-owner.

My reason for making this distinction is that our temptation to excuse Vicious contains a certain inconsistency. While we are tempted to exculpate people such as Vicious, who suffer from a very serious lack of moral motivation, but we do not feel the same way about people who suffer from the same condition as Vicious but to a lesser extent.

Consider Nasty, a person who is only slightly below average in his predisposition to be motivated by morally significant considerations. We would consider Nasty to be clearly culpable for any of his wrongful intentions. Indeed, other things being equal, as we move along a scale of decreasing moral motivation, from Nasty to Vicious, we find that people appear increasingly culpable until we get to the very end of the scale. In this way, our intuitions about people who lack moral motivation are strangely conflicted. We find moderate evil to be paradigmatically culpable but are then tempted to consider extreme evil a basis for exculpation. To be consistent our intuitions about the culpability of Nasty must be balanced against our temptation to exculpate Vicious.

This, of course, leaves open the possibility that there will be versions of Vicious's case that do deserve partial or complete exculpation. I am certainly not in a position to rule out this possibility, or to point to specific conditions or varieties of psychopathy that are or are not grounds for exculpation. All I can do is point in the direction that I think such answers might lie.

If there is a substantial difference of kind, and not just of degree, between Vicious's condition and Nasty's then this suggests that there might be grounds for exculpation that would not fall victim to this inconsistency. On the other hand, if Vicious's condition is just a more extreme version of the condition that made Nasty culpable, then I do not see how this could be the basis of an excuse.

Certainly, it will not suffice to point out that Nasty is only insensitive to some morally significant considerations while Vicious's lack of moral concern is total. This claim simply drives home the point that Vicious has more of the same characteristic that makes Nasty and all other wrongdoers culpable. Nor will it suffice to simply point out that Nasty is theoretically capable of recognizing his mistake – even if in practice this is unlikely – while Vicious is not. We could imagine that psychopathy was in rare instances suddenly cured, but this would not change the fundamental nature of the condition. It would be strange to suppose that the culpability of Vicious depends on whether she suffers from a strain of psychopathy that has been known to reverse in some cases.

I have not, in this thesis, had an opportunity to consider in detail the qualities that make a person the fitting object of judgments of culpability in the first place. If there are versions of Vicious's case that deserve exculpation, it is perhaps because these people, like children, lack some of the qualities necessary to be the fitting object of culpability judgments. This is one of the many places where a full understanding of culpability requires us to first understand something else. Here, as elsewhere, culpability is as much about the nature of agency and identity as it is about measuring moral fault.

4. Conclusion

This chapter adds more substance to the account of the kind of culpability that justifies liability, which I began in Chapter Five. In the first section, I discussed what it takes for an intention to manifest insufficient sensitivity to action-guiding moral reasons. I argued that we must consider both the content of the intention and the way in which the intention came about. While culpability is fundamentally a measure of our sensitivity to action-guiding moral reasons, measuring this sensitivity requires us to consider other things, such as beliefs and desires, to which action-guiding moral

reasons do not directly apply. In this way, my account of culpability sits between the rival volitionist and attributivist accounts of culpability.

In the second section, I showed how this account of culpability can easily explain the culpability of people who have wrongful intentions but have yet to commit a wrongful action. Finally, in the third section, I considered some hard cases raised by moral ignorance. Here, as elsewhere in the thesis, I was not able to provide definitive answers to these interesting questions. I simply tried to point in the direction in which these answers might lie.

Conclusion

The central theme of this thesis has been the relationship between culpability and liability. I argued, in Chapter One, that a liability is a reason to discount the interests of the liable agent when deciding what to do in cases of defensive harming. In Chapter Two I argued that the significance of choice can justify self-defense and other cases of liability. I showed how a liability can be generated by culpability, because the morally weighted costs of avoiding culpable actions are so low. This argument was supplemented in Chapter Three, where I argued that our duties do not directly give us our liabilities and that thinking about our duties can give us more reason to doubt the significance of causal responsibility in cases of defensive harming. In Chapter Four, I argued that causal responsibility for a threat is not a necessary condition of liability to be harmed in the prevention of that threat.

In this way, I argued that the significance of culpability is sufficient to generate liability, and to justify self-defense and other cases of defensive harming. Having argued for the importance of culpability, I used the remainder of the thesis, in Chapters Five and Six, to sketch an account of the kind of culpability that justifies liability. I presented a moral reasons based account of culpability, that sits between the rival volitionist and attributivist approaches.

I have endeavored to connect questions about the nature of liability with questions about the nature of culpability. These two sets of questions are usually treated separately, yet there is a benefit to studying them together. As I showed in Chapter Four, we can resolve some concerns about liability by looking more closely at the nature of culpability. Moreover, as I suggested in Chapter Five, we can make sense of the often amorphous concept of culpability by considering what culpability is for. Perhaps by focusing on breadth, by trying to connect usually distinct discussions, I

have missed opportunities to focus on depth. Throughout this thesis I have raised interesting questions for which I could not provide an answer. However, I believe my approach is justified because there is a certain kind of depth that can only be achieved by focusing on breadth. As I have argued, we cannot fully understand liability without an account of culpability, and we cannot fully understand culpability without an account of liability.

Bibliography

Alexander, Larry and Kimberly Kessler Ferzan. *Crime and Culpability*. Cambridge University Press, 2009.

Arpaly, Nomy. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press, 2003.

Christie, Lars. *Killing One to Save Another*. PhD Thesis: University of Oslo, 2015.

Coons, Christian and Michael Weber eds. *The Ethics of Self-Defense*. Oxford UP, 2016.

Crisp, Roger. "Review of Jon Kupperman, *Value...and What Follows*." *Philosophy* 75, 200 (2000) 458-462.

Duff, Anthony. *Criminal Attempts*. Oxford UP, 1997.

Fletcher, George. *Rethinking Criminal Law*. Oxford UP, 2000.

Frowe, Helen. *Defensive Killing: An Essay on War and Self-Defence*. Oxford UP, 2014.

Harman, Elizabeth. "Does Moral Ignorance Exculpate?" *Ratio* 24, 4 (2011) 443-468.

Hieronymi, Pamela. "The Wrong Kind of Reason." *The Journal of Philosophy* 102, 9 (2005) 437-457.

Husak, Douglas. *Ignorance of Law: A Philosophical Inquiry*. Oxford UP, 2016.

Husak, Douglas. "Intoxication and Culpability." *Criminal Law and Philosophy* 6, 3 (2012) 363-379.

Kamm, Francis. *Ethics for Enemies: Terror, Torture, and War*. Oxford UP, 2011.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor. Cambridge UP, 2012.

Kumar, Rahul. "Contractualism and the Roots of Responsibility." *The Nature of Moral Responsibility: New Essays* edited by Clarke, McKenna, and Smith. Oxford UP, 2015.

Lazar, Seth. "Necessity in Self-Defense and War." *Philosophy and Public Affairs* 40 (2012) 3-44.

Levy, Neil. "The Good, the Bad, and the Blameworthy." *Journal of Ethics and Social Philosophy* 1, 2 (2005) 1-16.

Levy, Neil. "Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick." *Ethics* 119, 4 (2009) 729-741.

Mason, Elinor. "Moral Ignorance and Blameworthiness." *Philosophical Studies* 172 (2015) 3037-3057.

McMahan, Jeff. "The Basis of Moral Liability to Defensive Killing." *Philosophical Issues* 15, 1 (2005) 386-405.

McMahan, Jeff. "Self Defense and Culpability." *Law and Philosophy* (2005).

McMahan, Jeff. "Preventive War and the Killing of the Innocent." *The Ethics of War* edited by David Rodin and Richard Sorabji (Ashgate Publishing, 2005).

McMahan, Jeff. "Proportionate Defense." *Journal of Transnational Law and Policy* 23 (2013).

McMahan, Jeff. "The Limits of Self-Defense." In *The Ethics of Self-Defense* edited by Christian Coons and Michael Weber. Oxford UP, 2016.

McMahan, Jeff. "Liability, Proportionality, and the Number of Aggressors." In *The Ethics of War* edited by Saba Bazargan and Samuel Rickless. Oxford UP, 2017.

Mele, Alfred. *The Springs of Action*. Oxford UP, 1992.

Moody-Adams, Michele. "Culture Responsibility, and Affected Ignorance." *Ethics* 104, 2 (1994) 291-309.

- Nagel, Thomas. "Moral Luck." *Proceedings of the Aristotelian Society, Supplementary Volumes* 50 (1976).
- Olsaretti, Serena. "Scanlon on Responsibility and the Value of Choice." *Journal of Moral Philosophy* 10, 4 (2013) 465-483.
- Overland, Gerhard. "Moral Taint: On the Transfer of the Implications of Moral Culpability." *Journal of Applied Philosophy* 28, 2 (May 2011) 122-136.
- Parfit, Derek. *Reasons and Persons*. Oxford UP, 1984.
- Parfit, Derek. *On What Matters*. Oxford UP, 2011.
- Peels, Rick. "What Is Ignorance?" *Philosophia* 57, 1 (2010) 57-67.
- Quong, John. *The Morality of Defensive Force* (Forthcoming).
- Renzo, Massimo. "Rights Forfeiture and Liability to Harm." *The Journal of Political Philosophy* 25, 3 (2017) 324-342.
- Rodin, David. "Justifying Harm." *Ethics* 122 (2011) 74-119.
- Rosen, Gideon. "Culpability and Ignorance." *Proceedings of the Aristotelian Society* 103 (2003) 61-84.
- Rosen, Gideon. "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18, 1 (2004) 295-313.
- Rosen, Gideon. "Kleinbart the Oblivious and Other Tales of Ignorance and Responsibility." *The Journal of Philosophy* 105, 10 (2008) 591-610.
- Scanlon, Thomas. "The Significance of Choice" *The Tanner Lectures on Human Values* 7 (1986) 149-216.
- Scanlon, Thomas. *What We Owe to Each Other*. Harvard University Press, 1998.

Scanlon, Thomas. "Thomson on Self-Defence" In *Fact and Value: Essays on Ethics and Metaphysics for Judith Jarvis Thomson* edited by A. Byrne, R. Stalnaker, and R. Wedgwood. MIT University Press, 2001.

Scanlon, Thomas. "Responsibility and the Value of Choice." *Think* (Spring, 2013).

Smith, Angela M. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, 2 (2005) 236-271.

Smith, Holly. "Culpable Ignorance." *The Philosophical Review* 92, 4 (1983) 543-571.

Smith, Holly. "Non-Tracing Cases of Culpable Ignorance." *Criminal Law and Philosophy* 5, 2 (2011) 115-146.

Smith, Holly. "Dual-Process Theory and Moral Responsibility." In *The Nature of Moral Responsibility: New Essays* edited by Clarke, McKenna, and Smith. Oxford UP, 2015.

Steel, Sandy. *Proof of Causation in Tort Law*. Cambridge UP, 2011.

Stephenson, Robert Louis. *The Strange Case of Dr. Jekyll and Mr. Hyde*. Scribner's Sons, 1886.

Tadros, Victor. *The Ends of Harm*. Oxford UP, 2011.

Tadros, Victor. "Causation, Culpability, and Liability." In *The Ethics of Self-Defense* edited by Christian Coons and Michael Weber. Oxford UP, 2016.

Temkin, Larry. *Rethinking the Good*. Oxford UP, 2012.

Thomson, Judith Jarvis. "Remarks on Causation and Liability." *Philosophy and Public Affairs* 13, 2 (Spring, 1984).

Thomson, Judith Jarvis. *The Realm of Rights*. Harvard UP, 1990.

Thomson, Judith Jarvis. "Self-Defense." *Philosophy and Public Affairs* 20, 4 (1991) 283-310.

Tomlin, Patrick. "On Limited Aggregation." *Philosophy and Public Affairs* 45, 3 (2017).

Watson, Garry. "The Trouble with Psychopaths." *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon* edited by R. J. Wallace, R. Kumar, and S. R. Freeman. Oxford UP, 2011.

Watson, Garry. "Two Faces of Responsibility." *Philosophical Topics* 24, 2 (1996) 227–248.

Williams, Bernard. "Moral Luck." *Proceedings of the Aristotelian Society, Supplementary Volumes* 50 (1976).

Wolf, Susan. "Sanity and the Metaphysics of Responsibility." *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* edited by F. D. Schoeman. Cambridge UP, 1987.

Zimmerman, Michael. "Moral Responsibility and Ignorance." *Ethics* (1997) 410-426.

Zimmerman, Michael. *Living with Uncertainty*. Cambridge UP, 2008.