Beyond Moral Responsibility and Lesser-Evils:

Moral Desert as a Supplementary Justification for Defensive Killing

By

James T. Murray

A thesis submitted to the Graduate Program in Philosophy

in conformity with the requirements for the

Degree of Master of Arts

Queen's University

Kingston, Ontario, Canada

November, 2014

Copyright © James T. Murray, 2014

Abstract

In recent years, philosopher Jeff McMahan has solidified an influential view that moral desert is irrelevant to the ethics of self-defense. This work aims to criticize this view by demonstrating that there are cases in which moral desert has a niche position in determining whether it may be permissible to kill a person in self- (or other-)defense. This is done by criticizing McMahan's Responsibility Account of liability as being overly punitive against minimally responsible threateners (MRTs), and by demonstrating, through reference to Saba Bazargan's Hybrid Account of liability to defensive harm, that further justification than mere liability is necessary in the justification of killing these kinds of threateners. It is also argued that other kinds of justifications like lesser-evil justifications are not necessarily strong enough to justify *killing* MRTs, even on top of established liability. This paper puts forward the view that, where a strong enough moral difference between threatener and innocent victim cannot be established, a justification on the basis of the implications of prior moral acts might, on some occasions, give us reasons to kill or not kill MRTs in self-defense.

Acknowledgements

First and foremost I would like to thank my supervisor, Professor Kerah Gordon-Solmon for her support and help throughout this process, and whose enthusiasm for the ethics of self-defense inspired me to write on this topic.

I would also like to thank my parents, Shawn and Cindy Murray for their belief in my potential, and their support—both emotional and financial. I couldn't have done this work without them. I want to also thank my brother, Chris Murray for his support and for always having a joke at the ready on tough days.

Bev Benson and Nick Sorge for allowing me to impose on them in the late stages of the writing process. I would be up a creek without a paddle without them.

Finally, I want to thank all my friends and colleagues in the philosophy department at Queen's who came out to hear me present an early version of this work. They provided some great comments and it was monumentally helpful. I specifically want to thank Jamie Shaw, Ian Allan, Ryan McSheffrey, Torin Doppelt, and David Pitcher for their great comments throughout the process, and for helping me to develop and/or reject some of my more ridiculous ideas. You guys are the best.

Contents

Abstract ii
Acknowledgements iii
Introduction 1
Chapter 1. 3
McMahan's Responsibility Account 3
The Culpability Account 6
The Responsibility Account 8
Criticisms of the Moral Responsibility Account and Responses 10
Fact-Relative over Evidence-Relative Permissibility 10
The View is Overly Punitive 14
Chapter 2. 17
The Hybrid Account 17
The Complex Account 17
The Lesser-Evil Discounting View 20
Objection: The Weakness of Lesser-Evil Justifications 23
Chapter 3. 30
A Supplementary Justification for Killing in Self-Defense 31
The Implications of Culpable Acts 38
Conclusions 43
Bibliography 45

Introduction

The aim of this project is to explore and refute Jeff McMahan's view, advanced in his 2009 book *Killing in War*, that desert is irrelevant to the morality of self-(or other-) defense.[1] On McMahan's view, desert and liability are conceptually separate, and, where liability is present, desert is unneeded. I will agree with the position that these concepts are indeed separate, but I do not discount desert from the picture. On McMahan's account, responsibility for a threat, and the liability that comes with it are the only things required in order to justify defensively killing culpable and minimally responsible threateners. In contrast, I will suggest a pluralistic view that accepts responsibility as a necessary condition for defensive killing in these cases, but not always a sufficient one. The conditions for defensive killing may be numerous, and I will argue, they may include desert.

In chapter 1 of this project, I will discuss McMahan's positive account of the permissibility of defensive harm. This will involve a reconstruction of the Responsibility Account of liability to defensive harm. This view has gained traction in recent years, and is amongst the most popular accounts of both liability to defensive harm, and justification for killing threateners. I will exposit the view by comparing it to other widely-held views of liability. This will help demonstrate the strengths of the view, and why it has become so popular. However, following this reconstruction I will discuss some potential difficulties that this account faces. Doing this will not completely refute the claims made by McMahan, but will at least bring into question the strength of the view as a monistic justification for defensive killing.

[1] Jeff McMahan, *Killing in War*, (Oxford: Oxford University Press, 2009), 8.

In response to these difficulties, the second chapter of this project will discuss a promising alternative view held by Saba Bazargan, which he calls the Hybrid Account of liability to defensive harm. Bazargan thinks that while minimally responsible threateners (MRTs) are liable to defensive harm, they are not liable to being defensively killed. His view is pluralistic in that he thinks that we can also use a different justification, a lesser-evil justification, in combination with liability, in order to justify killing MRTs. In response, I will argue that this lesser-evil justification combined with his 'complex account' of liability is still not strong enough to justify killing MRTs, and that a further form of justification would be required in order for the killing of MRTs in self- (or other-) defense to be morally permissible.

In chapter 3, I will use this conundrum as a base-point from which we might explore what other justifications might be used in order to separate the moral status of the defender and the threatener in cases involving MRTs. I will argue along similar lines to John Gardner and François Tanguay-Renaud's argument, in their paper, "Desert and Avoidability in Self-Defense". I will show how their argument can be developed to solve the conundrum that I presented in the end of chapter 2. While they ultimately play-down the usefulness of desert considerations, I argue that desert could be used to morally differentiate between defender and threatener where all else fails, even in cases involving MRTs. That is, in a "moral stalemate" (where it would seem that there are insufficient moral differences between parties), desert can be used in combination with responsibility-based liability and other justifications for killing in self-defense. The view I propose gives a positive account of what justifications might be used in justifying the killing of MRTs where McMahan's responsibility-based account of liability, and Bazargan's hybrid account fall short.

Chapter 1.

McMahan's Responsibility Account

Moral liability within the context of self-defense is a justification for harming. To say that someone is liable to defensive harm is just to say that they have performed some action that for which they have waived their right not to be harmed in self-defense. This person would not be *wronged* by the self-defender harming them in this instance. Being liable also makes it the case that one loses the right to defend oneself against this defensive harm.[2]

Consider the case of *Villainous Threat*:

> A moustache-twirling villain decides that in order to cure his boredom, he will design a situation that pits your life against your best friend's. If you do not kill your best friend, he will kill you and if your friend does not kill you, he will kill your friend. If either of you are able to follow through with the act, he will let the other one go, and if neither of you kill the other, he will kill you both. You are able though—due to the conveniences of philosophical thought experiments—to pull a nearby lever that will drop a vat of boiling hot metal on the villain's head, killing him painfully and saving both of your lives. This is the only way to save both your lives.

Are you morally permitted to do this? The answer is a resounding "YES!", and one thing that even the most obdurate of ethicists would agree on. The villain has culpably threatened to kill you, and has therefore made himself liable to be killed in self- (and other-) defense.

That the villain is liable to be killed is undeniable. *Why* it is that the villain is liable is a bit more debatable. Plausibly though, he has forfeited his right not to be killed in this circumstance, while both you and your friend have retained your rights not to be killed. His rights were forfeited when he threatened to violate the rights of both you and your friend. He is culpable because culpability for an act entails that there is "both fault in the act and fault in the agent".[3] That is, the

[2] McMahan, *Killing in War*, 8.
[3] Jeff McMahan, "The Ethics of Killing in War", *Ethics*, 114(4) (2004), 693-733, 723.

agent is blameworthy for her action, and the action itself is impermissible. The villain easily fits this bill.

One does not need to be a moustache-twirling villain with dastardly schemes though, to be considered a culpable threatener. Consider the following case, *Drunk Driver*:

> A drunk driver is speeding down a residential street. The driver, not *intending* to harm anyone, loses control of his car, and is careening towards an innocent pedestrian. The pedestrian, mysteriously armed with a bazooka can fire a rocket at the car, killing the driver but stopping the car in its tracks and saving her own life.

In this example, we see that it is not intention that is necessary for culpability. Engaging in negligent or irresponsibly threatening actions with high risks of harming others can also make one culpable. The driver engages in an action that is impermissible—drinking and driving—and the driver is blameworthy for this action. A person is considered blameworthy when they are at fault for the outcome of a wrongful act. Since no other person forced him to drink and drive as some part of a different odd scenario, and he knows full well that what he does is dangerous, the driver is unmistakably blameworthy. Both the drunk driver, and the villain in the previous example, are considered *fully culpable* since they have no *justification*, *permission*, or *excuse*.[4]

Culpability is not always perfectly cut-and-dry as it is in the above examples. Culpability varies in degrees and can be applied to a wide range of acts. A person's degree of culpability is influenced by the strength of her excuses. One has an excuse for an action when the act is wrong but the agent is not fully blameworthy.[5] This is not to be confused for a *permission* to act. A person is morally permitted to act, according to McMahan, "when in the circumstances, it is not wrong to do it (even if it is of a *type* that is generally considered wrong)".[6] For example, killing is generally

[4] McMahan, *Killing in War*, 159.
[5] Ibid., 110.
[6] Ibid., 43.

considered wrong, but there are circumstances in which it is entirely permissible to act in that way. A person that acts without moral permission acts wrongly, although she may not be considered fully culpable. This person could also be deemed partially culpable, minimally responsible, or non-responsible.

Consider another scenario that I call, *Distracted Driver*:

> A person keeps her car well maintained and always drives cautiously and alertly. On one occasion though, she has been awaiting a very important phone call from a prospective employer. Upon seeing her cell phone ringing while driving, she is tempted to pick it up. She knows that this is dangerous, but she thinks, "It is just one time and I will be quick about it. People do this all of the time and nothing happens". However, due to this distraction, she loses control of the car. It has veered in the direction of the same pedestrian with a bazooka from the last example.

The Distracted Driver is what we might call, a *partially culpable* threatener. She is different from the drunk driver because, although she acts irresponsibly, she has an excuse for acting. This excuse is not particularly good, but at least it's somewhat exculpating. A partially culpable threat's excuses are too weak to meet the threshold of full exculpation however. According to McMahan, a person is partially excused if he is under duress and "the level of duress he is exposed could be resisted by a person of ordinary fortitude", if his is misinformed because "he has been negligent in investigating the facts relevant to the permissibility of his action," or if his capacity for responsible agency is partially, but not entirely diminished.[7] The distracted driver seems to fit this description due to a level of duress that she had difficulty in resisting. Many of us may have been able to resist taking the call, even if it was seriously important, but it is understandable that Distracted Driver had a lapse in judgement. This does not absolve her of guilt, but it would not be considered as morally reprehensible as deciding to drink and drive.

[7] Ibid., 160.

The distinctions between degrees of culpability are important because they are positively correlated with the degree of harm that one is liable to. This is because there are *proportionality restrictions* of killing in self-defense. For cases involving fully culpable lethal threateners of innocent victims, it is difficult to exceed these restrictions. Harms to these threateners are considered less morally weighty than the valuable goal of saving an innocent life because they have forfeited their rights not to be killed, and they are at fault for causing the harm. One might even argue that moral considerations against malicious threateners are so significantly weakened that there is nothing we may not do to them if required in self-(or other-) defense. However, we need not go this far, and I do not commit myself to this particular position.

We can say though, that for those who are only partially culpable, the proportionality constraints are tighter, since their liability is mitigated by excuses. For example, imagine someone is attacking you, on the grounds that he thinks you will otherwise harm a large group of people, and threatens to severely harm, but not kill you. Suppose he is understandably misinformed—a moderate effort to look into the matter will show that his source is not credible. "If the excuse significantly diminishes his culpability," writes McMahan, "killing him may be disproportionate".[8] If he had been fully culpable however, you would have been permitted to kill him.

The Culpability Account

The Culpability Account is the view that what makes a person liable to defensive harm is culpability for an unjust threat. Again, culpability requires blameworthiness for an unjust threat without a proper excuse. The cases of *Villainous Threat, Drunk Driver,* and *Distracted Driver* are

[8] Ibid., 161.

all ones in which, according to this view, the threatener is liable to be killed in self-defense. This seems to account for a very commonly shared intuition. It would be considered absurd that the innocent defenders in these situations would not be permitted to kill their threats in self-defense. The culpability of the threateners is a good reason to believe that they have forfeited their rights (at least temporarily) not to be killed in self-defense.

McMahan's objection to the Culpability Account is that it is not broad enough to account for the liability of minimally responsible threateners (MRTs). He wants us to consider the following case, *Conscientious Driver*:

> A person keeps her car well maintained and always drives cautiously and alertly. On one occasion, however, freak circumstances cause the car to go out of control. It has veered in the direction of a pedestrian whom it will kill unless she blows it up by using one of the explosive devices with which pedestrians in philosophical examples are typically equipped.[9]

McMahan argues that since the pedestrian has done nothing to make herself liable to be killed, and the driver has made himself liable, the pedestrian is permitted to kill the driver in self-defense. The driver is liable, despite not being culpable, because, "as a morally responsible agent, he voluntarily chose to set a couple tons of steel rolling as a means of pursuing his ends, knowing this would involve a tiny risk that he would lose control of this dangerous object that he had set in motion, thereby imperiling the lives of the innocent".[10] Despite both the driver and the pedestrian having engaged in generally morally permissible actions, the fact that the driver's action is risk-imposing makes him "responsible when the risks he imposed eventuate in harms".[11]

[9] Ibid., 393.
[10] Ibid., 394.
[11] Ibid.

The Responsibility Account

As the name suggests, the Responsibility Account of liability to defensive harm holds responsibility as the basis for liability. Specifically, "the criterion of liability to defensive killing is moral responsibility, through action that lacks objective justification, for a threat of unjust harm to others, where a harm is unjust if it is one to which the victim is not liable and to which she has not consented".[12] Lacking objective justification is important here in that it contrasts with what the driver is evidence-relative permitted to do.[13] The Responsibility Account sees the *fact-relative* wrongness of an act as more important to liability than the *evidence-relative* wrongness.

According to Derek Parfit, who originated terminology, an act is "*wrong* in the *fact-relative* sense just when this act would be wrong in the ordinary sense if we knew all of the morally relevant facts".[14] In *Conscientious Driver*, the fact that in this instance of driving, she will hit and kill a person is unknowable to the Driver, and from her point of view, she does nothing wrong in deciding to drive. Based on what she can tell, this instance of driving should be just like any other instance where she does not threaten anyone. However, it is objectively the case that the pedestrian is wronged, since she has not done anything to give up her right not to be killed.

In contrast, an act is "*wrong* in the *evidence*-relative sense just when this act would be wrong in the ordinary sense if we believed what the available evidence gives us decisive reasons to believe and these beliefs were true".[15] This is, if it were genuinely the case, as the Driver

[12] Ibid.
[13] The driver is only morally *permitted* to act here, rather than being justified, since according to McMahan, he "has no positive moral reason to engage in the activity that he knows has a tiny risk of unintentionally killing a bystander" (*Killing in War*, 165).
[14] Derek Parfit, *On What Matters*, (Oxford: Oxford University Press, 2011), 150.
[15] Ibid., 151.

believes, and has good evidence to believe (based on prior data), that she will not pose a threat to anyone, then she would be permitted to act. Her act would not be evidence-relative wrong because based on the evidence, her act is permissible. So, the Conscientious Driver is evidence-relative, though not fact-relative permitted to act.[16]

This is important to note since it appears that those who are considered culpable threateners are neither fact-relative nor evidence-relative permitted to act. By only taking into account the presence objective justification, the Responsibility account widens its scope as to who can be considered liable. The Conscientious Driver's intentions are only important insofar as they mitigate proportionality restrictions, but unimportant in whether she is saddled with at least some degree of liability.

MRTs, like the Conscientious Driver are not liable to defensive harm to the same degree as fully or partially culpable threateners. Recall, partially culpable threateners have partial excuses, and therefore, the proportionality restrictions on them are tighter than they are on those with no excuse at all. MRTs have full excuses for acting, which are derived from their subjective (or evidence-relative) justifications. This means that proportionality restrictions are especially tight on them. If it were possible that the driver and the pedestrian could split the harms, it would be the case that the driver would only be liable to *slightly* more harm than the pedestrian. If the pedestrian could refrain from killing the driver, and take upon herself a broken leg, the pedestrian would not be permitted to kill the driver. It is only because the harms cannot be divided, and that one or the other *must* die, that the driver, who is slightly liable, while the pedestrian is not liable at all, should

[16] The driver is only morally *permitted* to act here, rather than being justified, since according to McMahan, she "has no positive moral reason to engage in the activity that she knows has a tiny risk of unintentionally killing a bystander" (*Killing in War*, 165). For a justification to be present there must be both permission and a reason.

bear the harm. It is the mitigation of this unfairness that McMahan sees as the advantage of the Responsibility Account over the Culpability Account.

Criticisms of the Moral Responsibility Account and Responses

Fact-Relative over Evidence-Relative Permissibility

A question that can be asked about the Responsibility Account, is why fact-relative permissibility is held to be more morally significant than evidence-relative permissibility. This is the inspiration behind the above-mentioned Culpability Account. While it is the case that victims in MRT cases are wronged, it does not seem to be the case that it is because the MRT has done anything wrong. From the point of view of the MRT, they engage in actions that are entirely permissible. They bear no ill will towards those that they threaten, nor do they violate some prior duty.

This is especially true when there is no way of knowing that the harm they threaten is objectively wrongful, as in the case of the Dignitary. In *Dignitary:*

> Imagine that you extend your hand to shake the hand of some foreign dignitary at a reception. Unbeknown to you, a third party projects a stunningly realistic holographic image of a pistol onto your hand. The dignitary, who is accustomed to threats on her life, sees the hologram, forms the justified belief that you are about to assassinate her, and coolly draws a pistol in order to shoot you down in self-defense.[17]

Michael Otsuka reasons that the Dignitary is liable to defensive harm "because she is of sound mind, in control of her actions, and aware of their likely consequences".[18] Otsuka sees this case as an illuminating defense of the Responsibility Account, but I see it as obfuscating. This case involves an MRT since the Dignitary is justified in her belief that she is under threat, and that the

[17] Michael Otsuka, "The Moral Responsibility Account of Liability to Defensive Killing," *Bowling Green State University Workshop in Ethics and Public Policy* (Bowling Green, Ohio, 2013), 1-2.
[18] Ibid., 2.

only way to prevent it is to engage in a harmful action. Under the circumstance that the threat is real, this harmful action is permissible, but there was evidently a minimal risk that the Dignitary might shoot an innocent person. The Dignitary is fact-relative unjustified in acting, and minimally responsible for her action. So on the Responsibility Account, the dignitary, like the Conscientious Driver is liable to be killed in self-defense. However, this case seems to give weight to the notion that evidence-relative permission should be taken into closer account. What the Dignitary does is fact-relative impermissible. However, the Dignitary is overwhelmingly well-justified in acting in an evidence-relative sense.

When the dignitary is looking down the barrel of a gun, and she has every reason to believe that her life is in immediate danger, there is, in her mind, no time to consider the outside chance[19] that the gun is not in fact real. Liability on the Responsibility Account is not concerned with one's moral character or intentions, but with what might be outside the possible consideration of the agent involved. The fact that there is no failure in moral deliberation where MRTs are concerned makes it seem as though this is not a moral issue at all, but rather, one of very bad luck, which we might view as somewhat arbitrary.

In response to this objection, Otsuka brings to the table the Dworkinian notions of 'brute luck' and 'option luck'. He writes, "the charge that it would be unfair if people such as the dignitary were liable to be killed is answered by pointing to the fact that they gambled with their moral liability to being defensively killed by engaging in such activity that they knew to be risky".[20] He quotes Dworkin and defines option luck as "a matter of . . .whether someone gains or loses through accepting an isolated risk he or she should have anticipated and might have declined".[21] This is

[19] Here, 'outside chance' is really understating the unlikeliness of the scenario.
[20] Otsuka, "Moral Responsibility Account", 17.
[21] Ibid.

opposed to brute luck, which is "a matter of how risks fall out that are not in that sense… gambles"[22]. In this way, Otsuka thinks that MRTs have no claim to absolution from liability. It is generally risky to drive, or to carry around a gun with you. There is always a risk that you might be made liable to defensive harm by engaging in these activities, but we generally consider the rewards of these activities well worth the risk.

While McMahan and Otsuka both defend the Responsibility Account, they emphasize different things. McMahan places emphasis on fact-relative impermissibility as the indicator of responsibility and liability. Otsuka sees the brute luck/option luck distinction as being implicitly present in McMahan's version, but he thinks it is important enough to bring to the forefront. The importance of this foregrounding can be seen in the following example where MRTs are differentiated from non-responsible threats. *Cell Phone Operator:*

> A Man's cell phone has, without his knowledge, been reprogrammed so that when he next presses the "send" button, the phone will send a signal that will detonate a bomb that will then kill an innocent person.

This case is considered different from the other cases because the risk could not be anticipated the way that it could be in *Conscientious Driver* or *Dignitary.* Otsuka and McMahan refer to this kind of threat as a non-responsible threatener. If we apply the notions of brute luck and option luck to this case, we have to acknowledge that this is clearly a case of brute luck. While it is theoretically possible that the cell phone operator could have considered the possibility that his phone was connected to a bomb, expecting him to do so, and to refrain from acting is beyond the realm of plausibility. In McMahan's words, "he is nonculpably and invincibly ignorant *that he poses any kind of threat to risk of harm to anyone"*.[23]

[22] Ibid., 17-18.
[23] McMahan, *Killing in War*, 168.

In response to this, one might argue that, especially in the case of the dignitary, despite the fact that the aim in shooting in this case is to purposefully harm someone, the evidence is so stacked in favor of the Dignitary that the person she is shooting is a culpable threatener, that to call pulling out her gun and shooting in this instance a gamble is to stretch the notion of "gamble" so far as to make it no longer normatively useful. Recall, for some result of an act to be considered the result of bad option luck, an isolated risk must have been accepted that the person should have anticipated or might have declined. As Kasper Lippert-Rasmussen notes though, "being in a position to choose another gamble is not the same thing as being able to decline the former gamble… for it might be the case that whatever one does one will face the risk of an unacceptably bad outcome.[24] In the case of the Dignitary, there are two choices: (1) Take the impossibly small risk that the person she threatens is not in fact culpable, or (2) Take the very high risk that she faces certain doom. Choice (2) is the one that Otsuka suggests that the Dignitary may have made as an alternative, but she is being asked to take a serious personal risk with a very low probability of reward based on a ludicrously implausible scenario. The scenario occurring at all, should perhaps be chocked up to bad brute luck.

There is of course, always a possibility that carrying around a gun will result in the wrongful killing of an innocent. But considering that the Dignitary frequently faces threats on her life, and that she only draws her gun when faced with a credible threat, the isolated risk of her killing an innocent is quite low, but the personal risk to her in not carrying one is actually quite high.

The upshot of this discussion is that the distinction between bad brute luck and bad option luck in these cases is rather tenuous. Whether some outcome can be reasonably called the result of

[24] Kasper Lippert-Rasmussen, "Egalitarianism, Option Luck, and Responsibility", *Ethics* 111(3) (2001), 548-579, 557.

bad option luck when there is no plausible alternative to the otherwise morally permissible action seems questionable, especially when the scenario that the agent finds herself in seems to be the result of bad brute luck. Here, the fact-relative permissibility is determined by sheer luck, while the evidence-relative permissibility is carefully calculated, involving greater care on the part of the agent.

The *Conscientious Driver* scenario is more difficult to slot into this discussion, since, unlike the Dignitary, she acts in service of a minor (trivial) interest, and therefore, she does have a plausible alternative action (that is, not pursuing that interest) that she could have made in order to avoid imposing the small risk. Because of this, it *might* be plausible to consider the Conscientious Driver to be taking a gamble of the sort that could make here liable. If this is the case, then my first criticism might only apply to a narrow range of (unlikely) cases such as *Dignitary.* Given this point, our attention is drawn to the variability in the strengths of MRTs' moral reasons.

Although the Dignitary acts prudentially in pulling the trigger to save her life, she has a much stronger reason for acting than the driver. Given her stronger reason for acting, it would seem as though Dignitary and Driver should differ in their respective degrees of moral responsibility, Dignitary having a lower degree.[25] Problematically, under the Responsibility Account, in a scenario where harms cannot be distributed, Dignitary and Conscientious Driver are both liable to the same amount of harm—death. It is this fact that motivates my next criticism.

The View is Overly Punitive

The Responsibility Account draws its strength from its appeal to fairness. All things being equal, argue proponents of the Responsibility Account, it is unfair for the completely innocent

[25] McMahan, *Killing in War*, 167.

defender to bear a harm for which the threatener is at all responsible for. This is not to say that it is not *bad* that the threatener is killed. It is highly unfortunate that the threatener should be unlucky on this occasion and that she should find her minimally risky behaviour turn into a threat to the life of an innocent person. It would be a less fair outcome though, for the defender to die, since the scale of responsibility is slightly tilted toward the threatener.

A criticism that one might level against proponents of the Responsibility Account of liability is that it, in effect, treats all non-innocent threats equally. Despite the account's concern for fairness, there is another unfairness created whereby those who threaten culpably are liable to the same harms as MRTs as long as the harms that MRTs threaten are non-divisible.[26] On this view, it is the case that a drunk driver in a similar situation to *Conscientious Driver*, except that she threatens the pedestrian not due to mechanical failure, but because they are inebriated, is liable to the same harms as Conscientious Driver. It seems considerably more justifiable for Pedestrian to blow up Drunk Driver's car than for her to blow up Conscientious Driver's car since Drunk Driver is culpable for her actions. MRTs are treated the same way in certain circumstances as culpable threateners in terms of severity of harms. This makes the Responsibility Account come off as rather crude and imprecise.

Concluding Remarks

In this chapter I have reconstructed McMahan's Moral Responsibility Account of liability to defensive harm, as contrasted to plausible an alternative account of liability. The Culpability Account, although being intuitive when it comes to dealing with guilty threateners, does not make

[26] If harms were divisible, the story may be different, since defenders against culpable threateners might be justified in performing greater harm, even if the harm were theoretically divisible, since the defender has no obligation to take *any* harm upon herself. For example, if a fully culpable threatener threatens to kill me, but I can stop him by either killing him, or by breaking my leg, and his leg, then I may still be permitted to kill him to avoid my own broken leg.

the Conscientious Driver liable to any harms whatsoever since she is non-culpable. This seems counter-intuitive in itself, since the Driver forfeits her rights against at least *some* harms to save the pedestrian. This puts McMahan's Responsibility Account in strong standing except for the less intuitive claims as to the degree of the liability of MRTs. At this point, any degree of responsibility as grounding liability to defensive *harm* is a view that I endorse. However, the degree of that liability can justify too much on McMahan's view, as I have argued. Because of this, I will argue, the killing of MRTs cannot be justified on responsibility-based liability alone.

Chapter 2.

The Hybrid Account

In "Killing Minimally Responsible Threats", Saba Bazargan puts forward what he calls the hybrid justification of killing minimally responsible threats (MRTs) in attempt to replace McMahan's Responsibility Account.[27] In this chapter I will reconstruct Bazargan's argument for his view. In doing this I will discuss Bazargan's criticism of the simple account, his defense of the complex account, and his 'lesser-evil discounting view'. His complex account and lesser-evil discounting view together comprise his hybrid justification for killing MRTs.

Moral responsibility, liability, and culpability, on the Responsibility Account can be considered as part of a spectrum that measures degrees of liability. On the low end of the spectrum, 0, falls innocent and non-responsible threateners. They are both non-culpable, and entirely non-responsible. On the higher end, let's say 100, we have fully culpable and fully morally responsible threateners. MRTs fall somewhere in between, presumably closer to 0 on the spectrum, though not 0. MRTs are completely non-culpable because culpability requires both moral responsibility and blameworthiness. MRTs lack the latter. Taking into account the Conscientious Driver's minimal responsibility given that the risk of causing harm is very low, we can imagine her placement on the spectrum as being fairly close to 0. An understanding of this spectrum is essential in understanding the way that Bazargan's hybrid justification for the killing of MRTs differs from the responsibility-based liability justification.

The Complex Account

[27] Bazargan refers to the Responsibility Account as the "simple account", in contrast to the "complex account" that he proposes. For the sake of consistency I will continue refer to it as the Responsibility Account.

Bazargan's complex responsibility account (or "complex account") of liability is more stringent than the Responsibility Account. Where the simple account stipulates that "the party that is more responsible for a fact-relative wrongful threat is liable to suffer *that very degree* of harm, if necessary to prevent it from being imposed on her victim,"[28] the complex account is as follows: "If P is at least minimally responsible for an objectively unjust harm which she will impose on Q unless we pre-emptively harm P, then P is liable for no more than n% of the unjust harm for which she is responsible, where n is equal to the percent moral responsibility she bears for that unjust harm".[29] On this view, MRTs do not forfeit their right not to be killed. MRTs on the Responsibility Account, are *de facto* liable to the harm of death. Killing on the complex account violates narrow proportionality. In other words, it is disproportionate that the person who is only 5% responsible should be liable to 100% of the harm. This can be thought of as less draconian than the conventional view.[30] However, while the complex account does not wholly justify the killing of an MRT, the hybrid account that Bazargan puts forward can, and usually does have this outcome.

Before I move on to an explanation of the hybrid account, there is a stipulation about the MRT's degree of liability that needs to be fleshed out for the complex view. It is not made entirely clear by Bazargan but I will try to explain it. At the beginning of his discussion of the complex account, Bazargan writes, "the degree of harm to which an individual is liable is the harm that she threatens to cause *multiplied by* the percentage degree of her moral responsibility" (my italics).[31] Here it seems that he merely means that the percentage is a representation of degree of responsibility. If n=0, where *n* is a quantification of the degree of moral responsibility, then the threat is innocent, or completely non-responsible, and therefore non-culpable. If n=100 the

[28] Bazargan, "Killing Minimally Responsible Threats", 7.
[29] Ibid., 8-9.
[30] Ibid., 11.
[31] Ibid., 8.

individual is 100% responsible for a harm and therefore fully culpable. An MRT would be between, say, 5% and 10% responsible, and non-cuplable. This would presumably make her liable to be harmed in a way equal to between 5% and 10% of the harm that she is imposing on her potential victim.

On Bazargan's complex account, if 5% of death is (for the sake of this example) equivalent to a broken wrist, then the driver is liable to the harm she would be liable to is she were fully culpable for imposing a threat of a broken wrist. In preventing a person from fully culpably harming a person, according to McMahan, "it can be proportionate to inflict a significantly greater harm on a Culpable Threat if it is necessary to prevent him from inflicting a lesser threat on an innocent victim".[32] We may not be able to justify *killing* the threatener to prevent her from culpably breaking a person's wrist, but we may be justified in say, defensively lobbing off one of her limbs. This is a defensive harm 'M' times greater than the initial threat.

Here if M = 10 and h is a harm equal to 5% that of death, then the threatener is liable to a defensive harm equal to or less than (M x h) or 50% the threat imposed. This means that an MRT is liable to substantially more harm than we initially led to believe under Bazargan's initial articulation of his complex view.[33] However, in order to justify killing the MRT, this substantially higher amount of harm is still not enough. In this case, the threatened person or an onlooker is only justified in imposing a defensive harm on the MRT equal to half of the harm of killing. This is of little help to the potential victim if the only way for her to save herself is to kill the MRT. A proponent of the complex account may simply have to admit that it is morally wrong for the potential victim to kill the MRT in this instance. However, Bazargan thinks that there is a lesser-evil justification for the remaining harm that would lead to the killing of the MRT. When combined

[32] McMahan, *Killing in War*, 159.
[33] Bazargan, "Killing Minimally Responsible Threats", 14.

with his complex view, what he calls the "lesser-evil discounting view" becomes the hybrid account.

The Lesser-Evil Discounting View

The 'lesser-evil' justification, like liability, is a justification for harm. In this case, it is a justification for *infringing* upon a person's right not to be killed when the harm to be averted is significantly greater than the harm to be inflicted.[34] In cases where the lesser-evil justification comes into effect, the party whose rights are infringed retains their right not to be killed. However, where the lesser-evil justification is part of a hybrid account, the party does not retain their right not to be *harmed*, since they are liable to at least some extent.

An infringement of the right not to be killed would not be permissible in a case of self-defense against a *non-responsible* threatener who threatens only one person. Lives of individuals are usually weighed equally and killing one to save one is generally unjustifiable. For a rights-infringement of that magnitude to be justified it requires that there be enough at stake meeting a particular threshold. For instance, it may be impermissible to kill one to save one, but it may be permissible to kill one in order to save five. Thought experiments like the famous trolley problem are instrumental in theorizing what this threshold might be.[35] For those unfamiliar, the trolley problem is as follows:

> There is a trolley hurtling toward a group of five people stuck on the track. There is however, a side track onto which the trolley can be diverted before reaching the five. You can pull a lever to change the track, but there is a lone worker stuck on the side track who you would be kill by doing this.

[34] Ibid.
[35] See: Foot, Phillipa, "The Problem of Abortion and the Doctrine of Double Effect", the Oxford Review, Number 5, 1967.

With the trolley problem we are faced with a difficult decision: kill one or allow five to die. The difficulty in this decision stems from the moral relevance of the doing/allowing distinction. Killing a person outright seems much worse than allowing a person to die. The numbers at play however, weigh against the wrongness of the action. There is, according to common intuitions, some number of people that it would be morally impermissible to allow to die to avoid killing one.[36] This number is **R**, a threshold that when passed, makes infringing on a person's right not to be killed the 'lesser-evil', and therefore, justified. This threshold does not need to be quantified in terms of number of people at stake. Rather, that number of people of people at stake is equivalent to a badness that ought to be prevented.

Returning to our discussion of MRTs, a comparison between an MRT and a non-responsible threatener is important here because an MRT is only slightly (though non-marginally) more responsible for the threat. This means that as with the non-responsible threat, threshold R must be met to justify infringing on an MRT's right not to be killed.

Given that the MRT is not completely blameless, and that the complex account has already justified harming the MRT up to 50% of the threatened harm, R is already going to be a lower threshold than it would be if the MRT was a non-responsible threat. Bazargan uses the following example, "Suppose [an MRT's] potential victim is 20 years old and would lose 60 years of life by being killed. Accordingly, the lesser-evil justification has to be sufficiently strong to justify inflicting on the MRT a harm to which she is not liable that is equivalent to the loss of 30 years of life".[37] However, this is without having even factored in the moral relevance of the doing/allowing distinction. Bazargan elaborates:

[36] Note: This is highly controversial. On certain accounts, such as some strict Kantian views, there may be no number of people at risk high enough to justify the act of killing.
[37] Bazargan, "Killing Minimally Responsible Threats", 15.

> For example, might say that we have a lesser-evil justification for *intentionally killing* (my italics) a person who is not liable to be killed, only if that were necessary to save the lives of *ten* other non-liable individuals. But if we take a multiplier of ten as required for a lesser-evil justification [of intentional killings], we would have a lesser-evil justification for killing the MRT only if doing so would prevent the victim from losing 300 years of good life (since 30 years times 10 is 300 years). [38]

We can see then, that it is much more difficult to justify harm to an MRT via lesser-evil justification than it is to justify it on the Responsibility or complex accounts of liability. This is true at least where the multiplier is necessary. However, Bazargan still thinks that it is the lesser evil to kill the MRT rather than the totally innocent person and proposes the 'lesser-evil discounting view'.

The lesser-evil discounting view is explained as follows: "When determining whether there is a lesser-evil justification for imposing on an MRT a defensive harm greater than that to which she is liable, we ought to *discount* the disvalue of that harm relative to the weight of the harm that the threatener would otherwise impose on her potential victim".[39] That is, we ought to discount the badness of the harms to the more responsible party.

In order to understand why this is so, we are asked to "weigh these two exclusive options against each other: a) the badness of imposing harms on an MRT more severe than that to which she is liable, and b) [the badness of] allowing those harms to befall the MRT's non-responsible victim".[40] Since the MRT bears some moral responsibility for the situation that they both find themselves in, and her victim, by stipulation bears none, it would be totally unfair for the victim to bear the cost. It is the badness of the unfairness of the victim dying rather than the MRT that makes killing the MRT the lesser evil. Therefore, we ought to discount the lesser-evil—that is—

[38] Ibid.
[39] Ibid.
[40] Ibid.

forget about the injustice to the minimally responsible party for the sake of the totally innocent party. This relies of course, on the assumption that the one evil is greater to a significant enough degree as to be morally relevant. Bazargan sees the liable/non-liable distinction as filling this role. For him, "killing someone who is morally liable to sub-lethal harm is the lesser evil relative to the option of killing someone who is not morally liable to be harmed at all".[41]

To recap, Bazargan endorses two views in his paper: the complex account of liability to defensive harm, and the lesser-evil discounting view. These are combined into the pluralistic "hybrid account" in order to supplant the Responsibility account of liability. While the Responsibility and hybrid accounts often yield the same outcome—the permissible killing of the MRT—the hybrid account does a better job at taking into account just how minimally responsible the MRT is. Importantly, Bazargan accentuates the injustice involved in killing a minimally responsible (and completely non-culpable) threat. It is made clearer that killing an MRT in self-defense is not the killing in self-defense of a *bad* person. The MRT is non-malicious in her intent and therefore, there really are two potential evils in a scenario of kill-or-be-killed involving an MRT. The Responsibility Account can often be seen as indifferent to this fact.

Objection: The Weakness of Lesser-Evil Justifications

The main criticism I wish to level here is concerning the proposed strength of the lesser-evil justification for killing MRTs. Much emphasis, in this project, and in Bazargan's paper, has been put on just how minimal the moral asymmetry is between MRTs and those that they threaten. The question is whether the badness of killing someone who is wholly non-liable to any harm is significantly worse than killing someone who is minimally responsible so as to pass threshold 'R'

[41] Ibid., 18.

(the amount of badness that it must be possible to prevent in order to justify infringing on a person's right not to be killed). I argue that in many (if not all) cases, it is not.

To understand why I believe that the lesser-evil justification is not strong enough to meet threshold R, consider Bazargan's rationale for why we ought to apply a lesser evil justification. He quotes Dave Rodin, "It is implicit in a lesser evil justification that harm inflicted on an innocent bystander is a greater evil than an equivalent harm inflicted on a person liable to that harm".[42] At the same time though, he acknowledges that lesser-evil justifications must do a lot of work. Even where an MRT is made liable, due to the complex account, to 50% of the harm of death—in order to justify killing the MRT, that other 50% needs to be accounted for.

By invoking the trolley problem, we see that the doing/allowing distinction restricts the use of lesser-evil justifications so that to justify killing one person, five equally innocent people must be under a threat that can be prevented by that person's death. This is enlightening because, although our case does not involve a doing/allowing distinction, we get a sense of the difficulty involved in invoking lesser-evil justifications. That said, an MRT is not an equally innocent person. She is liable to 50% the harm of death to prevent that harm befalling an innocent. This is accounted for though, by halving the threshold number from 5 to 2.5.[43] Threshold R is multiplied by the percentage of threatened harm that the MRT is liable to since R represents the amount of harm that it must be possible to prevent in order to justify killing *one* person. Since we have already justified

[42] Bazargan, "Killing Minimally Responsible Threats", 18.

[43] In actuality, a person who is liable to 100% the harm that they threaten must still be threatening at least one person in order for us to justify killing them because of the 'avoidability argument'. This is the notion that we may not kill a person just because they are liable to harm. There must be a reason to harm for a justification to apply. This makes the baseline 1 rather than 0. So, if we can justify killing an innocent person if and only if it will save 5 innocent people under the banner of lesser-evil, this has 0% liability matched up with a threshold of 5. If we are to assume this scale moves proportionally, a 25% liable person's threshold would be reduced to 4, a 75% liable person's threshold would be 2, and in the middle, a 50% liable person's lesser-evil threshold would be 3 rather than 2.5 as I stated above for simplicity's sake. This does not necessarily change things significantly since varying intuitions about the trolley problem might place the initial threshold somewhere other than at 5. This is merely to point out the way in which a person's percentage liability influences the reduction of threshold R for an MRT.

through alternative means (liability) harm equal to half of killing a person, R is now equal to the amount of harm that it must be possible to prevent to justify *half* the harm of killing one person. This means that we are taking into account the liability to harm that the MRT has already accrued by being minimally responsible, and reducing the restrictiveness of what is needed for other justifications proportionally. We can imagine this as reducing the MRT to the status of non-responsible (although, she is by definition still responsible). The MRT has been punished for her responsibility, and has therefore "paid her debt" so-to-speak. Considering her still non-innocent after punishments have been applied is much like convicting a person for a crime, sending her to prison, having her do her time, and then still treating her as if she must be punished further. To do this would be unjust. The caveat of the punishment is that where we can prevent the equivalent badness of no less than the deaths of 2.5 innocents (rather than the normal 5), it would be considered a lesser evil for her to be killed instead.

Bazargan's reasons for discounting the badness of the harms to MRTs where killing them is necessary to save the lives of those that they threaten revolves around the fact that there is a serious injustice at play where the innocent person is killed instead of the MRT. Both she and the MRT are suffering more harm than that to which they are liable. This is what Bazargan calls the *absolute* (or intrapersonal) unfairness of the situation. The difference, Bazargan says, is that the defender suffers two different kinds of unfairness. The second unfairness is that "there is someone else who is *more* responsible… and who is getting away 'scot-free'".[44] This is the *comparative* (or interpersonal) unfairness. The question then is whether the comparative unfairness of the situation is significant enough to require rectification via lesser-evil justification. Surely the situation is

[44] Bazargan, "Killing Minimally responsible Threats," 19.

worse for the defender from a distributive justice standpoint, but there is no explanation from Bazargan as to how severe the badness of this distributive unfairness is.

This is a significant problem for Bazargan because, as I alluded to above, the badness of the harm to be prevented by killing the MRT must exceed threshold R (the amount of badness that needs to be prevented in order to make some action a lesser evil) *after* R has been multiplied by the percentage of threatened harm that the MRT is liable to. In this particular instance we have assumed that that badness must be equal to or exceeding the badness of the equivalent of the death of 2.5 innocent people. To be clear, this badness can come in any form. Bazargan argues, "More than numbers matter when assessing the lesser evil".[45] He means that some ways of bringing harm into being are worse than others, regardless of the quantity of the harm that comes about, and this can contribute to the badness of the harm.

For example, harm intentionally inflicted carries with it more badness than harm brought about as a foreseen but unintended side effect.[46] It might be considered worse that 1000 innocents are purposefully killed in order to stop a war, than if, in the bombing of a munitions factory, a bomber foreseeably but unintentionally kills 1000 innocents. Although the outcome is the same, there is a certain amount of badness created from the choice to intentionally kill. This is the intention/foresight distinction, and Bazargan sees it as relevantly similar to a liable/non-liable distinction.

Bazargan attempts to make the point here that "killing someone who is morally liable to sub-lethal harm is the lesser evil relative to the option of killing someone who is not morally liable to be harmed at all", [47] and surely it is a *kind* of lesser evil in that it is otherwise interpersonally

[45] Ibid., 18.
[46] Ibid.
[47] Ibid.

unfair. Bazargan is making the argument that the injustice of the unfairness here contributes significantly to the badness of the situation. In response to Bazargan, I argue that this unfairness has to be very bad to justify as the lesser evil contravening a person's right.

Certain final states of affairs are worse than others, but this does not mean that their prevention is always the lesser evil. In a trolley problem where there are two innocents under threat and we can kill one innocent to save them, the final state is arguably better if, all things being equal, we save the two people since one more person is allowed to live. But there are other considerations at work that deem this morally wrong. In the same way, there *is* an unfairness being prevented in killing the MRT, and this is, all things being equal, a better state of affairs.

However, we need to keep in mind that after the complex account of liability has come into play, justifying imposing a harm with the badness equivalent to paraplegia on the MRT, there is a significant differential of badness between paraplegia and death. What Bazargan is arguing is that the badness of paraplegia plus the badness of the unfairness to the potential victim are combined, greater than the badness of the death of the MRT. This is simply implausible.

The greatness of evils has to be measured in degrees. Killing one innocent to save five innocents appeals to a lesser-evil justification. But killing one to save ten is also appeals to this justification. The latter is clearly easier to justify than the former. In the same way, killing a minimally responsible person is a lesser evil than killing a non-responsible person, but killing a fully responsible person is a *much* lesser evil than killing a non-responsible person. The way that lesser-evil justifications are presented by Bazargan, we are made to believe that either some state of affairs would be justifiably brought about by appealing to lesser evils, or there is no state that would be a lesser-evil at all. It is not so obvious that this is the case. For a lesser-evil justification

to obtain, the harm to be prevented must be *much worse* than the harm inflicted. There is a difference between lesser-evils that can be justifiably acted upon, and those that we must let go.

Where responsibility is minimal, the difference between the two parties in terms of innocence is not enough to justify calling the killing of the MRT a lesser evil. Even though threshold R has been reduced by the MRT's liability, that threshold is still equivalent to a large amount of badness to be prevented. The badness that is prevented in preventing the interpersonal unfairness in this case is not as great as it that.

Say the MRT is liable to be harmed up to 50% the threat of death, and let us presume that this harm is a harm as bad as paraplegia. In killing the MRT, we impose on her the harm differential between paraplegia and death. As we have discussed, the lesser-evil justification badness threshold (R) must exceed badness of the deaths of 2.5 people (since an innocent's threshold is 5 and the MRT is 50% liable). This is, according to the above reasoning, the badness of imposing paraplegia on 5 innocent people (since paraplegia is presumed to be half as bad as death).[48] For Bazargan to claim that preventing the comparative unfairness of imposing death on an innocent person rather than a minimally responsible person is a badness to be prevented worthy of invoking a lesser-evil justification, he has to be ready to admit that that unfairness is *worse* than the badness of imposing paraplegia on 5 innocents. The moral asymmetry between the MRT and the innocent person is simply not that great. The comparative unfairness is bad, but not *that* bad.

This is not to say that the lesser-evil justification cannot do any work. Perhaps the unfairness is a badness on par with the badness of a broken arm. If we can prevent this unfairness

[48] The harm of death is actually, based on intuitions, probably many times worse than the threat of paraplegia, since people with paraplegia often go on to live happy lives. But this intuition that the threat of death is significantly worse only serves to bolster my claim. Killing a person who is not liable to be killed proves to be a very significant badness. But even the badness of imposing paraplegia on five innocents seems great enough to support my following claim.

from occurring by breaking the MRT's arm, we may have a lesser-evil justification for breaking the MRT's arm even though she is not liable to this harm, and we may discount the badness of this extra harm. But this still leaves us justifiably imposing paraplegia and a broken arm if it can prevent the death of the defender. But the Hybrid Account, as Bazargan presents it simply cannot justify killing an MRT.

Chapter 3.

In this chapter I want to explore a different, often rejected avenue for going about justifying defensive harm—that the harm is deserved. McMahan is one that would have discussions of desert remain entirely separate from discussions of self-defense. Liability and desert are clearly separate. In "The Basis of Moral Liability to Defensive Killing", McMahan starts the discussion by writing, "The claim that someone deserves to be killed implies that there is a reason to kill her even if it is possible for no one to be killed; but the claim that someone is liable to be killed has no such implication".[49] What's more, in *Killing in War*, McMahan further differentiates liability from desert by reference to the distinction between the instrumental and the non-instrumental. He writes:

> Desert is noninstrumental. If a person deserves to be harmed, there is a moral reason for harming him that is independent of the further consequences of harming him... Although a deserved harm is bad for the person who suffers it, it is, from an impersonal point of view, intrinsically *good*. By contrast, a person is liable to be harmed only if harming him will serve some further purpose—for example, if it will prevent him from unjustly harming someone… Harms to which people are liable are bad not only for those who suffer them but also from an impersonal point of view.[50]

This is definitely troubling from the perspective of someone who wants to argue that desert is the main source of justification for harming in self-defense. Liability and desert are separate concepts. One can be liable to be defensively harmed regardless of desert. What's more, one can be deserving of harm but not be liable to it.[51] But here I do not wish to make the claim that desert should replace liability. Rather, I will argue that desert may play a role in justifying self-defensive killing where simple liability (and lesser-evil justification) does not.

[49] McMaham, "Basis of Moral Liability," 386.
[50] McMahan, *Killing in War*, 8.
[51] Though McMahan does think that desert implies liability to punishment (McMahan, 2005).

In building this argument, I will refer to John Gardner and François Tanguay-Renaud's paper, "Desert and Avoidability in Self-Defense". In this paper, they argue that McMahan misunderstands desert, and that desert can have subsidiary roles in determining the proportionality of defensive actions. They do not go as far as I will in applying desert considerations to situations involving minimally responsible threateners, but I argue, this is the logical next step if the argument I made in Chapter 2 is accepted.

A Supplementary Justification for Killing in Self-Defense

Killing culpable threateners, under any of the accounts discussed in previous chapters is fully justified. Even where the Hybrid Account is skeptical about the ability for the Responsibility Account to justify the killing of MRTs, it has no problem with its ability to justify the killing of culpable threateners. They are simply liable to be killed—no further justification required. What it is that makes them liable, for all intents and purposes, is unimportant. The culpability account does not see the culpable threatener's responsibility as the determining factor, but rather his guilt due to blameworthiness for wrongdoing. He has exercised positive agency in doing something morally wrong.

The Responsibility account expands the scope of those that may be killed in self-defense. All parties responsible for a potential harm are liable to be killed in virtue of their responsibility. The culpable party's guilt is a manifestation of his degree of responsibility. The further fact that he is culpable is beside the point when it comes to whether he can be killed in a situation where it is kill-or-be-killed. What matters is that he is responsible.

Gardner and Tanguay-Renaud deny that this is the case. When it comes to killing culpable threateners, they maintain that it is not responsibility that is the important differentiating feature

between the defender and threatener, but rather, the culpable threatener's guilt, and the defender's innocence.

Gardner and Tanguay-Renaud claim that the difference between the defender and the threatener in instances of permissible self-defense is that "what is done to [the threatener] by [the defender], unlike what would otherwise be done to [the defender] by [the threatener], is *deserved*. For [the threatener] is the guilty (aka culpable, blameworthy, faulty) wrongdoer, and [the defender] is his innocent (aka nonculpable, blameless, faultless) victim".[52] This view attempts to account for "one aspect of the moral situation of [the self-defender], namely the existence of the moral asymmetry between her and [the threatener].[53] They call this view, self-deprecatingly, "the vulgar proposal" since it seems to track very simple, straightforward, and commonly-held intuitions about the proper distribution of harms. People are inclined to believe that it is a good state of affairs that good things happen to good people, and that bad things happen to bad people. Conversely, it tends to be thought of a as a bad state of affairs that bad things happen to good people and good things happen to bad people.

The vulgar proposal is formulated by Gardner and Tanguay-Renaud in terms of a norm of non-comparative justice, and its scoping rule. This is done because we might be inclined to express the proposal in the following way, Vulgar Proposal*: "It is morally permissible intentionally to inflict harm only on the guilty and in proportion to their guilt".[54] But in the interest of preserving express reference to desert, Gardner and Tanguay-Renaud insist on expressing it as follows:

[52] John Gardner and François Tanguay-Renaud, "Desert and Avoidability in Self-Defense," *Ethics*, Vol. 122, No. 1, Symposium on Jeff McMahan's Killing in War (October 2011), 114. My italics.
[53] Ibid., 116.
[54] Ibid., 120.

Norm: It is morally permissible intentionally to inflict harm only on those who deserve such an infliction and only to the extent that they deserve it.

Scoping Rule: Those who deserve such a harm are all and only guilty wrongdoers, and only to the extent that the infliction is proportionate to their guilty wrongdoings.[55]

There are two major objections that can be raised here. The view is broad in that it seems to justify too much, but also too narrow in that it does not fulfil the role that we expect of a theory of the moral permissibility of self-defense.

The first objection says that the vulgar proposal violates avoidability restrictions. Intuitively, killing a threat when it is avoidable, regardless of culpability, is unnecessary and not morally permissible. Imagine the following scenario: A hitman is coming to kill me. Unbeknownst to him, I caught on to his plan and removed the bullets from his gun when he wasn't looking. When he goes to shoot me, I know that he is already neutralized. If I know that I can easily restrain him at this point without killing him, it seems as if I am not morally permitted (that is, regardless of whether it is illegal) to kill him to prevent him from pulling the trigger. This is despite the fact that he might be deserving of harms due to his morally reprehensible actions. That killing him would be morally wrong is what Gardner and Tannguay-Renaud call McMahan's "avoidability argument".

The second objection says the view is too narrow in that it cannot provide an account of the moral permissibility of self-defense. Other views take into account moral restrictions on the infliction of harm as a side-effect of defensive measures. But the vulgar proposal does not concern itself with these. For example, "Perhaps D's self-defensive plan only involves crushing E's hand

[55] Ibid., 117.

in the elevator mechanism (depriving him of his hand); but D knows that it is highly likely that in the process E will be crushed to death".[56] The fact that E might be killed as a side-effect on other views makes it potentially impermissible since proportionality restrictions might be violated. The vulgar proposal's non-comparative norm does not take this into account. It says that we may intentionally inflict harm on those who deserve that harm. E may deserve the harm of his hand being crushed, and so it is permissible to crush his hand intentionally. But, D may also unintentionally kill E, a harm that E does not deserve. The vulgar proposal has no way of dealing with this.

But neither of these issues are actually a problem since the aim is not to defend desert as a replacement for other, more comprehensive views. Rather, desert is a consideration to take into account where there is a vacuum created by the lack of other justifications for killing. Desert takes no precedent over other more basic considerations like liability. Desert cannot be a consideration that takes precedent because oftentimes it does not qualify as a justification (since justifications are permissions with reasons. It may be the case that we have). Here we have to draw the line between permission to harm, and reasons for harming. Gardner and Tanguay-Renaud write, "Let's accept there is a moral reason to inflict deserved harm—namely, the fact that it is deserved—even when harm is avoidable. We can accept this without conceding that this reason makes the avoidable infliction of the deserved harm morally permissible".[57]

If one does have a reason, and it is not otherwise defeated by a countervailing consideration, then the action is justified. If we take the above hitman example again, we can say that I have a reason for harming him—desert-based punishment, even though it may serve no

[56] Ibid., 118.
[57] Ibid., 124.

greater purpose. However, there is a much stronger reason *not* to harm him. There are humanitarian considerations to take into account, (as well as other considerations such as maintaining the rule of law, or the desirability of not living in a vigilante state). Since, by stipulation, killing the hitman serves no purpose but vengeance (presumably), and it is to take the higher ground not to kill him, we may say the desert reason is defeated.

The vulgar proposal does not give us reasons for harming. It is merely permissive. It tells us why a threatener might be a morally suitable target for a defender's self-defensive actions but it does not supply the reasons for the defender to perform the action. What it does say is that if there is such a reason "and the reason is otherwise undefeated, then it is permissible to inflict the harm for that reason, so long as the infliction is deserved".[58] This is why the avoidability argument is not really a problem. It may be the case that a person deserves to be harmed, or killed, and the vulgar proposal provides a reason to do it. It would be the case that this person could be permissibly harmed, but the defeasibility of that desert reason makes it the case that there is no justification for the action.[59]

Self-defense might be considered a proper reason for engaging in an action where a person is deserving of harm. When a defender is faced with a threat from a threatener, she may ask herself, "do I have a reason to harm this person?", and her response might be, "yes, and the reason is a self-defensive one". If this is the case, she may then ask, "Am I morally permitted to do this?" that permission may be derived from liability, lesser-evil, desert, or some other justification.[60]

[58] Ibid., 125.
[59] McMahan, *Killing in War*, 43.
[60] This is not to make a positive claim about the existence of further justifications, but just that we should not rule them out.

Reasons go both ways though, and there may be reasons for not killing a threatener even where there is a self-defensive reason. Consider for example, a scenario where killing the threatener involves shooting him from a distance, leaving you safe. However he is wearing a heart-rate sensor that, if it falls to 0, sets off a bomb strapped to his chest, killing all those around him. If you let him kill you, he will leave everyone else alone, and retire from his career of killing. While there is a self-defensive reason for acting here, if there are say, at least 5 people at stake, then there is a lesser-evil justification for not killing the hitman. This lesser-evil justification defeats our self-defensive reasons for acting.

As with culpable threateners, responsibility-based liability may be completely sufficient to justify killing, even without the further fact that they may be deserving of punishment. This background fact may be there, but the fact that the threatener is deserving of being killed overdetermines the defender's permissibility since liability was already sufficient.

However, we can see from chapter 2, responsibility may not be enough to justify killing when the threat is only minimally responsible. Moreover, the further lesser-evil justification is not strong enough either. But what if the background fact that the threatener is deserving of harm is also present in an MRT case? This is not to suggest that MRTs might be deserving of harm due to their responsibility for the situation. Rather, the moral character of both the MRT *and* the innocent victim might be taken into account in justifying the killing of the MRT or in determining that the MRT may not be killed in self-defense.

Consider the case of *Nazi Lawnmower*:

> A person keeps her car well maintained and always drives cautiously and alertly. On one occasion, however, freak circumstances cause the car to go out of control. It has veered in the direction of a man who is casually mowing his lawn in full Nazi SS uniform. He will

> be killed unless he blows up the car with his prized anti-tank weapon that he proudly carries around with him to remind him of the war, and all the atrocities he proudly committed.

In this case, as in *Conscientious Driver*, the defender has done nothing to set in motion the events leading to his being threatened. Both the Nazi and the Driver have done nothing to forfeit their rights not to be killed in this particular situation. The Nazi is not liable to defensive harm since he is in no way responsible for the events that are transpiring. As before, the Driver in this case is an MRT and liable to some portion of harm, but she is not liable to be killed in self-defense. Even if we accepted Bazargan's claim that it would be a lesser evil to kill the Driver rather than the potential victim, we might be made to feel as if this further justification is neutralized by the status of the victim as a war criminal. It feels wrong that a person's minimal responsibility for a situation might justify her being killed rather than a Nazi war criminal who has managed to avoid legal sanction. Where Bazargan appeals to the badness of the unfairness of the non-responsible victim bearing the full harm while the MRT gets off 'scot free', this unfairness pales in comparison to the badness of the Nazi getting off 'scot free' while the Driver bears the full harm.

Assuming though that my criticism of Bazargan in the previous chapter is correct and that while the lesser-evil justification may apply, it is too weak to justify killing the Driver (even when her victim would not be a vicious Nazi), let us switch the roles. The situation is the same as *Conscientious Driver*, but the Driver has the moral character of the Nazi in *Nazi Lawnmower*. In this situation, it seems as if we have a good reason to favour the pedestrian over the driver above and entirely separated from the fact that the Nazi driver is minimally responsible. It seems intuitively easier to justify the pedestrian's killing Driver since the Driver's moral character is so abhorrent. She is only *liable* to a portion of the threatened harm because liability only tracks responsibility, but there is justification for her to be killed because there is a reason for acting—

self-defense—and, along with being liable to a portion of harm, she is deserving of further harm. It does not have to be the case that she deserves to die, and, depending on where one stands, we might not think that *anybody* deserves to die. But it is likely not that controversial to think the Nazi is deserving of at least some degree of severe harm. And since the driver is already liable to some significant sub-lethal harm (and with Bazargan's lesser-evil justification we might be able to tack on a bit more), a further justification for significant harm on separate grounds could justify killing this minimally responsible threatener.

The Implications of Culpable Acts

Gerhard Øverland supports this view in his paper, "Moral Taint: On the Transfer of the Implications of Moral Culpability". Rather than culpability tracking causal impact for the harm at hand, as the Responsibility and Hybrid accounts do, he proposes that a person's moral culpability to a particular victim should transfer to other conflict situations that the wrongdoer might find himself in.[61] This is desert in a wider sense than what McMahan might have in mind in his denial of desert's value in self-defensive situations. The claim is not being made that an MRT is deserving (or not deserving) of harm because of her responsibility for the situation. Rather, it is desert in a wider context that comes into play. Consider the following example that Øverland uses,

> *Pool*: Bill culpably kills Alice, but has no intention of killing anybody else and there is no reason to expect that he will. A bit later Bill and Cathy are about to drown through no fault of their own. There is only time to save one.[62]

It seems obvious that Bill and Cathy should not be given equal consideration. But Bill has nothing to do with the threat to Cathy's life. If it is possible to save both of them, we seem to have a duty to do this, but where it is only possible to save one, Cathy is the obvious choice. What seems clear

[61] Gerhard Øverland, "Moral Taint: On the Transfer of the Implications of Moral Culpability," *Journal of Applied Philosophy*, 28(2) (2011): 122-136, 122.
[62] Ibid.

is that although Bill's death is bad, it is not nearly as bad as Cathy's death because of Bill's status as a vicious person.

Øverland makes a further assertion that "time might erase the significance of moral wrongdoing" and that "the implications of moral culpability for harm should transfer to situations occurring more or less immediately thereafter and to cases in which the wrongdoer expresses neither regret nor repentance".[63] While I agree with the view that proper regret and/or repentance might absolve a person from the normal implications of past culpability (namely lowered considerations for their wellbeing in moral dilemmas), I am less convinced by the significance of time where regret or repentance are not present. *Nazi Lawnmower* seems to be a case where time is not of much consequence, although the previous level of culpability (from the committing of war crimes) is possibly significant enough to outweigh those effects. In Øverland's defense, we could imagine a case where Bill finds himself to be a pedestrian in *Conscientious Driver*. His culpability for a single crime twenty years prior could plausibly not make a difference in whether he should be killed rather than an MRT.

Let us consider another example of Øverland's:

> *Window:* Bill has culpably killed Alice. David is about to shoot Cathy from a narrow basement window. Cathy cannot shoot David but can kill Bill, thereby saving herself from David by causing Bill to slump to the ground in front of the window.[64]

Here, Bill's culpability, like in *Pool*, weakens his claim against being killed for the achievement of some good. This is not to say that he is liable. He has not forfeited his right against Cathy not to be killed. But, this is a justification akin to a lesser-evil justification. And, in the same way that Bazargan proposes that that we discount the impersonal badness of the situation in which the MRT

[63] Ibid., 123.
[64] Ibid., 127.

is harmed rather than an innocent victim, we ought to discount the badness of the situations in which those with vicious moral characters are harmed rather than innocents. This view lines up with our intuitions about the badness of harms to vicious people. While Bill is still wronged by being used to defend Cathy against David, this outcome is justified because the badness of this rights-infringement is diminished.

Øverland proposes that "the additional means that may be used against a culpable threatener X to protect a prospective victim,[65] as compared to the means that may be used against an innocent contributor for the same purpose,[66] may also be used against the person X in order to save someone other than his or her victim".[67] Øverland sees the implications of culpability as transferable. He writes, "If it were permissible to impose 10 cost-units on an innocent contributor and 30 on a culpable contributor to prevent a particular harm, 20 units would be transferable to a new situation where this wrongdoer needs to suffer a similar harm".[68] Here Øverland is making an assertion that I find to be intuitively plausible. The identity of the person a culpable party is harmed to save does not seem to matter.

In the present example, Bill was liable to be harmed 30 units for culpably killing Alice, and therefore, there is, in a sense, a 'debt' to be paid. Cathy imposes a harm on Bill, a person that is not presently threatening her, and therefore, the units that we could permissibly impose on an *innocent contributor* are removed from Bill's debt. The harm that Cathy imposes on Bill by killing him clears his ledger. But imagine instead that Cathy could have caused Bill's body to slump in

[65] This is, the harm they threaten, multiplied by 'M'—the amount of times over we may permissibly harm culpable threateners more than the harm they threaten.
[66] This is a harm that the victim might be morally required to share if possible.
[67] Ibid., 128.
[68] Ibid.

front of the window by merely injuring him. In this case, Bill still 'owes' 20 units, and therefore, we may impose those 20 units on Bill to prevent harms to another non-culpable person.

This reasoning may sound somewhat odd and even esoteric but it can be defended by thinking of culpable harms as not victim-sensitive. Who the victims are, if they are innocents, seems unimportant to the moral status of the action and the perpetrator. The act is wrong regardless. This is antithetical to McMahan's view of culpability as being derived from responsibility for violating the right of a particular individual, but it is consistent with Gardner and Tanguay-Renaud's suggestion that the grounds for harming culpable threateners and MRTs may in fact, be different. Culpable acts, including negligent ones are in a different class from the acts of MRTs because of a culpable person's failure of a duty to respect the rights of others. Negligent acts are a good example of the victim-insensitive nature of culpable threats. A negligent, though not malicious act generally has no particular person as a target. The victims could presumably be anybody, and therefore, their identity is inconsequential. It makes sense then, to think that a culpable person has some duty to subsequent innocents who might be saved at his expense, as long as the implications of his culpability have not been expunged by time, repentance, or some other redeeming factor.

That a person retains implications of culpability, again, as explained earlier in the discussion of the vulgar proposal, is not a justification for killing this person in any circumstance whatsoever. A good reason for harming is required for a justification. It may be the case that we are permitted to take punitive actions against Bill, since he has culpably killed Alice. But we have good reasons not to allow just anybody to exact upon Bill the outstanding deserved harm that remains from his getting away with murder. There are good reasons against vigilante justice, namely that it is contrary to the bounds of our system of justice. But this reason that normally holds

against regular people harming Bill is outweighed by a countervailing reason when that reason is self- or other-defensive. What is normally just a restricted permission becomes a justification for Bill's harm.

Conclusions

The discounting of the badness of harms to those who retain the implications of previous culpable actions is, in effect, an account of a role for desert in self-defense. It is a kind of justification for harming. It may not be enough to justify killing a person, but in combination with a complex account of liability, and other justifications, this discounting view can "put a thumb on the scale" in determining what party might justifiably be killed in a self- or other-defensive situation involving MRTs.

In this project I have argued that McMahan's assertions about desert as being irrelevant in self- (and other-) defense are false. In Chapter 1, I argued that McMahan's Responsibility Account does not properly account for the minimal responsibility of MRTs. In response, I turned to Bazargan's hybrid account as a plausible response to the Responsibility Account. This view accounted for minimal responsibility of MRTs, and was not as overly punitive as McMahan's account. While I see Bazargan's complex account of liability as plausible, I remain unconvinced that combining it with the lesser-evil discounting view has the strength to justify the killing of MRTs. Instead, I proposed in chapter 3 that desert may be a deciding factor in whether or not an MRT can be killed in self-defense. Given that the MRT is only minimally responsible, the moral asymmetry between parties cannot justify either person's death. However, if either party retains implications of culpability from previous actions, this justification, along with liability considerations, can provide good reason to kill the MRT, or to allow the victim to be killed.

If it is the case either that an MRT is not liable to be killed, or that there is no desert-based justification for killing her, then we may have to accept that it may not be permissible to kill either party. I maintain that a third-party does not have enough reason to come to the lethal other-defense of the potential victim of an MRT if there is not sufficient secondary justification for harming that,

combined with liability, gives permission to violate the right of the MRT not to be killed. Rather, the third-party may be morally required to stay out of things. Both the MRT and her victim retain their right not to be killed.

One might be tempted to wonder then, what either party *can* permissibly do in one of these kill-or-be-killed situations involving an MRT. Without committing to any positive claim here, I would suspect that this becomes a situation of excused symmetrical self-defense. The potential innocent victim is not morally permitted to kill the MRT, but she does have a fully exculpating excuse for doing so. Harming would certainly be permitted, and given the opportunity, I don't presume to think that any reasonable person would allow herself to be killed. But this is also the case for the MRT. It seems as though there is no morally permissible solution to the situation, as hard as McMahan and Bazargan may try to find one. Sometimes people are unfairly thrust into situations where their choices include engaging in a morally impermissible act, or allowing themselves to be killed.[69]

[69] For an interesting view on this topic, see Jonathan Quong, "Liability to Defensive Harm," *Philosophy & Public Affairs* 40(1) (2012), 45-77.

Bibliography

Bazargan, Saba. forthcoming. "Killing Minimally Responsible Threats." *Ethics.*

Foot, Phillipa. 1967. "The Problem of Abortion and the Doctrine of Double Effect." *The Oxford Review* (5): 5-15.

Gardner, John, and François Tanguay-Renaud. 2011. "Desert and Avoidability in Self-Defense." *Ethics* 122 (No. 1, Symposium on Jeff McMahan's Killing in War): 111-134.

Kagan, Shelly. 2005. "The Geometry of Desert." *The Lindley Lecture.* The University of Kansas.

Kagan, Shelly. 1992. "The Structure of Normative Ethics." *Philosophical Perspectives* 6: 223-242.

Lippert-Rasmussen, Kasper. 2001. "Egalitarianism, Option Luck, and Responsibility." *Ethics* 111 (3): 548-579.

McMahan, Jeff. 2011. "Duty, Obedience, Desert, and Proportionality in War: A Response." *Ethics* 122 (1): 135-167.

—. 2009. *Killing in War.* Oxford: Oxford University Press.

McMahan, Jeff. 1994. "Self-Defense and the Problem of the Innocent Attacker." *Ethics* 104 (2): 252-290.

McMahan, Jeff. 2006. "The Basis of Moral Liability to Defensive Killing." In *Philosophical Issues, I5, Normativity*, edited by Ernest Sosa and Enrique Villanueva, 386-405. Boston: Wiley-Blackwell.

McMahan, Jeff. 2004. "The Ethics of Killing in War." *Ethics* 114 (4): 693-733.

Otsuka, Michael. 1994. "Killing the Innocent in Self-Defense." *Philosophy and Public Affairs* 23 (1): 74-94.

—. 2013. "The Moral Responsibility Account of Liability to Defensive Killing." *Bowling Green State University Workshop in Ethics and Public Policy.* Bowling Green, Ohio.

Øverland, Gerhard. 2011. "Moral Taint: On the Transfer of the Implications of Moral Culpability." *Journal of Applied Philosophy* 28 (2): 122-136.

Parfit, Derek. 2011. *On What Matters.* Vol. 1. Oxford: Oxford University Press.

Quong, Jonathan. 2009. "Killing in Self-Defense." *Ethics* 119 (3): 507-537.

Quong, Jonathan. 2012. "Liability to Defensive Harm." *Philosophy & Public Affairs* 40 (1): 45-77.