

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Imperfect Victims and The Right to Self-Defence

Haeg, Jonas

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Imperfect Victims and The Right to Self-Defence

Jonas Haeg

A thesis submitted for the degree of

Doctor of Philosophy (PhD)

Department of Philosophy,

Faculty of Arts and Humanities,

King's College London

Acknowledgements

First of all, I would like to thank my supervisors. Sarah Fine's enthusiasm and support during my BA made it an easy decision to return to King's for my PhD. Her continued help and encouragement during the project have been invaluable. Massimo Renzo has always pushed me to improve my arguments and sharpen my thoughts. He has also been able to link my research to many other areas in philosophy I was unaware of and could learn a lot from. David Owens agreed to take over as a supervisor when Sarah moved to Cambridge, and I have benefitted greatly from our meetings. Especially his questioning of the many underlying assumptions popular in the literature on defensive harming has been eye-opening. I am grateful to all of them. The thesis has benefitted tremendously from their different perspectives and expertise.

I have also been fortunate to present my work at various conferences and workshops around the world, including U.C. Berkeley, Stockholm University, University of Reading, University of Oxford, University of Colorado Boulder, University of Manchester, the Institute of Philosophy in Zagreb, University of St. Andrews, the Philosophy, Politics and Law Graduate Forum, the Philosophy, Ethics and Killing workshop, and the London Graduate Workshop on Moral and Political Philosophy. And, of course, several presentations and discussions in the Advanced Research Seminar at King's. I am incredibly grateful to all the event participants, especially those who provided detailed written and/or verbal feedback on my work there, including Cecile Fabre, Connor Kianpour, Emily Podhorcer, Helen Frowe, Kim Ferzan, Lisa Hecht, and Penny Oderberg.

Special thanks go to Gunnar Björnsson and Theron Pummer. First, during an extended period in Stockholm, while I was finalising the thesis, Gunnar helped set me up with an office space and invited me to present my work at several seminars at Stockholm University. My time there was excellent! Second, after Theron commented on my paper at the CEPPA Graduate Conference, we continued to stay in touch. He ended up reading drafts of more than half of my thesis. I am so thankful for his excellent feedback and encouragement.

I am also lucky to have such a supportive family back home in Norway. When I left Oslo at 19 years old to study in London, I do not think any of us expected I would stay for a decade and end up writing a PhD thesis in moral philosophy. But I would not have been able to do any of it without their support. One highlight of going back home has also been to visit

and discuss philosophy with my wonderful friend Jon H. Marthinsen. Our many long discussions and disagreements have helped make philosophy fun.

Finally, my biggest debt of gratitude is owed to Romy Eskens. There is no one I would rather discuss philosophy with than her, and for years I have been so fortunate to get her thoughts on everything from half-baked thought experiments to nearly completed chapters. Throughout the many ups and downs while writing the thesis, Romy has always encouraged me, pushed me to continue, and reminded me to relax and reflect on my achievements. I doubt there would be a thesis without her.

Abstract

This thesis investigates the right to self-defence by examining a range of cases involving “imperfect” victims – that is, victims of wrongful threats who contributed to, or could have avoided, being wrongfully attacked. It asks when, if at all, this fact about a victim should impact or compromise their right to defensively harm the aggressor. Despite the booming literature on the ethics of defensive harming, this question has received little attention compared to questions about how the fault and responsibility of the aggressor affect the right to self-defence. It is often ignored or treated as an afterthought. But it merits more attention. There are cases in which it seems intuitive that a victim’s actions negatively affect their right to self-defence – for instance, cases involving provocation – but unless we are careful in spelling out the reasons for this, we risk ending up with overly harsh and “victim blaming” theories of self-defence. Moreover, partly because victims have received little attention, some of the ways that the principles and conditions for permissible self-defence are phrased in the literature also imply too harsh restrictions on the right to self-defence when applied to cases involving imperfect victims.

The thesis aims to help fill this gap. One of the main themes that emerge is that we should not simply focus on whether a victim could have (even easily and without much cost) avoided, or avoided contributing to, a threat situation, nor simply on whether they were culpable for not avoiding this. Instead, we should focus on whether their behaviour violated specific duties to avoid contributing to situations in which they need to impose defensive harm. As I explain, there are different possible grounds for this type of duty, which impacts the duty’s stringency. Understanding these different grounds helps us see more clearly when and to what extent an imperfect victim’s right to self-defence is compromised. From this approach, we get a unified and nuanced view according to which it is generally difficult to compromise one’s defensive rights. Taking risks, even culpable risks, has little or no compromising impact on one’s right to self-defence against culpable aggressors. However, wrongfully interfering with others’ fair opportunities to avoid defensive harm, or manipulating them into attacking, can radically alter one’s right to self-defence.

Table of Contents

1	Introduction.....	7
1	The Ethics of Self-Defence.....	7
2	The Imperfect Victim.....	12
3	The Right to Self-Defence	17
4	Compromised Defensive Rights	23
5	Fairness and Responsibility	26
6	Scope, Assumptions, and Method.....	28
7	Overview of Chapters	31
2	What Do We (Not) Owe to Aggressors?	35
1	Introduction.....	35
2	Conditional Threats and Escalations.....	40
3	Necessity and Proportionality	42
4	The Submission Argument	46
5	Responsibility	51
6	Denying One's Own Agency	54
7	Intervening Agency.....	57
8	Clarifications and Implications	63
9	Conclusion	65
3	A Duty to Avoid Trouble and Its Limits.....	66
1	Introduction.....	66
2	Unnecessary Harm	67
3	The Draconian View	75
3.1	Continuity and Breach	77
3.2	Responsibility and Rights	82
3.3	Value and Initial Costs.....	84
3.4	Culpability.....	85
4	Objections to the Duty to Avoid Trouble	87
4.1	Costs Over Time	90
4.2	Rights and Dignitary Interests	93
4.3	Degradation.....	96
4.4	Socially Unjust Threats.....	99
5	Conclusion	102
4	The Irrelevance of Risk-Taking.....	104
1	Introduction.....	104
2	The Less Wrong Claim	106
3	The Proportionality Claim	110
4	Luck Egalitarian Worries.....	113
5	The Easier to Justify Claim.....	117
6	The Challenge	119
7	Some Suggestions	123
7.1	Luck Egalitarianism Revisited.....	123
7.2	Cost Imposition.....	124
7.3	Unequal Weighing but Equal Rights	126

8	Holism about Risk-Taking	128
9	An Objection	135
10	Conclusion	137
5	Provocation and Duties to Protect from Harm	139
1	Introduction	139
2	Two Preliminaries	140
2.1	The Concept of Provocation	140
2.2	Compromised Defensive Rights	144
3	Excuses, Desert, and Honour	146
4	The Partial Responsibility View	149
5	The Protective Duty View	154
5.1	Duties to Protect from Harm	155
5.2	Fair Opportunities to Avoid Harm	157
5.3	The Moderate View	162
6	Conclusion	166
6	Defensive Entrapment and Manipulation	167
1	Introduction	167
2	Alexander's Challenge	169
3	The Entrapment Solution	172
4	A Theory of Entrapment	174
4.1	Existing Views	174
4.2	The Manipulation Account: Entrapment	180
4.3	The Manipulation Account: Punishment	186
4.4	Objections	192
5	Defensive Entrapment	198
6	Conclusion	203
7	Conclusion	204
1	Summing Up	204
2	Looking Ahead	205
	Bibliography	207

1

Introduction

1 The Ethics of Self-Defence

Here is what seems like a clear-cut example of a situation in which someone has the right to impose serious harm on someone in order to defend themselves against a threat (I will call this “defensive harm” for short):

Perfect Victim: Brian is asleep at home when Alexis smashes his bedroom window, jumps in, and tries to beat him up because she hates him. To defend himself, Brian must seriously harm Alexis.

It is possible to be a pacifist concerning killing and harming more generally, but I set that possibility to the side here.¹ Instead, I follow the majority of philosophers working on the ethics of self-defence in holding that it can be permissible to harm, indeed kill, in self-defence (and other-defence, for that matter). *Perfect Victim* seems to be a paradigmatic example of a scenario in which this is the case. Seriously injuring Alexis in self-defence is permissible. Even defensively killing her might be permissible if the potential beating is severe enough.

However, suppose that we start adding specific details to Brian’s backstory. First, we learn that (i) Alexis only got access to his property because Brian did not bother to close the gate. Then we learn that (ii) he failed to do so despite being warned about a potentially dangerous person roaming nearby. Furthermore, we learn that (iii) Brian had earlier angered Alexis by yelling disparaging and mocking remarks at her from his balcony. Lastly, we learn that (iv)

¹ However, for a recent argument in favour of a very strong pacifist position, see Hereth (2017).

Brian did all of this (mocked her and failed to lock his gate) because he hoped Alexis would show up and attack him so that he could defensively harm her.

These facts make Brian appear less like the perfect victim he initially seemed to be – sound asleep and completely unaware of the attack until it occurred – and more like an imperfect one. The question I am interested in here is whether any of these details matter morally. We can ask that question in a variety of ways. Do they matter for how *blameworthy* Alexis is for attacking Brian? Do they matter for how harshly we should *punish* Alexis? Do they matter for how much *sympathy* we should have for Brian? Do they matter for how *blameworthy* Brian is for his injuries? Do they matter for Brian’s claim to *compensation* from Alexis and possibly the state? My focus in this thesis, however, is on the defensive question. Do facts like (i) to (iv) (and possibly other similar facts) affect Brian’s *right to defend himself against Alexis*?

There are two reasons for focusing on this particular question. First, the literature on the ethics of self-defence has been highly influential. It has given rise to a framework for approaching many other ethical questions. For instance, many have used insights from thinking about self-defence to shape the literature on punishment and just war theory.² Key concepts from this literature – such as “liability”, “proportionality”, and “necessity”, which I come back to later – have also been applied fruitfully to a wide variety of other issues, such as (online) public shaming,³ free speech suppression,⁴ rioting and uncivil disobedience,⁵ immigration and emigration restrictions,⁶ sting operations,⁷ and mandatory vaccinations.⁸ Thus, focusing on

² For the former, see, for instance, Quinn (1985), Farrell (1985; 1990; 1995), Montague (1995), and Tadros (2011). For the latter, see, in particular, McMahan (1994; 2004a; 2004b; 2009) but also Rodin (2002), Fabre (2012), and Frowe (2014), as well as numerous edited volumes, such as Fabre and Lazar (2014), Frowe and Lang (2014), and Strawser, Jenkins, and Robillard (2017).

³ See Billingham and Parr (2020a; 2020b).

⁴ See Howard (2019a).

⁵ See Pasternak (2019).

⁶ See Hidalgo (2015; 2021) and Oberman (2012), respectively.

⁷ See Nathan (2017).

⁸ See Flanigan (2014).

whether (and if so, how) Brian's behaviour leading up to the threat in *Perfect Victim* affects his right to self-defence also has the potential to impact many discussions to which insights from theories of defensive harming have been applied.

The second and more important reason for focusing on the defensive harming context is that the question of how, if at all, a victim's behaviour leading up to an attack can affect the morality of self-defence is under-discussed and often glossed over in the literature. Although the ethics of self-defence literature has been booming over the last few decades, it has undeniably been heavily *attacker-focused*.⁹ One possible reason for this is that people have been wary of victim blaming. I respond to that worry below. The more manifest reason for this trend, however, is that, over the last few of decades, a popular view on what makes it permissible to seriously harm or kill an aggressor in a case like *Perfect Victim* is that Alexis has made herself *liable* to be defensively harmed. This, in turn, is often understood as the claim that she has *forfeited* her right not to be harmed or killed (for this defensive purpose, at least) and that this is why it is permissible for Brian to harm or kill Alexis to save himself.¹⁰

A lot of the work in this literature has therefore been focused on aggressors and on what is necessary and sufficient for them to forfeit their rights against defensive harm. Almost everyone agrees it is necessary that the agent is causally related to the threat in question in some way, but is anything else necessary?¹¹ Consider the following variants of the original case:

Perfect Victim (Causation): As in *Perfect Victim*, but Alexis is not operating as an agent. She is unconscious, and an evil villain controls her body.

⁹ I borrow the term from Lang (2017, p. 249).

¹⁰ As far as I know, the term liability to defensive harming comes from McMahan (2005a, p. 386) but is now ubiquitous in the literature. The idea that liability involves *forfeiture* of rights is also very popular, though some prefer to understand rights as already specified in a way that gives way to liability (see, for instance, Uniacke (1994) and Draper (2016a)). For more discussions and some critiques of the notions of liability and forfeiture, see, for instance, Tadros (2012), Lang (2014; 2017), Ferzan (2016), Dempsey (2016), and Renzo (2017).

¹¹ Though some, like Christie (2020), argue that being causally related to a threat should not even be necessary for being liable to defensive harming.

Perfect Victim (Responsibility): As in *Perfect Victim*, but Alexis has received convincing evidence that Brian is plotting to kill her and that the least harmful way of preventing this is to scare him by breaking into his house and beating him up now.

One view is that all that is necessary to forfeit one's right against defensive harm is that one is causally responsible for, or (at least) casually posing, a threat to an innocent person. Thus, that Alexis or her body is posing a wrongful threat to Brian suffices to make her liable.¹² She is therefore liable to defensive harm in all three cases considered so far. Others argue that Alexis cannot be liable in *Perfect Victim (Causation)* because she does not operate as an agent and therefore cannot forfeit any moral rights either. Instead, she is liable only in the original case and *Perfect Victim (Responsibility)* because she is *morally* responsible for those threats to Brian. In both cases, she voluntarily tries to harm Brian. Although her decision is not culpable in the latter case, she ought to have known there was a risk that Brian was innocent and, so, that she would be posing an unjust threat.¹³

Others again think that Alexis is liable only in the original case, for only in that case is she *culpably* responsible for the threat to Brian.¹⁴ That one is morally responsible for a wrongful threat is insufficient to make one liable. One must also have acted in some morally blameworthy manner. Still, on this view, Alexis, in *Perfect Victim (Responsibility)*, could be liable if (say) she had a chance to double-check the evidence and chose not to do so.

There are also other views on what grounds liability to defensive harm.¹⁵ But the main point here is to show that a large part of the literature on defensive harming has been devoted to this attacker-focused liability question. In pursuing that question, it may seem sensible not

¹² This view, in some version, is often attributed to Thomson (1991) though she does not use the term "liability". See also Hossein (2017) and Leverick (2006).

¹³ For defences of the moral responsibility view, see, for instance, McMahan (2005a; 2005b), Otsuka (1994; 2016), Draper (2016a), and Gordon-Solmon (2018).

¹⁴ For a defence of this type of view, see, for instance, Montague (1981; 2010) and Ferzan (2005; 2012).

¹⁵ See, for instance, Quong (2012; 2020) and Tadros (2011).

to focus so much on cases involving victims with complicated backstories and facts such as (i) to (iv). Excluding such facts, and focusing on thought experiments only involving complicating facts relating to the aggressor, makes it easier to hone in on precisely what it is about the *aggressor* and their relation to the threat that makes them liable to defensive harming.

As suggested, however, the downside of this approach is that it is easy to lose sight of the possibility that facts about the victim's behaviour can also affect the morality of defensive harming. It is thus unsurprising that the very attacker-focused literature on the ethics of self-defence has so far had comparatively little to say about this topic.¹⁶ This is made even more unfortunate by two facts. First, it seems to me very intuitive that in at least some of the variations of *Perfect Victim*, such as the one in which Brian has goaded Alexis in an effort to create an opportunity for defensively harming her, Brian has compromised or perhaps even lost his right to harmfully defend against Alexis. Thus, we need an account of why his behaviour has that effect. Second, few real-world victims seem as "perfect" as Brian is in the original case. Often, the threat that a victim faces will be partly *self-generated*.¹⁷ For instance, it will partly be the result of taking a more dangerous than necessary way home, failing to avoid or retreat from a dangerous environment, a confrontation gone wrong, and so on. On the face of it, it is counterintuitive that one should compromise one's defensive rights in all such cases. Together with the first intuition, this increases the pressure on us to think more and harder about exactly what kinds of behaviour from a victim should and should not affect their right to defend themselves and why.

The aim of this thesis is to help fill this gap in the literature by doing that – consider a variety of ways in which a victim may be considered imperfect and look at which features of

¹⁶ As a further illustration of this trend, consider Rodin's (2011) seminal article, "Justifying Harm", which provides a general framework for what is relevant to justifying harm. Roughly ten pages are devoted to the aggressor-centred factors relevant for liability (p. 82-91), while less than half a page is devoted to the relevance of the victim's behaviour (p. 95).

¹⁷ I borrow the term from Leverick (2006, p. 109).

their behaviour and contribution to the threat situation may, or may not, affect their defensive rights. I am not suggesting that nothing has been said about this issue in the past. It has been touched on occasionally, but those discussions are often brief.¹⁸ There are also some longer exceptions to this.¹⁹ There have also been more written on issues which are closely related to imperfect victims.²⁰ I will draw on and engage with many of those resources throughout the thesis. Nevertheless, a systematic and nuanced treatment of the various ways victims can affect the ethics of self-defence is so far lacking.

First, I begin by clarifying some assumptions and views I will be working with: (1) What is an imperfect victim? (2) What is the broader theory of self-defence I am working within? (3) What does it mean to say that one's right to self-defence has been compromised? I address those questions in the following three sections. Then, I consider a possible critique of my project here. Lastly, I go over some more general assumptions about philosophical methods and ethical frameworks that I employ, and provide an overview of the chapters to come.

2 The Imperfect Victim

One initial worry, briefly flagged above, concerning the project is that using terms such as “imperfect victim” and focusing on victims’ behaviour leading up to an attack involve a form of victim blaming. For instance, Martha Gill uses the term “imperfect victim” in the context of the #MeToo movement to highlight derisive description of victims of sexual assault and misconduct “who were wearing the wrong thing, or were drunk, or were promiscuous, or loved

¹⁸ For instance, Rodin (2002, p. 79; 2011, p. 95), Frowe (2014, p. 134) and Hurd (2016, pp. 271-70) all have a couple of paragraphs on “provocation”. Draper also often devotes a few pages to “provocation” and “responsible victims”. See, for instance, Draper (1993, p. 77; 2009, p. 75-6; 2016a, pp. 76-8; 2018, pp. 92-5).

¹⁹ See, for instance, the issue of “culpably causing the conditions of one’s own defence” which is broader than the issue of defensive harm but is often discussed in relation to that. Relevant papers here include Robinson (1985), Raymond (2010), Alexander (2013a; 2021), Ferzan (2013), Farrell (2013), and Hecht (2019a).

²⁰ Especially on issues such as “stand your ground”, “duty to retreat”, and “conditional threats”, both in individual self-defence and war scenarios. See, for instance, Alexander (1999; 2013a; 2013b), Leverick (2006), Øverland, (2010), Lazar (2014), Rodin (2002; 2014a), Finlay, (2015), Hurd (2016), Ferzan (2019), Steinhoff (2013), and Lang (2022).

their perpetrator, or had previously broken the law, or had lied before, or had a bad character [...]” (2022). Similarly, criminologist Nils Christie coined the related terms “ideal victim” and “non-ideal victim” to highlight and criticise how people’s attitudes towards actual victims, and readiness to give them “the complete and legitimate status of being a victim” (1986, p.18), tend to be shaped by stereotypes about victims as being weak, vulnerable, completely faultless, and so on. The less an actual victim fits the ideal, the less likely we are to offer them sympathy and compassion.

The worry that the project here thus inherently involves a kind of victim blaming is not without precedence either. From the 1940s and onwards, some criminologists (called “victimologists”) began paying increasing attention to the role of the victim in creating crimes. Several of them created topologies of different types of victims, often distinguished by the level of fault or responsibility they were perceived to have had for the crime occurring. For instance, Benjamin Mendelsohn (1976) would categorise victims according to their degree of “guilt” for the crime as compared to the criminal’s guilt. Similarly, Stephen Schafer (1968) classified victims according to degree of responsibility and used terms such as “provocative victims” and “precipitative victims”. However, one repeated criticism of this trend in criminology has precisely been that it engages in, or legitimises, *victim blaming*.²¹ After all, the terms and classifications used have strong blame-related connotations. Some early victimologists also directly contributed to this worry with their own words. For instance, Menachem Amir’s (1971) asserted that a sizeable portion of rapes was “victim precipitated” because they occurred after the woman had engaged in behaviour that “could be taken as an invitation to sexual relations” (1971, p. 266).

However, my use of the term “imperfect victim” here is only meant to pick out those who do not fit the stereotype of the victim attacked completely out of the blue. It does not *entail*

²¹ See, for instance, Eigenberg and Garland (2003), Karmen (1991), and Peterson (1991).

blameworthiness on the part of the victim. In fact, one of the central questions of the thesis concerns when an imperfect victim's behaviour is blameworthy and, further, what this should and should not entail for their right to self-defence. Blameworthiness is therefore not built into the concept.

I will call a victim of an unjust attack “imperfect” if they can be held, in some sense, partially responsible for the threat or threat situation they are now facing. The sense of responsibility I have in mind is a minimalist conception of responsibility, which (again) does not entail blameworthiness, and which some call agential or agent-responsibility. To quote Seth Lazar, “one is agent-responsible for doing x if one voluntarily chose to x , and, when choosing, one met the (appropriately specific) minimum standards of responsible agency” (2009, p. 702). By the latter, Lazar has in mind standards such as “minimal rationality, physical self-control and nonmanipulation” (2009, p. 706).

Thus understood, however, agent-responsibility only applies to actions themselves, and that is not what is most relevant here. After all, in *Perfect Victim*, Brian is agent-responsible for *going to sleep*. Even if this act caused Alexis to break into his house and attack him (for instance, she saw him go to bed and then broke in), Brian is not an imperfect victim. In other words, that Brian is agent-responsible for x ing and his x ing happens to cause the threat is insufficient.

A victim is imperfect only if they are also partially agent-responsible for the *outcome* – the threat situation. For this to be true, there needs to be a more agential, as opposed to merely causal, relationship between their responsible action and the outcome.²² I think the necessary element is *foreseeability*: the threat must have been, in some sense, a reasonably foreseeable outcome of the action that the victim was agent-responsible for. In other words, it is necessary

²² Here, I am outlining roughly the type of responsibility that McMahan, Otsuka, and other proponents of the moral responsibility view of liability (see above) think is essential for liability.

that “[i]t was foreseeable to the agent that her action might result in the kind of outcome that it actually resulted in” (Burri 2020, p. 385).²³ It is, of course, difficult to determine exactly what this requires. To see this, consider the following examples concerning responsibility for creating threats to innocent people:

Gun Shop Owner: Alex has a gun shop. One day, he sells a gun to someone who matches the description of a person the police have warned gun shop owners about because they know he is planning a violent bank robbery.

Cell Phone Operator: Unbeknownst to Alfred, his phone has been tampered with so that when he presses the “send” button, the phone will detonate a bomb that will kill an innocent person (McMahan 2009, p. 165).

Conscientious Driver: Alice always keeps her car well-maintained and drives carefully, but one day “a freak event that she could not have anticipated occurs that causes her car to veer out of control in the direction of a pedestrian” (McMahan 2009, p. 165).

Suppose the unfortunate outcomes occur in all cases. The actual outcome was clearly reasonably foreseeable for Alex, and it was clearly not reasonably foreseeable for Alfred. The former had good evidence suggesting that selling the gun to the customer would contribute to a robbery, but Alfred did not have any reason to suspect that using his phone would cause an explosion. Having one’s phone tampered with in that way is so out of the ordinary and far-fetched that no reasonable person could have expected their phone use would pose a risk of causing this outcome. Things are more complicated with regard to Alice in CONSCIENTIOUS DRIVER. Jeff McMahan (2009, p. 165) argues that it was not foreseeable *that* driving on that day would cause harm. However, the outcome was foreseeable in the sense that everyone knows (or should

²³ Note that Burri is critical of the notion of foreseeability. See also Quong (2020, pp. 29-32) for criticism of the notion.

know) that driving is a dangerous activity and always carries *some* risk of causing that kind of outcome.

I am less sure what to conclude about liability in *Conscientious Driver*, but nevertheless think that the notion of foreseeability helps elucidate the concept of an imperfect victim. Part of what makes Brian in the original case seem like a perfect victim is that the threat situation was not a reasonably foreseeable outcome of his decision to go to sleep. By contrast, some of the cases of imperfect victims that I will talk about involve the victim intentionally provoking and goading another person. Being threatened is surely a reasonably foreseeable result of such behaviour. Other cases I have in mind include, for instance, those in which a victim walks alone through a dangerous park at night or quarrels with someone outside a pub in the evening. At least in many instances, I take it, being mugged or ending up in a physical altercation are reasonably foreseeable outcomes of those actions.

As said, however, there is a difficult question of where to set the bar for reasonable foreseeability. Does, as McMahan seems to suggest, the mere fact that one knows that an activity carries *some* risk of ending one up in a dangerous situation suffice for the victim to be partially agent-responsible for the situation if it occurs? In one sense, almost anything we do in life carries *some* risk of resulting in danger. Plausibly, the magnitude of the risk must be more significant than McMahan suggests. Although I am unsure where exactly one should set the bar, I will try to avoid some of these complications by focusing mainly on cases in which the victim *knows* or considers it *likely* that their behaviour will put them in danger.

Another key part of being an imperfect victim is that one had adequate opportunities to avoid acting in the “imperfect” manner. Again, it is difficult to determine what exactly counts as an adequate opportunity in this context, but the basic idea should be intuitive. Consider again someone contemplating walking through a dangerous park at night to get home. I take it that they have an adequate opportunity to avoid the dangerous park if they can take a few minutes

detour around the park, go through a safer park next to it, or hitch a ride with a friend driving to their destination anyway. The reason is that the costs of avoiding the danger, or significantly reducing the risk of danger, are pretty small. However, there are more complicated cases. Suppose the only way of avoiding the dangerous park is to pay for a taxi. Exactly how expensive must the trip be before that alternative counts as inadequate? It is difficult to say – and it may depend on individual circumstances (for instance, how much one’s salary is). But, again, I will try to avoid some of these complications by focusing on cases in which I stipulate that the alternatives to exposing oneself to the relevant dangers are adequate.²⁴

To summarise, I will understand an imperfect victim as a victim of a wrongful attack who voluntarily chose to *x*, had adequate opportunities not to *x*, and could reasonably foresee that *x*ing would lead to the present threat situation. This casts a very wide net, but that is partly the point. As Gill says of perfect victims, “they barely exist” (2020). (And recall, my use of the term imperfect victim is not, in itself, intended to cast any blame on them). There are often some grounds for assigning some partial agent-responsibility to them. That is why it is particularly important to understand precisely when an imperfect victim’s behaviour compromises their right to self-defence and when it does not.

3 The Right to Self-Defence

We also need a basic framework for approaching these questions. That is, we need some basic idea of the nature and grounds of people’s right to self-defence. Then we can start looking at how the behaviour of imperfect victims may or may not complicate things.

²⁴ However, this will not always be possible as one can expose oneself to the risk of being attacked for different reasons. For instance, one might do it to assert or stand up for one’s rights. If that is the goal of going into the dangerous park, then going around it will not be an adequate alternative, even if it is easy and not very costly. I return to this more in Chapter 3.

One can have different starting points, but mine is a rights-based one. According to a rights-based account, the right to use defensive force against someone is ultimately grounded in the fact that we have *rights* not to be harmed and that the other person is trying to violate one or more of those rights. First, then, let me say more about what I take these underlying rights to be. Although there are many different rights one may have, and these may have different grounds, I am here primarily concerned with a subset of rights related to well-being and autonomy: the right not to be mugged, not to be beaten up, not to be raped, not to be shot, not to be tortured and so on. For simplicity, I will refer to these as rights against being harmed.

I consider these rights grounded in our interests in well-being and autonomy, and the fact that these things (mugging, torture, being shot, etc.) negatively affect our well-being and autonomy. Borrowing from Cecile Fabre: “rights are interest-based, and the interests they protect may be interests in what happens to us as well as interests in deciding what will happen to us” (2000, p. 15). More precisely, I have in mind what is, since W. N. Hohfeld (1919), often called claims or claim-rights. These are rights which correlate with *duties*. A person *V* always has a claim-right to something which correlates with another person’s, *X*’s, duty to ensure that thing. Following Joseph Raz (1984, p. 195), on the interest-based conception of rights favoured here, the possession of such rights is explained by the claim that the interests at stake are important enough to ground a duty on someone else’s part.

The concept of a right is important because it picks out a particular kind of wrongdoing. Some wrongful actions do not necessarily have any victims. Littering might violate a duty (for instance, one imposed by the state), even if it does not make anyone worse off. Other wrongful actions may harm someone without being wrong (primarily) for that reason. *X*’s kidnapping of *V* may cause *V*’s neighbour emotional distress, but that is not (primarily) why *X*’s actions are wrongful. Instead, when *X* kidnaps *V* and thereby violates one of *V*’s rights, he is guilty not just of acting wrongfully but of *wronging V*. He violates a *directed duty* – a duty owed to *V*

grounded in V's interests. The wrongful act thus has a special moral significance in how it relates the wrongdoer and the victim.

More than that, possessing a right is often thought to come with various morally significant features, which are not present when other duties (not correlated with rights) are at stake. Lisa Hecht summarises many of these suggestions succinctly:

[R]ight-holders may enforce their right and they can claim compensation or an apology for rights infringements. They may resent the infringement and forgive those who wronged them if they so choose (2022, p. 10).²⁵

I will not delve into the thorny debates over whether one of these features, or perhaps another, is essential to possessing a right and present in all cases involving a right. Though, it seems plausible that all of them are relevant to a wide range of rights.

The most relevant feature in our context is the first feature mentioned – enforceability. That feature is intimately connected with the subset of rights that I am interested in here.²⁶ In my view, part of the very function of our rights against torture, beatings, and so on is to provide us with normative protections against those setbacks to our interests, and that includes some permission to use force if those interests are threatened. At least, these rights all seem *presumptively* enforceable. I add that qualification because, as I will argue, imperfect victims can compromise the right to enforce their rights without necessarily lacking the right not to be harmed.²⁷

²⁵ See also Owens (2012, pp. 44-67) for an overview of several suggestions concerning what is special about wronging someone as opposed to merely doing wrong. One suggestion not mentioned by Hecht is the idea that right-holders have some special kind of standing to demand or claim compliance with the correlative duties and to blame the wrongdoer in case of non-compliance.

²⁶ It is popular to link, at least many, claim-rights and enforceable duties. See, for instance, Cruft (2013), Wenar (2013), and Flanigan (2019).

²⁷ Draper also makes such a qualification “because there can be moral considerations that might require one to forgo enforcing a right [...]” (2016a, p. 49).

I believe our right to impose defensive harm on aggressors is fundamentally tied to these enforceable rights. This position, in some form or another, is popular in the literature. Here are, for instance, Gerald Lang and David Rodin:

The Defender has interests which her rights serve to protect. The central function of her rights is to protect those interests and to give her normative remedies for preventing their frustration. As a rights-bearer, the Defender can typically be expected to have permissible means to frustrate those frustrations, [...] (Lang 2017, p. 245).

Defensive rights seem to be entailed in a very basic way by rights *to* things. Thus if I have a right to X, then it seems to follow as a simple corollary that I have the right to take measures to prevent my right to X from being violated (Rodin 2002, p. 37).²⁸

As stated, though, nothing in these claims holds that one can only use force when doing so is *proportionate* to the threat one is facing and *necessary* to avert it. I will come back to these later, as they are complicated conditions. For now, I will just add that I take it for granted that enforcement is only permissible within those constraints.

More importantly, there is an ambiguity here in how someone's right against harm grounds their right to self-defence. On one view, the enforceable right is essentially all we need to explain the right to use defensive force against another person. Doggett (2017) and Lang (2017) seem to have something like this in mind. For both of them, the fact that a victim is *non-liable* is essential for explaining their right to self-defence. Being non-liable is broader than having a right against harm since one can be non-liable to being crushed by a rolling boulder, but one does not have a right against the boulder that it does not kill one. However, the notions are intimately connected. On Lang's (2017, p. 245) view, being non-liable entails having a

²⁸ For others who also make rights central to the right to self-defence (but in various ways), see, for instance, Narveson (1965), Thomson (1991), Draper (2009), Quong (2020), and Burri (2022).

permission to take (necessary and proportionate) steps to neutralise threats to protected interests that are either wrongings (i.e., rights violations) *or* not-permissible-but-not-impermissible (e.g., a boulder about to kill one). Consider:

Falling Man: Bex is standing at the bottom of a well. Alec, an unconscious individual, has been pushed over by a gust of wind and is falling towards Bex. If he falls on her, she will die, and he will live. But Bex can vaporise him with her ray gun, thereby killing him but saving herself.²⁹

Alec is an innocent or non-responsible threat here. It is by no means settled whether Bex can kill him in self-defence or what the grounds for this permission would be. Part of the disagreement here concerns whether Alec can be said to *violate* Bex's right not to be killed if he falls on her. In favour of thinking so, some may highlight that Bex is innocent, and Alec does not have a *right* to crush her, while others may insist that Alec is not operating as an *agent* at all and thus cannot be said to *violate* Bex's rights.³⁰ On a non-liability view, however, this debate is avoidable. The critical fact is that Bex is *non-labile* to be crushed by the threat Alec (or his body) poses. That is sufficient to give her permission to try to neutralise this threat.

By contrast, I believe the right to self-defence does not rest solely on people's rights not to be harmed or non-liability more broadly. One reason is that, on this view, it seems that Bex's right to kill Alec is identical in ground and scope to Brian's right to kill Alexis (if she tries to *lethally* beat him up) in *Perfect Victim*: they are both upshots of Bex's and Brian's being *non-labile* to be killed. But that, I think, is wrong. Even if, all things considered, Bex is permitted to kill Alec to save herself, that permission differs from Brian's. First, for instance, I am inclined to think that if he woke up and saw Bex about to vaporise him, Alec could also try to defend himself against her. This is not the case in *Perfect Victim*: Alexis cannot try to prevent

²⁹ This case is originally due to Nozick (1974, p. 34).

³⁰ See, for instance, Thomson (1991) for the former view and McMahan (2005a) for the latter.

Brian's defensive force. Second, it seems preferable that (if possible) Bex and Alec *divide* the unavoidable harm between them in some fashion – perhaps perfectly equal or perhaps somewhat shifted in Bex's favour.³¹ But that seems false in *Perfect Victim*.

When I talk about a right to self-defence in the remainder of the thesis, I will (unless stated otherwise) have in mind the kind of permission to use force that Brian has – which is more asymmetrical and robust than I think Bex's permission is. What sets this right to self-defence apart is that it targets a *liable* aggressor – someone who has forfeited their rights against defensive harm.³² But, again, a person's possessing a right against harm is key in explaining another person's liability. To borrow from an earlier paper by Lang, we can say that an aggressor's liability and forfeiture of rights “is *made* true by the claim that [he] threatens to violate the Victim's rights” (2014, p. 58).³³ Susanne Burri makes this connection particularly strong:

It is *constitutive* of our rights against being harmed that they have an enforcement dimension: if they are about to be violated, there are certain things that we—or third parties acting on our behalf—are permitted to do in order to prevent the violation. *Liability to defensive force* is an element of this enforcement dimension (2022, p. 546; emphasis added).

However, this does not necessarily commit us to the broad view of liability which entails that Alec in *Falling Man* is *liable* to be defensively harmed. As Burri explains, this rights enforcement view is open to different views on liability because various views can be re-described in its language. Recall the views I mentioned above, which focused on different kinds of responsibility (causal, moral, culpable) for a threat. In Burri's view, these can be captured by the rights

³¹ See Øverland (2011a) for a more sophisticated proposal for dividing harm between agents in these situations.

³² As said, views involving liability and forfeiture are very popular in the literature, and I am following them here. However, as also said, I am skipping over some of the deeper issues relating to liability and forfeiture. See footnote 10 above.

³³ This is also what most of those quoted earlier in relation to the rights-based view seem to have in mind – for instance, Rodin, Quong, Thomson, Draper and Burri. They all focus on the violation of a right, or the threat of violating a right, as grounding the aggressor's liability.

enforcement view. It just depends on what one thinks is necessary and sufficient for threatening to *violate* a right. Some may think that simply casually posing a threat to an innocent person's interests, as Alec does, is sufficient. But some may say that one only violates a right if one can be held morally responsible for the threat to the other person's interests. Others again might hold that one only violates the rights of another person if one culpably contravenes them. Given what I said above, I am sceptical of a causal account of rights violations. Apart from that, I do not intend to defend a more specific view on liability. My view is more in line with the moral responsibility view. However, in the remainder of the thesis, I will avoid some of the complications in this debate by focusing primarily on culpable aggressors.³⁴

To summarise, my starting point for thinking about the right to self-defence grounds it in our rights against harm. First, part of having a right is to be permitted to try to prevent the interests it protects from being set back. Second, this includes a permission to impose (non-trivial) *harm* on aggressors because they are liable to defensive harm in virtue of threatening to violate those rights.³⁵

4 Compromised Defensive Rights

Strictly speaking, my view is slightly more complex. To see this, we can consider a potential objection to the project I am undertaking here. If V's right to self-defence against X and X's liability to defensive harm is grounded in V's rights and the fact that X is threatening to violate one or more of those, then the issue of imperfect victims becomes a "non-problem". Either the imperfect victim V will have done something so morally egregious that he has forfeited his *right* not to be harmed by X (for instance, trying to attack X, in which case V is the aggressor

³⁴ However, in Chapter 5, I look at cases in which the victim's behaviour affects how culpable the aggressor is.

³⁵ Does working within this approach to the right to self-defence radically limit the relevance of the discussions to come? No. First, rights-based approaches are very popular in the literature on defensive harming. Second, the arguments and views to come could, I believe, be restated in terms compatible with other approaches to the grounds of the right to self-defence.

and X the victim), or the imperfect victim V has not acted so egregiously that he has forfeited his right not to be harmed by X, in which case we should just treat the case as we would any case involving a culpable attack on a perfect victim.

The objection oversimplifies things. Recall the qualification I briefly mentioned earlier: our rights are *presumptively* linked to enforceability. To be more precise, we have three relevant concepts on the table: “V’s right” (not to be beaten, killed, raped, and so on), “X’s being liable to defensive harm”, and “V’s right to impose defensive harm on X”. Although the last two spring from the former, there are only presumptively linked. The latter two can come apart from the underlying right and from each other.³⁶ And this is precisely what opens up the possibility that imperfect victims can affect their right to self-defence in different ways.

Part of the goal of the thesis is to discuss these possibilities in more detail. But we can illustrate the possibility of these features coming apart with other examples. For simplicity, I will focus on cases in which they come apart voluntarily. Imagine that V has a bully, X, but he also feels sorry for X’s mother, Y. He knows she will be extremely sad if X is harmed and therefore promises not to harm her son even though the bullying is severe enough to make defensive force (otherwise) legitimate. In this case, V may waive his right to use defensive force against X, but it does not involve waiving the underlying right (it is still wrong for X to bully V). Moreover, I do not think that it involves removing X’s liability. If V were to wrongfully impose defensive harm on X after all, it would be wrong since it broke a promise to Y, but it would not *wrong* X.

It is also possible for V’s right to use force against X to be compromised because of duties owed to X. Imagine that X is aware of his tendency to lose his temper easily. In

³⁶ Thus, I disagree with Rodin’s (2002, p. 75) claim that X’s forfeiting his right against harm and V’s right to harm him cannot “part company”, and Burri’s (2022, p. 546) claim that X’s liability is a constitutive part or upshot of V’s having a right. However, to reiterate, this is not to suggest that they can come easily apart. The standard scenario is that they come together, and only come apart when complicating factors are in the picture.

recognition of this, V may promise not to use defensive harm against him if X should lose his temper with V in the future. Again, this does not make it permissible for X to attack V, but it can make it wrong for V to use (otherwise permissible) force against X. Whether this also affects X's liability is also a further question. One might hold that, *qua* wrongful aggressor, X is liable to defensive force, yet that it is impermissible to act on this because of a contrary duty. Alternatively, one might think that the contrary duty also affects his liability.³⁷

Furthermore, the right to self-defence may not be simply either lost or not. It can come in degrees. A promise not to defend against X can be justifiably broken. Suppose V has promised to never defensively break X's fingers. If X poses (say) a *lethal* threat, the promise may be justifiably broken. Plausibly, how severe X's threat must be for it to be permissible to impose this defensive harm after all will also depend on the grounds of the initial duty. If the promise was made to X by V out of sheer kindness, it might be easily broken permissibly. However, suppose the promise was made to X as part of an agreement for V to use X's property for his business venture. Intuitively, it is more difficult to justifiably break the promise in that case, and thus more difficult for V to justifiably defend himself against X.

The point here is simply to illustrate the theoretical possibility of altering one's right to self-defence and (possibly) the aggressor's liability without necessarily altering the underlying right not to be harmed. The "non-problem objection" is therefore misplaced. The links between rights, enforceability, and liability are tight and the starting point for the right to self-defence, but they can come apart under certain circumstances.³⁸ For simplicity, I will refer to cases in

³⁷ See also Clark (2023) who claims that right-holders can decide whom their aggressors are liable to be harmed by. According to him, if X attacks V, and Y can save him by harming X, it is possible for V to refuse protection from Y without waiving his right not to be harmed by X and without waiving his own right to defensively harm X. In that case, Clark claims that X liable to be harmed by V but not liable to be harmed by Y.

³⁸ Similarly, while Ferzan grounds an aggressor's liability in their culpably threatening to violate someone's rights, she claims that rights-holders can "lose the standing to complain about being wronged (and lose the ability to prevent being wronged) while still being wronged" (2020, p. 126). Recently, Kamm (2022, pp. 349-51) and Clark (2023, pp. 46-50) also claim that rights-holders can lose the standing to demand compliance with, and the right to enforce, a right without losing the underlying right.

which this occurs, for whatever reason, as ones in which the victim's right to self-defence has been *compromised*. The question we are interested in is whether the behaviour of an imperfect victim can involuntarily compromise their right to self-defence and, if so, how and to what extent.

5 Fairness and Responsibility

There is also another possible objection to the project I am undertaking here, which comes from the opposite direction. Many view the ethics of defensive harming as an issue of distributive justice. In short, a threat situation is a *forced choice*: it is unavoidable that someone will be harmed (unless it is possible to avert the threat without violence). Thus, there is some harm that must be distributed amongst the parties involved. The fairest way to sort out this distributive challenge, furthermore, is to adopt a principle that distributes the harm in line with how responsible or at fault the different parties are for the situation. More precisely, the fairest distribution might be the one suggested by Kaila Draper:

[O]ther things being equal, if more than one person is to blame for the fact that someone must sustain a cost, then that cost should be divided among those persons in proportion to their blame or, if the cost cannot be divided, then the one who is most to blame should sustain the cost (1993, p. 77).³⁹

Later, Draper rephrases this same basic idea in a right-based form and focuses on responsibility instead of fault.⁴⁰ Applying this to the issue of imperfect victims, the philosopher's job becomes (seemingly) relatively easy. Consider the (i) to (iv) variations of the *Perfect Victim* case, for instance. Our job is to determine Brian's comparative degree of responsibility for the threat

³⁹ Many others have also approached the ethics of self-defence through a distributive fairness lens. See, for instance, Montague (1981), McMahan (2005b), Segev (2005), Farrell (2013), and Gordon-Solmon (2018). Rodin (2011, p. 95) also suggests something like this view in relation to provocation cases in particular.

⁴⁰ See, for instance, Draper (2009, p. 76; 2016a, p. 79).

situation in these cases. The more responsible he is, the more his right to self-defence will be compromised. If he is most responsible, he cannot use any defensive force. If he is responsible but not most responsible, the amount of defensive harm he can impose is restricted in proportion to his responsibility for the situation. End of story.

I disagree – matters are not so simple. First, I am sceptical that the ethics of self-defence, in general, can be so easily reduced to a matter of forced choices and distributive fairness.⁴¹ Second, it is not obvious how to determine comparative degrees of responsibility in this context. Draper suggests something like the following. When two people, X and V, are responsible for an outcome, O, the most responsible one is the one who is blameworthy for their contribution to O. If both are blameworthy “but X is more blameworthy for the role X plays than [V] is for the role [V] plays, then X is more responsible than [V] for O” (2016a, p. 79). Suppose that V steps into X’s line of sight, knowing that X is looking to kill him, just because V insist on crossing the road here and now, and that X then tries to kill V. Who is *most culpable for their contribution* to the threat here? It seems difficult to determine pre-theoretically. This is partly because it also depends on whether we think that exposing oneself to wrongful danger in this way, knowing that one might need to use defensive force, is itself a source of culpability.⁴² The matter is also made even more difficult by Draper’s suggestion that blameworthiness, in this context, is not limited to moral blameworthiness but also foolishness and imprudence. But it is very unclear how to determine the “general culpability” of an immoral act compared to an imprudent one. At the very least, the philosopher’s job is more complicated than suggested. We would need a more fundamental theory on the table in order to determine the relevant sources of different degrees of responsibility and fault in this context.

⁴¹ For general critiques of the distributive fairness approach, see, for instance, Wasserman (1987), Miller (1992), Kaufman (2008), Quong (2020), and Burri (2022).

⁴² As we will see, there is widespread disagreement on this.

Relatedly, a distributive fairness approach is at risk of simplifying or flattening the moral landscape here. As illustrated in the previous section, I believe imperfect victims' actions can (at least in theory) compromise their right to self-defence in different ways. It is not just a matter of degree. For instance, there is a difference between V compromising his right to use defensive force against X without affecting X's liability, and V compromising that right by affecting X's liability. Of course, one might say that this is just the difference between V sharing responsibility and V being most responsible for the threat. However, we still need to know why this is the case. And again, it seems possible that there are different reasons why these facts hold. As suggested above, V might have different duties not to defensively harm X, some owed to X and some not. These differences also matter for the stringency of V's duty not to use defensive force and for other moral upshots (such as counter-defence, compensation, and so on).

Thus, even if it might be possible to describe our judgements about the cases to come in terms of comparative degrees of responsibility and fault, this should not be our starting point but the end product. We would first need more precise arguments about what counts as relevant to comparative degrees of responsibility and fault. We also need, I believe, arguments that are nuanced enough to capture the various ways in which a victim can increase their overall degree of responsibility in this sense. Part of the aim of this thesis is precisely to provide such arguments. Ultimately, then, this version of the non-problem objection to the project also seems mistaken.

6 Scope, Assumptions, and Method

I opened this introduction by saying that the topic of imperfect victims – or the role that victims play in determining the right to self-defence – has received comparatively little attention in the literature on the right to self-defence and that I aim to help fill this gap. A full theory of

imperfect victims and defensive harming must consider imperfect victims in all sorts of threat scenarios. However, I primarily limit my focus to cases involving culpable aggressors.⁴³ This is an important starting point. First, the right to self-defence is most non-controversially taken for granted in cases of culpable aggression, and I am mainly interested in the possible restrictions on the right to self-defence that our own behaviour can create. Second, as I hope to illustrate, enough puzzles and complications arise in this subset of cases to fill a thesis project (or more). Moreover, I also believe that the fundamental arguments and claims formulated here can be appropriately extended and applied to other threat scenarios, thus contributing to our understanding of the defensive rights of imperfect victims more generally.

In addition to the theoretical assumptions and commitments already mentioned, I will be working with several others which are popular and common in the literature on defensive harming. First, as might be obvious, I set aside issues concerning determinism and free will scepticism. I take for granted that we can make sense of, and apply, terms such as responsibility and culpability. At some level, a complete investigation of the ethics of self-defence needs to grapple with these issues head-on, but I am not taking up that task here.⁴⁴ Next, as might also already be obvious given my focus on rights, I assume that utilitarianism and consequentialism, more generally, are false.⁴⁵ Instead, I work within a broadly moderate deontological approach to ethics, which is popular in this literature. For instance, I believe everyone has certain rights, but these do not typically ground absolute prohibitions on acting against those rights.

Take a classic trolley scenario in which someone X can divert a runaway trolley onto another track, killing V but saving a large number of lives. It might be permissible, even though V has done nothing to lose his right not to be killed. In this case, since X is acting with a

⁴³ Though, as flagged earlier, in Chapter 5 I look at cases in which the aggressors are at least not fully culpable.

⁴⁴ See, for instance, Tadros (2016a, pp. 67-90) for a discussion of some of these issues in the context of criminal (and defensive) liability.

⁴⁵ See also Draper: “the ethics of self-defence is itself a good reason to reject act-utilitarianism” (2016a, p. 66)

justification, he is only *infringing* V's right. If he has no such justification, however, he is *violating* V's right. To describe cases of acting against a right without specifying whether there is a justification, I will talk about *contravening* the right.⁴⁶

Moreover, I also think that *modes of agency* matter. For instance, the fact that someone has a right not to be killed does not tell us everything we need to know about the prohibitions against killing them. It matters *how* the killing will come about. Most importantly, there is a difference between killing V as a *side-effect* of doing something good and killing V as a *mere means* to that good. For instance, even if it is permissible to divert the trolley onto V's track to save five people on the other track, it might not be permissible to *lethally use* V for the same goal (for instance, to push V onto the track, killing him, in order to stop the trolley before it hits the other five).

Methodologically, the thesis also follows the majority of the existing literature in being heavily case-based and intuition-driven. I use (often highly) stylised examples to start discussions, test principles, and illustrate arguments, and I often use my intuitions about cases to influence the direction of my arguments. Although there are criticisms of both aspects of the methodology, I set those aside here.⁴⁷

However, it is important to highlight that this method does not reduce our theories and positions to mere intuitions. Intuitions help us start thinking about cases, but they cannot stand alone. They must also be backed up by careful analysis and argumentation. One also needs to check intuitions, and the conclusions they suggest, against broader ethical principles. The aim, in short, is to find a reflective equilibrium. This process involves, as Quong puts it, “going back and forth between particular judgements and more general principles, along with more general theories, with the aspiration of reaching a position where particular judgements cohere with

⁴⁶ The infringe-violate distinction is, as far as I know, due to Thomson (1977, p. 47), but I borrow this particular way of drawing it, as well as the “contravene” term, from Quong (2020, p. 14).

⁴⁷ However, see Quong (2020, pp. 14-17) for a discussion of, and some responses to, various worries.

more general principles and theories” (2020, p. 15). But this is not to deny that the method I am employing here will inevitably lead to occasional stalemates due to irreconcilable intuitions about cases. In those cases, the best I can hope to do is to offer as plausible arguments as possible to support my intuition.

7 Overview of Chapters

The thesis consists of five main chapters, roughly separated into two parts. Chapters 2 to 4 make up the first part which is the more critical. They look at three possible grounds for why the defensive rights of otherwise morally conscientious, imperfect victims might be compromised and largely reject these suggestions. Chapters 5 and 6 turn to cases involving more morally devious, imperfect victims and provide novel explanations for why their actions compromise their defensive rights.

Chapter 2 – “What Do We (Not) Owe to Aggressors?” – starts the investigation by asking what a potentially imperfect victim owes a potentially culpable aggressor. In particular, do we owe it to potential aggressors to stay out of trouble because of the defensive harm we otherwise likely have to impose on them? This matters because it suggests that if one wrongfully fails to stay out of trouble, there is a stringent duty not to use defensive harm. The chapter argues that we do not typically owe it to potential aggressors to stay out of trouble for their sake. The challenge is to explain why this is not the case. I survey and criticise several suggestions before landing on a view that holds that the wrongful intervening agency of the aggressors protects us from owing it to them to stay out of the trouble they will cause.

Chapter 3 – “A Duty to Avoid Trouble and Its Limits” – claims that one might nevertheless have *a* duty to stay out of trouble because of the inherent value of aggressors’ interests and lives. This duty, I suggest, is best characterised as an imperfect, charitable duty. Even so, I argue that it entails that the defensive rights of imperfect victims who violate it are at least

somewhat compromised. Towards the end, I consider possible objections to the duty to avoid trouble. Although I believe they do not rule out the duty completely, I argue that the considerations they rely on contribute to reducing its scope and stringency.

Chapter 4 – “The Irrelevance of Risk-Taking” – focuses instead on imperfect victims themselves. Does the fact that an imperfect victim voluntarily exposes himself to an increased risk of being wronged make it less wrong to attack him, thereby reducing the amount of force he may use to defend himself? The suggestion here – that voluntary risk-taking reduces the wrongness of attacking someone – can be found both in the philosophical literature and beyond. It is made more plausible by the fact that another claim – that a person’s voluntary risk-taking makes it easier to justify harming them – seems intuitive and intimately connected with the first. I agree with the latter claim but not the former. Part of the chapter canvasses and rejects some possible attempts to account for this. Instead, I suggest a view according to which someone’s voluntarily exposing themselves to a risk of harm only attenuates the reason not to harm them in forced choices.

Chapter 5 – “Provocation and Duties to Protect from Harm” – focuses on a more objectionable type of imperfect victim: the provoker. Although all the preceding chapters suggest that imperfect victims rarely compromise their defensive rights (or do so only to a small extent), I argue that people who knowingly provoke attacks seriously compromise their defensive rights. To explain this, I first look at legal perspectives on provocation, which focus on how a provoked attack can be partially excused and/or partially justified. I argue they cannot account for many core intuitions about provocation and the right to self-defence. Next, I turn to an account which focuses instead on the provoker being *partially responsible* for the attack he is facing. I suggest that there are some thorny counterexamples to this and that it fails to capture an interpersonal element of the reason why a provoker’s right to self-defence is compromised: he owes it to the provoked aggressor not to harm him. Instead, I outline an account based on duties to

protect from harm, which holds that the provoker must “protect” the aggressor from defensive harm because his behaviour reduced the aggressor’s fair opportunities to avoid attacking him and thus to avoid defensive harm. Importantly, however, this is only true insofar as the provoker affects the culpability of the aggressor.

Chapter 6 – “Defensive Entrapment and Manipulation” – presents a potential objection to the previous chapter. Suppose that the prospective victim provokes the aggressor so mildly that he does not reduce the culpability of the attack, but that he acts only because he wants to create a situation in which he can defensively harm the aggressor. Surely, this imperfect victim also compromises his defensive rights, to some extent, even though he does not affect the aggressor’s culpability. I agree, but think it is true for reasons other than the ones highlighted above. First, I argue that we should call these cases instances of “defensive entrapment” because they are relevantly similar to instances of legal entrapment. Second, I outline an account of legal entrapment, according to which entrapment is, at its core, a form of (non-culpability-affecting) manipulation which morally taints certain further actions, like the act of punishment.⁴⁸ Then, I apply this theory to defensive entrapment. Accordingly, the aggressor may be liable to defensive harming in the end, but the preceding manipulation has morally tainted the use of force. This compromises the right to self-defence in an importantly different way than in the provocation cases discussed in the previous chapter.

Taken together, the chapters provide a comprehensive overview and analysis of some of the core features that (potentially) bear on imperfect victims’ defensive rights. A general lesson is that we should look at the nature of an imperfect victim’s duties vis-à-vis the prospect of having to use defensive harm. Understanding the different possible grounds of such duties helps us clarify when (and, if so, how and to what extent) their behaviour affects their

⁴⁸ This part of the chapter overlaps with my article: Haeg, J. (2022). “Entrapment and Manipulation”, *Res Publica* 28(4): 557-583.

subsequent right to self-defence. The resulting view is a unified account that provides victims with robust protections against wrongful harm while also explaining why, in especially wrongful cases, the right to self-defence can be severely compromised.

2

What Do We (Not) Owe to Aggressors?

1 Introduction

We are interested in imperfect victims and their rights to self-defence. For simplicity, in this chapter and the next, I borrow and slightly modify a stylised thought experiment from Jonathan Quong:

Dangerous Party: Betty is invited to a party tonight, but she has been reliably informed that Albert will try to kill her if she attends. At that point, it will be necessary for Betty to defensively kill him. Instead of attending the party, she could stay home and finish a novel she is enjoying. Still, she decides to go to the party because she was invited and wants to see her friends. Albert then attacks her (2020, p. 133-4).

Betty fits the description of an imperfect victim. Her choice of going to the party was voluntary, and she could reasonably foresee that it would lead to a threat situation. She also seems to have had adequate opportunities to avoid this threat situation because she could have stayed home instead, which is not a particularly high cost. We can even stipulate that there are many more parties to come that she can attend without problem. Furthermore, the case does not involve any obviously nefarious intentions on Betty's part (for instance, that she goes to the party because she wants to kill Albert). It is therefore a nice and clean starting point for our investigation.

There are two morally important moments in this story. First, what should we say about Betty's decision to go, at time t_1 , knowing what is to come? Second, if she goes and is attacked,

at t_2 , what should we say about her right to defensively harm Albert? Although the two are intimately connected, the primary focus of the thesis is the second question. So, let me begin by outlining some of the possible answers one may give to that one. To start, one may disagree about whether Betty is permitted to defensively *kill* Albert. However, this disagreement obscures the fact that there can be several different, more precise views about Betty's defensive rights at t_2 underlying one's position on this.

To see this, we can imagine many versions of *Dangerous Party* in which the defensive options open to Betty differ. For instance, in one version, Betty's only way of defending herself is to shoot Albert in the knee. In another, she can save herself from being killed by either killing Albert or jumping out of a window, breaking her legs. In fact, we can imagine all sorts of variations of the case in which the possible options open to Betty differ in how much harm they will do to Albert and how much sacrifice they will require of her. In asking whether Betty's defensive rights are compromised, we are interested in all of these scenarios and what would be permissible for Betty to do in them.

With that in mind, one may claim that Betty in *Dangerous Party* is permitted to kill Albert because one believes in the following more general view:

The Lenient View: Betty's right to harm Albert is *uncompromised*.

By "uncompromised", I mean that the permissible defensive options open to Betty are identical to those open to her "perfect victim" analogue – that is, a version of Betty who unforeseeably ends up facing the same threat. In other words, in all possible versions of *Dangerous Party* in which the available defensive options differ, Betty will be permitted to choose exactly the same ones as the perfect victim analogue would be permitted to choose. On this view, Betty's behaviour leading up to the threat has no negative impact on her right to self-defence.

However, the belief that Betty may permissibly kill Albert is compatible with a less lenient view. Some may think Betty's defensive rights should be somewhat less permissive

than the perfect victim's rights. If push comes to shove and one has to die, as in the original case, Betty may permissibly kill Albert. But if certain other options are available, she might not be permitted to kill him. Imagine, for instance, that she can avert the threat by killing Albert or by jumping out of the window and breaking her legs. Suppose that the latter option is too harmful to demand that the perfect victim make such a sacrifice to avoid killing Albert.¹ Still, some might insist that because of the way Betty has (partly) brought the threat upon herself, this kind of sacrifice could be morally required. In that case, we can say that her defensive rights are "somewhat" compromised to signal that the possible permissible defensive options are worse than the ones she would have had if she had been a perfect victim. Call this:

The Moderate View: Betty's right to harm Albert is *somewhat* compromised.

At the same time, this view can come in different versions. Someone may agree that Betty's defensive rights are worse than the perfect victim's rights and also believe that Betty is *not* permitted to defensively kill Albert in the original case. However, they may hold that she would be permitted impose some serious but lesser defensive harm (if possible), like shooting him in the knee, to prevent the attack. Alternatively, one could also claim that Betty may not kill Albert because one embraces an even more radical view:

The Draconian View: Betty's right to harm Albert is *fully* compromised.

On this view, Betty has no permissible harmful defensive options. Even if she could avert his attack (say) by just punching Albert in the face, it would be forbidden.²

In the remainder of the thesis, I focus on these more general views on how an imperfect victim's defensive rights have or have not been compromised. But it is impractical to consider all possible versions of *Dangerous Party* separately at every stage. Thus, for simplicity, I will

¹ Those who disagree may increase the cost she would have to suffer to avoid killing Albert.

² Ferzan's (2013) view on provocateurs comes close to this. I return to this in Chapters 5 and 6.

mainly use the original one, and when I talk of whether Betty's right to self-defence is "compromised" in that case, I will not simply be making a claim about whether she has a right to kill instead of being killed, but also about what other defensive options would or would not be permissible if they were available. In short, the more compromised her right to self-defence is, the worse her possible permissible defensive options are compared to those a perfect victim analogue would have in the same situation. When necessary, I will specify those options as well.

Any view that holds that Betty's defensive rights are compromised must explain why this is so. More precisely, if Betty's permissible defensive options are worse than those a perfect victim version of her would have, it must somehow be because she decided to go to the party at t_1 , foreseeing that it was likely that Albert would attack. This brings us back to the first question: what should we say about Betty's decision to attend the party? In this chapter and the next two, I look at three different arguments concerning the moral relevance of that choice which might have a compromising effect on her subsequent right to self-defence. One plausible assumption is that Betty's defensive rights may be compromised if it was *morally wrong* to go to the party in the first place because she foresaw that she would likely have to use defensive force there.³ Although, in Chapter 4, I look at an argument that diverts from this assumption (it focuses only on the possible relevance of an imperfect victim's voluntarily exposing themselves to the risk of harm), here and in the next, I focus on possible arguments for concluding that Betty's decision to attend the party is wrong and therefore impacts the right to self-defence.

There are different ways that her initial decision can be wrong for reasons related to the prospective need to defensively harm Albert. First, one could argue that it is wrong for Betty

³ The "because she foresees the need to use serious defensive force there" qualification is important. There may be other reasons why it is wrong to attend the party. For instance, if she had promised to attend someone else's party instead. But I do not think this kind of wrongdoing negatively impacts her right to defend against Albert. However, see, in particular, Chapter 5 for more on the (ir)relevance of whether Betty's decision is "independently" culpable in this sense.

to go to the party because she *owes* it to him at t_1 not to set in motion a sequence of events likely to end with his being killed or seriously harmed. In other words, Albert is wronged by Betty's deciding to go to the party and harm him. Second, one might argue that Betty's decision at t_1 instead violates some more "charitable" duty which does not correlate with any right that Albert has against her. Part of the lesson of the thesis is that our view here is important, in general, for whether we should also think that imperfect victims' defensive rights are compromised and to what extent.

In the next chapter, I argue in support of the second suggestion above, but here I focus on and criticise the first suggestion. More precisely, I focus on what I call the Submission Argument, according to which Betty's decision to go and kill (or significantly harm) Albert is wrong for the same reason that a paradigmatic instance of disproportionate defensive harm *wrongs* the aggressor. If true, Betty would have a stringent duty owed to Albert not to go and kill him.⁴ In that case, wrongfully going to the party would also, I argue, significantly compromise her right to harmfully defend against him. My response to the argument is to take Albert's wrongful intervening agency seriously. It effectively insulates Betty from having to justify, to Albert, her decision at t_1 with respect to the harm she expects it will ultimately cause him. Thus, at least as far as what Albert is owed, Betty's defensive rights at t_2 are uncompromised.

The chapter proceeds as follows. Section 2 highlights a similarity between *Dangerous Party* and conditional threats and escalation cases. Section 3 distinguishes two ways of arguing that Betty should not attend the party. Section 4 outlines the Submission Argument in more

⁴ Betty could owe it to Albert not to attend the party for different reasons unrelated to his interest in not being defensively harmed. For instance, many argue that doing wrong is bad, harmful, or otherwise negatively affects the value of a wrongdoer's life. If so, one might say she owes it to Albert not to attend the party to avoid triggering his wrongdoing and saving him from the "harm" of immorality. I set this question to the side here, however. For the suggestion that doing wrong negatively affects the wrongdoer, see, for instance, MacLagan (1955, p. 219), Boxill (1976, p. 60), Thomson (1990, pp. 268-9), Bloomfield (2007; 2008; 2017), McMahan (2010, p. 60), Howard (2016, pp. 28-31) Parr (2016, p. 994), Tadros (2016a, pp. 1-2; 2020a, pp. 228-234), Brownlee (2019, pp. 110-115), Mund (2021, pp. 328-332), and Duus-Otterström and Page (forthcoming).

detail and clarifies its potential implications for Betty's right to self-defence. Sections 5 and 6 critically assess two possible ways of responding to the argument and argue that they fail because they are too radical. Instead, Section 7 develops a more moderate response which focuses on the aggressor's wrongful intervening agency. Section 8 introduces some clarifications and implications of the view which are relevant to later chapters.

2 Conditional Threats and Escalations

Before I outline the Submission Argument, let me quickly note that the structure of *Dangerous Party* is similar to other cases discussed in the literature on self-defence. *Dangerous Party* is what we may call a *triggering* case. When Betty makes her initial decision, there is no threat and, if she stays home, there will be no threat. A threat arises only if Betty attends the party. In this sense, Betty's action triggers the threat. We can imagine that Albert has not already decided to try to kill her, and not even contemplated doing so, when she is making her decision. He might think that Betty has moved abroad and that he will never see her again. Still, seeing her at the party will predictably cause him to decide to attack her. Now compare the following two cases:

Party Coercion: Betty is at a party, and Albert wants her to leave because he hates her. He issues a threat: if she does not leave within the next few minutes, he will try to kill her. At that point, it will be necessary for her to kill him instead. Alternatively, she can go home and finish a novel.

Party Escalation: Betty is at a party, and Albert wants her to leave because he hates her. He is trying to push her out the door. Betty knows that if she refuses to budge, he will try to use more force to get her out. At that point, she will have to use some defensive means proportionate to the new threat. But she knows that he is then likely to use even more force, making it necessary for her to increase the defensive force, and so on, until she has to kill him to stop him

from killing her. Alternatively, she can allow him to push her out and finish a novel at home instead.

The former is a *conditional threat* case. A threat to Betty already exists, but Betty's life will only be in danger if she chooses not to comply with the coercive demand issued by Albert. The latter is an *escalation* case. There is no conditional threat *issued* by Albert and, so, no attempt to coerce Betty by explicitly threatening to do more harm. Still, Albert is already wrongfully trying to push Betty out the door, and her life will be in danger only if she resists and sets in motion an escalation sequence.

These cases have a longer history in the literature than triggering cases like *Dangerous Party*.⁵ Although they differ from that case in the ways just highlighted, they all share the same basic structure. Betty could avoid a threat to her life and having to kill Albert by allowing herself to suffer a relatively small cost (going home or staying home). Thus, looking at arguments and views related to *Party Coercion* and *Party Escalation* is helpful in thinking about our main case. I will also occasionally use escalation and conditional threat cases, as it is sometimes easier to illustrate points and arguments with them. However, this is not to say that there are no morally relevant differences between these cases. When necessary, I will highlight them.⁶

⁵ See, for instance, McMahan (1994, p. 196), Rodin (2002, pp.132-38; 2014, pp. 79-88), Øverland (2010), Fabre (2014, pp. 103-13), Lazar (2014, pp. 25-31), Finlay (2015, pp. 63-69), and Lang (2022).

⁶ For instance, Quong (2020, p. 134n12) also notes the similarity between *Dangerous Party* and conditional threat cases but argues that an important difference is that, in the latter, the aggressor is already liable to some harm in virtue of wrongfully trying to coerce the victim. By contrast, as said, we can imagine that Albert in *Dangerous Party* is not already doing anything wrong since he has not decided to try to harm Betty yet.

3 Necessity and Proportionality

Suppose that someone insists that Betty does *not* wrong Albert by attending the party and killing him in *Dangerous Party*. There are (at least) two kinds of worries one may have with that.

To see this, consider the following two cases:

Two Options: Betty is at a party when Albert wrongfully tries to kill her. To avert the threat, she has to either retreat home (leaving Albert unharmed) or defensively kill him (Quong 2020, p 134).

Party Thief: Betty is invited to an invite-only party. Albert really wants to go but lacks an invitation. He therefore approaches Betty to steal it from her purse, but he is not willing to kill her to take it. Betty realises he is too strong for her to push him away and that the only way to prevent the theft is to shoot him dead first.

Many would say that Betty has a duty to *retreat* from the threat and a duty to *submit* to the threat, respectively, in these two cases. That is, in the first case, Betty is required to run home instead of using lethal force to avoid the lethal threat. In the second, she is required to allow Albert to take her invitation instead of using lethal force to stop him.

Based on those judgements, there are two arguments about *Dangerous Party* (as well as *Party Coercion* and *Party Escalation*) that one might make:

The Retreat Argument:

- (i) Betty is required to retreat in *Two Options*.
- (ii) If Betty is required to retreat in *Two Options*, then Betty is required to stay home in *Dangerous Party*.
- (iii) Therefore, Betty is required to stay home in *Dangerous Party*.

The Submission Argument:

- (i) Betty is required to submit in *Party Thief*.

- (ii) If Betty is required to submit in *Party Thief*, then Betty is required to stay home in *Dangerous Party*.
- (iii) Therefore, Betty is required to stay home in *Dangerous Party*.

We can flesh out the two arguments in more detail as follows. First, Betty is required to retreat from the threat in *Two Options* because, in the circumstances, killing Albert is *unnecessary*. According to the Retreat Argument, going to the party and killing Albert in *Dangerous Party* also counts as imposing unnecessary force.⁷ Second, Betty is required to submit to the threat in *Party Thief* because it is *disproportionate* to kill someone to stop them from stealing a party invitation. According to the Submission Argument, going to the party and killing Albert in *Dangerous Party* is tantamount to imposing disproportionate harm.⁸

This way of explaining the two arguments relies on a distinction between the necessity condition and the proportionality condition for permissible self-defence. As it is often put, the former condition is meant to ensure that a defender does not use more force than necessary to avoid or avert a threat. In contrast, the latter is meant to ensure that the defender does not use more force than is appropriate relative to the magnitude of the threat. A popular way of illustrating the difference is this. If X wrongfully tries to pinch V, then it might be necessary for V to kill X to prevent the threat if he has nowhere to run and cannot fight back non-lethally. However, killing X will be disproportionate because it only averts a very minor threat. In the other direction, if X wrongfully tries to kill V, it is proportionate to defensively kill him, but unnecessary if V can run away or disarm X and thereby avert the threat instead.

Things quickly become complicated, however. For instance, in the literature, some people talk about the duty to retreat as a corollary of the necessity condition while others talk about

⁷ Tadros (2014, pp. 381-83) suggests an argument with this structure, and, more recently, Schwartz (2020, pp. 593-597) explicitly makes it. It is also the worry that Quong (2020, pp. 134-37) has concerning *Dangerous Party*.

⁸ See also, for instance, Alexander (1999, p. 1479), Rodin (2002, pp. 132-38; 2014, pp. 79-88), McMahan (2014, pp. 146-49), and Lazar (2014, pp. 30-31) for arguments with a similar structure. Lang (2022, pp. 227-285) also considers and rejects both kinds of arguments in relation to conditional threats and escalation cases. I discuss this more below.

it as a corollary of the proportionality condition.⁹ Some also suggest that once we dig deeper into the two conditions, we should realise that the distinction between the two is small or not fundamentally important.¹⁰ What matters is whether a defensive harm is permissible. Thus, the above way of distinguishing the Retreat Argument and the Submission Argument – as corollaries of two different conditions for permissible self-defence – may not be so straightforward after all.

However, I think we should keep them apart. Labels themselves are not that important, but the proportionality and necessity conditions, in my view, track (or approximate) a more important distinction concerning different grounds for why defensive harming may be wrong.¹¹ For instance, certain defensive harms are wrongful because they *wrong* the aggressors. This is most clearly the case when the defensive harm is said to violate the proportionality condition. An example is *Party Thief* in which Betty kills Albert to prevent him from stealing her party invitation, or the case in which V kills X to stop X from pinching him. It is not just the case that the victim should not impose such harm. We should also intervene on the aggressor's side if the victim tries to use disproportionate force.¹²

By contrast, although I think that proportionate but unnecessary harm (such as killing Albert instead of retreating in *Two Options*) is wrong, I am less convinced that it *wrongs* the

⁹ For talk about the duty to retreat as a corollary of the necessity condition, see Quong (2022, p.134), McMahan (2016, p. 195), Arneson (2018, p. 256), Rodin (2002, p. 128), Fishback (2016, p. 280), and Lang (2022, p. 281). For talk about it as a corollary of the proportionality requirement, see Alexander (1999, p. 1479; 2013b, p. 1057), McMahan (2004a, p. 76), Ferzan (2019, p. 736), Hurd (2016, p. 256), and Steinhoff (2013, p. 1020). It is perhaps not surprising that many papers on conditional threats and escalation cases discuss both proportionality and necessity. See, for instance, Arneson (2018, pp. 252-261), Ferzan (2019, pp. 735-737), and Lang (2022, pp. 276-285).

¹⁰ See, for instance, Lazar who says that there is a “subtle difference between” the two conditions on his view of the necessity condition (2012, p. 21) and that “[p]roportionality and necessity are superficially distinct” and may be subsumed under one broader criterion (2017a, p. 44).

¹¹ Quong (2020, p. 130) also claims that we should keep the proportionality and necessity conditions apart because they pick out two importantly different ways in which harming an aggressor can be wrong. Though, for him, these are two different ways of *wronging* an aggressor, while (as I say below) I think they track the difference between wronging an aggressor and wrongfully harming an aggressor without wronging them, respectively.

¹² This is a popular intuition but not universally accepted. See, for instance, Schopp (1998, p. 84), Alexander (1999, p. 1502) and Steinhoff (2017, pp. 276-79) for scepticism.

aggressor. Many believe that it does, but it is more controversial than the view that disproportionate harm wrongs the aggressor, at least.¹³ I am inclined to think that other moral principles better explain the wrongness of such defensive harm, not that it violates a right that the aggressor has against the harm. For instance, it is less clear to me that we should intervene on the aggressor's side in these cases. At least, the harm we could use to stop the victim would be *significantly* less than what we could, intuitively, impose to stop someone from disproportionately harming an aggressor.¹⁴

I will say more in defence of this claim in the next chapter. For now, the main point is that because I believe the wrongs of unnecessary defensive harm and disproportionate defensive harm differ in nature, they can underpin two different kinds of arguments for why going to the party and defensively killing Albert in *Dangerous Party* is wrong. According to the Submission Argument, going to the party and killing Albert *wrongs* him because it is morally equivalent to using disproportionate harm. According to the Retreat Argument, it is wrong but does not necessarily wrong him as it is morally equivalent to using unnecessary harm.

As said, my focus here is on the first argument (I return to the second in the following chapter). If going to the party and killing Albert would wrong him and be equivalent to killing Albert in *Party Thief*, Betty has a very stringent duty to stay home, and others could potentially use significant force to prevent her from going. Additionally, I will suggest, the argument can be used to claim that Betty's right to defend herself if she goes to the party is also seriously

¹³ This is evidenced by McMahan (2016, pp. 195-210) and Draper (2016b, pp. 173-80). They argue that both unnecessary defensive harm and disproportionate defensive harm wrong aggressors. However, they spend most of the pages arguing for the first claim, seemingly taking it for granted that the second claim is more obvious and widely shared.

¹⁴ My view here is inspired by Frowe's (2014, pp. 91-106; 2016, pp. 154-59) arguments for the claim that necessity is "external" to liability. On that view, aggressors are liable to whatever harmful means are proportionate to stop their threats, even unnecessary ones. This is in contrast to, for instance, McMahan's (2016, pp. 195-97) suggestion that necessity is "internal" to liability, meaning that aggressors retain their rights not to suffer unnecessary defensive harm. See also Firth and Quong (2012, pp. 691-700) and Quong (2020, pp. 131-133) for a kind of middle position according to which aggressors forfeit some of their rights against proportionate but unnecessary defensive harm, but retain a "humanitarian right" or "right to easy rescue" to be spared from such harm.

compromised. Thus, it is important to explain exactly where the argument goes wrong if we want, as I believe we should, to avoid that conclusion.

4 The Submission Argument

First, let me explain in more detail why someone might find the Submission Argument plausible. According to it, recall, Betty is required to submit to the invitation-theft in *Party Thief* because the alternative, killing him, is disproportionate. Similarly, the argument goes, Betty is required to stay home in *Dangerous Party* because the alternative of going to the party and killing him is disproportionate. Why is it disproportionate? Of course, Betty's *life* is threatened if she goes which is different from *Party Thief*. However, the thought goes, if preserving her life is her goal, she can just stay home. The only thing she cannot secure without going to the party and killing Albert is attending the party. Thus, going to the party and killing him really only achieves the goal of allowing her to attend a party. But that is the same goal Betty tries to achieve by killing Albert in *Party Thief*. If killing for the sake of getting to attend a party is disproportionate in the latter case, it should also be so in the former case.

Similar concerns are found in the literature on conditional threats and escalations. For instance, Lazar describes a case in which X wrongfully wants to pinch V. If V resists, this will set in motion an escalation sequence consisting of the two of them going back and forth, at each step increasing the threatened harm and the defensive harm until, finally, V has to kill X to save himself from being killed. Suppose, importantly, that each of V's defensive acts, considered in isolation – that is, considered in isolation from the further threats and defensive harm it will create – is proportionate to the immediate threats. If satisfying proportionality at each step of the escalation is sufficient to satisfy proportionality overall, then we seem to end up saying that “it is proportionate for the victim to kill the aggressor, ultimately, to prevent him from pinching her” (2014, p. 31). Surely, something has gone wrong here.

Larry Alexander presents the same kind of worry in the context of another case which is perhaps even more analogous to *Dangerous Party*:

Suppose A is coming to shoot B on Main Street, where B is currently shopping, and B knows this. B can avert the danger now if he goes home. If he ‘retreats’ to his home, B will have sacrificed an interest to which he is entitled, namely, the liberty to shop on Main Street. If he does not retreat and chooses to kill A in self-defense when A arrives, he will have taken a life. If taking a life is not a proportional response to a threat to a few minutes or hours of liberty to shop on Main Street, then even if it *is* a proportional response to a threat to B’s life, B should be required to retreat before A arrives (1999, p. 1479).¹⁵

The thought here naturally applies to *Dangerous Party*. If Betty stays home, she sacrifices her interest in attending a party she is invited to. However, if killing someone to avert a threat to one’s freedom to attend a party is disproportionate (as in *Party Thief*), it should be wrong, for the same reason, to go to the party and kill Albert in self-defence. In both cases, one is trading a life for a party.

It is important to explain why Betty is not required to stay home in *Dangerous Party* because of the proportionality condition. However, our primary interest is in Betty’s right to defend herself *if* she goes to the party. On the face of it, the argument above seems to jeopardize Betty’s right to self-defence at the party. To see this, consider *Party Escalation* in which Albert tries to push Betty out of the party at t_1 and, if she resists, this will set in motion an escalation sequence ending with her having to decide between killing him or allowing herself to be killed at t_2 . According to the Submission Argument, the proportionality condition effectively starts applying to Betty at t_1 when she faces the initial threat and makes it wrong to resist because the

¹⁵ Note that Alexander does not think this worry applies to culpable aggressors because he seems to doubt that there should be a proportionality constraint on our defensive responses to such aggressors. See Alexander (1980, pp. 209-14; 1986, pp. 180-82; 1991, pp. 232-28) for more on this view. By contrast, I follow the (arguable) majority of philosophers in thinking that disproportionate harm wrongs aggressors.

significant defensive harm that resistance will ultimately necessitate is disproportionate to the threat of getting kicked out of a party. If she nevertheless refuses to obey the demands of proportionality at t_1 , this does not make it proportionate to impose *more* defensive harm later on. Instead, the maximum amount of permissible defensive harm at t_2 is capped at what was proportionate initially – that is, the amount of harm that is proportionate to use against the threat of getting kicked out of a party.¹⁶

However, some may resist this argument and insist that the Submission Argument has no implications for Betty's right to self-defence at t_2 . They may agree, in line with the initial argument above, that it is disproportionate for Betty to refuse to leave the party because we must factor in the fact this will make killing Albert necessary, and all she gains from resisting is getting to stay at the party. Nevertheless, they might argue that one needs a new proportionality calculation each time there is a new immediate threat. So, if Betty wrongfully refuses to leave and the lethal threat becomes actual, then it is proportionate to defensively kill Albert after all. In other words, although she wrongs Albert by not leaving the party, knowing what is to come, she does not wrong him by defensively killing him once he tries to kill her.¹⁷

I am sceptical of this position. We can cast some doubt on it in the following way. The core idea behind the proportionality condition is that it is impermissible to impose defensive harm which is excessive compared to “the evil averted” (Lazar 2017a, p. 44). Ordinarily, when facing a wrongful lethal threat, the evil to be averted is one's own wrongful death, so it is proportionate to kill the aggressor. But this may not strictly be true in Betty's case. Consider *Party Escalation* again. If she wrongfully (according to the Submission Argument) initiates the escalation and thereby puts her life “on the line”, anticipating having to kill to preserve it, then

¹⁶ Lang (2022, p. 278) calls this “intertemporal proportionality”.

¹⁷ This is close to Arneson's (2018, p. 256) position, I believe. In similar cases, he argues that victims owe aggressors a duty not to set in motion sequences (like escalations) which will end with them being *permitted* to defensively kill them.

that is on her, so to speak. Despite actually facing a lethal threat at t_2 , she cannot, in good faith, claim that the evil prevented is her wrongful death. As suggested above, if preserving her life was truly her aim, then she ought to have gone home at t_1 . Instead, since the only truly unavoidable threat to her is the threat to her safely attending the party, the proportionality condition at t_2 should be determined relative to that threat only. Ultimately, we end up with the same conclusion as the initial argument concerning Betty's defensive rights above.

This would entail something close to the Draconian View. The proportionality condition for what is permissible for her to do once lethally attacked at t_2 operates "as if" she is only facing a threat to her freedom to attend the party. Only minimal defensive harms would seem proportionate to defend herself, therefore. However, we can also consider a slightly less extreme version of this view. Even though Betty should not have put her life on the line, it is now at risk, and she cannot undo that. A more forgiving view might account for this by holding that the fact that Betty will be wrongfully killed now matters for proportionality but, since this threat was not unavoidable and she ought to have left the party, we should *discount* the disvalue of her being killed in determining how much defensive harm is proportionate at t_2 . If we do not discount to zero, we get a version the Moderate View instead. It may not be proportionate to kill Albert, but it could be proportionate to (say) shoot him in the knee (which seems disproportionate if she was only defending herself from being pushed out of the party).

This line of thinking seems attractive in certain other cases, at least. Consider, for instance:

Bystander Theft: A mugger, Albert, approaches a group of ten people and says, "your money or your kneecaps". The ten people understand that he will try to shoot them in the kneecaps and permanently disable them if they do not give him their money. At that point, one of them will have to throw a grenade at him to protect the group. However, they also know that a foreseeable side-effect of this is that the innocent bystander, Vera, will suffer substantial harm from the

grenade blast. Still, the ten people do not want to give up their money and refuse to hand it over.

Let us suppose that the harm to Vera is substantial enough to make it disproportionate for the ten to impose it as a side-effect of keeping their wallets. Say, both her legs will be broken, and she will be physically impaired for months. But suppose it is ordinarily proportionate to impose this as a side-effect of saving ten people from being permanently physically disabled. Given the way that the ten put their kneecaps on the line is it still proportionate to impose this harm on Vera to protect their kneecaps?

One could reasonably doubt this for the reasons given above. First, per the Submission Argument, it is disproportionate to refuse to give up their wallets, knowing that this will create a situation in which they have to impose severe harm on Vera to protect themselves from getting shot in the kneecaps. Second, if they refuse to comply, they undermine the proportionality of harming Vera to protect their kneecaps at t_2 . One could think either that what counts as proportionate now is identical to what would be proportionate to impose on Vera to defend against the original threat at t_1 or that the (*seriously*) *discounted* value of their kneecaps determines the proportionality. In either case, the group's right to impose harm on Vera while defending themselves against substantial harm seems seriously compromised. So, why should the same not apply to Betty in *Dangerous Party*, *Party Escalation*, and *Party Coercion*?

Here is one suggestion. In *Bystander Theft*, the group already owe it to Vera at t_1 to comply with the mugger's demands because they foresee that refusing will make it necessary to impose substantial harm on her to protect themselves against a worse threat – harm which is disproportionate relative to what they gain from refusing to comply now (keeping their wallets). That is why, if they fail to comply with this duty, they owe it to her at t_2 not to impose substantial harm on her or, at least, to seriously limit the harm. For by refraining from substantially

harming her at t_2 , they avoid imposing the very harm the prospect of which grounded the original duty owed to her at t_1 .

This reasoning is related to what is sometimes called the Continuity Thesis. In short, the idea is that the reasons that underpin a duty can continue existing after one violates or disrespects the duty and can ground even more stringent secondary duties. I return to this in much more detail in the next chapter. Here, I aim to explain why the above reasoning does not apply to Betty's situation in the three cases because the initial premise is false. In none of the cases does Betty owe it to Albert at t_1 not to attend the party or refuse to leave *because* of the substantial harm she otherwise has to impose on him to protect herself. Thus, I think, even if the form of argument above is valid in *Bystander Theft*, it does not entail that Betty owes it to Albert not to kill him or to harm him less severely at t_2 . Contrary to how it works in *Bystander Theft*, the proportionality condition for permissible defensive harming does not apply to Betty already at t_1 . Determining what Betty owes Albert at t_1 does not require factoring in the defensive harm he may end up suffering as a result of her decision at that time. The question is, *why not?*

5 Responsibility

There is a fundamental difference between Albert and Vera vis-à-vis the prospective harm at t_2 . In *Bystander Theft*, Vera cannot be held responsible for the harm she might suffer at t_2 , and the harm is not small enough to justify the group's setting in motion a sequence of events which will lead to her suffering it. That is why the group owes it to her to do as the aggressor wants and, if they refuse, will have their justification for harming her to save their kneecaps undermined. In *Dangerous Party*, by contrast, Albert can be held responsible for the harm he would ultimately suffer because it will be the result of his own wrongdoing. That is why Betty does not owe it to Albert to stay home to avoid having to impose such harm on him later.

This is on the right track. However, the focus on Albert's wrongdoing and responsibility is ambiguous between a forward-looking and a backward-looking version of the suggestion. To see this, compare the following case:

Dangerous Party Machine: As in *Dangerous Party*, but Betty knows that her life will be in danger at the party because Albert has culpably set in motion an "assassination robot" which is programmed to kill Betty if she shows up. Albert cannot shut it down until the party is over. If Betty shows up before that, she will have to kill him to avoid being killed by the robot.

On some views, this case is morally indistinguishable from the original case. That Albert has automated the conditional threat to her does not change Betty's reasons, or absence of reasons, not to attend and kill him in self-defence.¹⁸ The proportionality condition does not apply to Betty at t_1 when deciding what to do, or does not factor in the downstream harm she would have to impose on Albert at t_2 . That harm is irrelevant to what she owes Albert at t_1 since it is harm he will be responsible for in virtue of having wrongfully made it so that she has to impose it on him to save herself at t_2 .

However, I think this is too harsh. It does not properly account for the fact that Albert cannot undo the potential threat anymore. In a way, when Betty is deciding, Albert is just as vulnerable to what Betty decides as Vera is vulnerable to the group's decision-making in *Bystander Theft*. Of course, he is responsible for putting himself in this vulnerable position vis-à-vis Betty. But that is in the past and cannot be undone by him. In my view, that suffices to grant him a right against Betty going to the party and killing him.¹⁹

¹⁸ See, for instance, McMahan (1994, p. 196) for this view in the context of conditional threats. More recently, in a discussion of cases like *Dangerous Party*, Hecht (2022, pp. 9-17) argues for a similar view. In short, one does not owe it to Albert to save or protect him from the prospective defensive harm if he culpably put himself in the position of needing to be saved from it.

¹⁹ This is also what Quong (2020, p. 142n25) would say. See also Wasserman (1987, p. 372) for the view that it is more difficult to justify harming an aggressor who can no longer undo his threat.

To support this, imagine that Albert in *Dangerous Party Machine* could deactivate the robot after all, but that it would require sacrificing all his limbs. According to the view here, if Albert refuses, Betty does not wrong him if she attends the party and defensively kills him because we assign this outcome to his earlier unjustified action of activating the robot. This gives Betty a tremendous amount of normative power over Albert. Effectively, by going to the party, she can force Albert into having to choose between sacrificing his limbs or being defensively killed by her – just so that she may attend the party safely. However, I do not think she should have such normative power over him. Indeed, under these circumstances, going to the party and killing him seems to me wrong precisely because it seems disproportionate. We ought to intervene on Albert's side and (if possible) at least distribute the unavoidable harm more evenly between them. Additionally, if Betty managed to force Albert to choose to sacrifice his own limbs instead of being defensively killed, I believe she would subsequently owe him an apology for doing so.

A plausible explanation for this looks at Albert's duties after activating the robot. By doing this, he has made it more difficult for Betty to attend the party safely or without causing him serious harm. For that reason, he should ideally deactivate the robot or thwart the automated threat in some other way. But Betty's interest in going to the party is not sufficiently important to make Albert required to sacrifice his limbs to deactivate or undo this threat – even factoring in the fact that he is at fault for having created the automated threat in the first place. Since he does not have a duty to reverse the threat at such high costs, it would be disproportionate for Betty to go and kill him or force him to sacrifice his limbs for this sake.²⁰ The same is true in the original *Dangerous Party Machine* in which Albert's only way of stopping the threat is to allow Betty to kill him. She would violate the proportionality requirement by going and killing Albert in both cases.

²⁰ I am indebted here, in particular, to Tadros' (2011) way of thinking about enforceable duties and liability.

In other words, the present suggestion for why Betty does not wrong Albert in *Dangerous Party* is too broad. The obvious alternative view, which separates the original case from the automated version, is a forward-looking responsibility view which highlights that, in *Dangerous Party*, Albert's wrongful action lies in the *future*, not the past, of Betty's decision to go to the party. If that is the case, then there is no worry that Albert's fate would be vulnerable to Betty's decision at t_1 . He is perfectly able to, and required to, avoid threatening her later on.

6 Denying One's Own Agency

The forward-looking alternative is found in different versions in the literature as well. Christopher Finlay, in discussing an escalation case, asks whether the victim owes the aggressor a "duty to avoid the risk that the situation will unfold to the point where [he] will have to kill him" and answers that he doubts it precisely because "the factors that will lead to this escalation consists of [the aggressor's] own choices" (2015, p. 70).

Similarly, in relation to escalation cases, Lang discusses the view that escalating and defending against the final threat violates the necessity and proportionality conditions. He objects that such suggestions ignore the aggressor's future wrongful agency. Commenting on a case structurally similar to *Party Escalation*, for instance, he says that the view in question would essentially impose on Betty a duty "to spare [Albert] from a more serious defensive harm that will only be inflicted if [Albert] decides, impermissibly, to escalate" (2022, p. 285). The mistake, he claims, is that in making such a demand of the victim, one would incorrectly treat the aggressor's future wrongful escalation as a "*fait accompli*" (2022, p. 285).

Quong's (2020) view of *Dangerous Party* is similar. He focuses on the necessity worry, but it is relevant to our discussion since his concern is also whether Betty would wrong Albert by going to the party and defensively killing him. For him, the relevant question is whether Betty at t_1 owes Albert a *duty of easy rescue* – more precisely, a duty to rescue him from the

defensive harm she otherwise has to impose on him. Quong's answer is no. The reason is that Albert's demanding that Betty stay home at t_1 to save him would violate what is sometimes called the Interpersonal Test.²¹ In Quong's words, the test is this:

If a moral demand depends on an empirical premise, and the person issuing the demand (or on whose behalf it could be issued) is also the person whose future decision will make the empirical premise true, then if that person cannot provide a satisfactory justification of why she will choose to behave in a manner that makes the empirical premise true, this undermines the moral demand that the premise supports (2020, p. 138).

Its relation to *Dangerous Party* is this. Albert's demanding at t_1 that Betty stay home because of the defensive harm he will otherwise suffer depends on an empirical premise: that he will try to kill her at t_2 . However, Albert obviously cannot justify making this premise true, as it involves trying to violate Betty's right to life. Thus, Albert has no right against Betty at t_1 to be rescued from that harm – and so no right that she stays home. Moreover, one therefore cannot cite her having violated a duty owed to Albert as a ground for why her defensive rights should be compromised once he attacks her.

Although the focus on the forward-looking responsibility of Albert seems correct, we can still ask *why* it matters whether Albert's demanding that Betty stay home for his sake passes the Interpersonal Test. To answer this question, Quong seems to lean on the more fundamental idea that there is something morally objectionable about Albert "denying his own agency" (2020, p. 137). In demanding that Betty stay home for his sake, he would treat "his own future decisions as events over which he has no control or no responsibility" (2020, p. 137). Echoing Lang, we might say that Albert would be wrongfully treating his own future wrongdoing as a *fait accompli*.

²¹ The test comes from Cohen (2008, pp. 27-46) and is also discussed at length in Frick (2016, p. 235-65).

This is also how G. A. Cohen and Johan Frick describe the thought behind the Interpersonal Test. Cohen, for instance, claims that taking one's own free agency for granted in this way involves an illicit "alienation from [one's] own agency" (2008, p. 66). Likewise, following on from Cohen, Frick says that it involves taking "an essentially impersonal attitude towards [one's] own future actions, casting them as something outside [one's] control and thus immune to demands for justification". But since those actions are not in fact outside one's control, "this retreat to the impersonal point of view seems morally illegitimate – an evasion of responsibility for [one's] own actions" (2016, p. 238).²²

One worry, however, is that this account jeopardises the view that paradigmatic instances of disproportionate harm – such as killing Albert to prevent him from stealing a party invitation in *Party Thief* – *wrong* aggressors. Imagine that Albert is chasing Betty around, trying to steal her invitation. Realising she cannot keep him away much longer, her only options are to give up the invitation or to shoot him. The latter is, of course, disproportionate. But can Albert *demand* that she not do it? The Interpersonal Test suggests no. His demand that Betty not shoot seems undermined for reasons very similar to those which undermine his demand that she not go to the party and kill him in *Dangerous Party*. Although the demand does not take his *future* wrongdoing for granted, it takes his *present* wrongdoing for granted. When demanding that she not go and kill him in *Dangerous Party*, Albert is perfectly able *not* to threaten her in the future. Similarly, when demanding that she not shoot him in *Party Thief*, Albert is perfectly able to *cease* trying to steal her invitation and avoid being harmed in that way. Thus, in making this demand of her, Albert also seems to take a morally suspect impersonal point of view on his own wrongful agency, and to treat his refusal to cease threatening

²² The supposed illegitimacy of treating one's own future actions as fixed or beyond one's control is also found elsewhere in the literature. For instance, White (2017a, pp. 220-24) thinks coercion is wrong (in part) because it involves the coercer denying his own agency and shifting the responsibility for *his* future wrongdoing onto the victim. Similarly, White (2021, p. 826) argues that using one's own predictable but voluntary future moral failures as justifications for acting sub-optimally at present is often illegitimate because it denies one's own agency.

Betty as a given, a *fait accompli*, or “events over which he has no control or responsibility” (Quong 2020, p. 137).²³

In other words, if the problem in *Dangerous Party* is that it is illegitimate to deny one’s own wrongful agency while making a demand of someone else, then the same problem seems to exist in *Party Thief*. To avoid the more radical conclusion that even paradigmatically disproportionate harm does not wrong aggressors, we need an account of *Dangerous Party* which more properly focuses on the fact that Albert’s wrongdoing lies in the *future*.

7 Intervening Agency

In this section, I want to offer a slightly different spin on the forward-looking responsibility view. Instead of asking first what Albert, the aggressor, can or cannot reasonably demand of Betty, let us start with what *Betty* does or does not owe Albert. Although I think these questions ultimately track the same things, how we phrase the questions may affect our intuitive responses. Let us start with why it wrongs Albert to kill him instead of letting the invitation go in *Party Thief*. One explanation is that we owe it to everyone, including aggressors, not to harm or cause them harm unless there are sufficient reasons in favour of doing so. That is why, for instance, it seems that we wrong aggressors when we impose completely gratuitous defensive harm:

TWO BUTTONS: As in *Two Options*, Albert tries to kill Betty, but she can protect herself by pressing Button A or Button B on a device. The former will disable Albert’s gun; the latter will kill him. Pressing either button is equally easy and will keep her equally safe.

²³ Fabre (2022, pp. 531-33) makes the same kind of critique against Quong’s view in relation to the necessity condition. Wasserman (1987, p. 372) also critiques the forced choice or distributive justice model of the ethics of self-defence in this way. He points out that, from the aggressor’s perspective, a threat situation is rarely a forced choice since they can choose to stop threatening the victim.

Killing Albert here is gratuitous and not just “unnecessary” in the technical sense of the word that people often have in mind in the literature, which means something more like a defensive option which does not impose “the least morally weighed harm overall”.²⁴ In the original *Two Options*, there is *a* reason for using lethal defensive force – the alternative is retreating home. In *TWO BUTTONS*, by contrast, there is no reason in favour of killing him.²⁵ Plausibly, I think, no one forfeits their rights against such completely gratuitous harm.

This pushes us to answer the more difficult question: when there is *a* reason to harm someone in a threat situation, what counts as a *sufficient* reason? By answering this, we can get an idea of the proportionality condition. One might say, for instance, that for there to be sufficient reason to impose a given harm, the violation of the right which will otherwise happen must be of comparable importance to the harm one contemplates imposing. (Perhaps factoring the aggressor’s degree of culpability as well). Those sympathetic to the idea that unnecessary harm wrongs the aggressor might say something similar as well. One view is that whether the reason to impose a given harm on the aggressor is sufficient also depends on the other defensive alternatives and the difference in cost (to victim and aggressor) between those options.

The main point is that, fundamentally, this burden – a justificatory burden – on the victim applies to him when he wants to cause harm to the aggressor. It most clearly applies to him when he wants to perform actions that “constitute” harm – such as shooting the aggressor in the leg, stabbing him, etc. We might call these instances of *doing* harm. But ordinarily, the burden also applies in instances where the victim’s actions would “merely” *cause* harm. Suppose, for instance, that in *Party Thief*, Betty could stop Albert from taking her invitation by stabbing him in the neck with a syringe she keeps in her purse and injecting the substance into his body. This act, alone, is enough to freak Albert out and leave her alone. However, suppose

²⁴ See, for instance, Lazar (2012, p. 7) and McMahan (2016, p. 186) for this.

²⁵ Setting aside possible retributive reasons and reasons relating to standing up for one’s honour and dignity. I discuss the latter suggestion in more detail in the next chapter.

that Betty also knows that the substance will lead to his muscles deteriorating and him dying over time. Surely, she wrongs Albert here because she causes disproportionate harm, even though the preventive act considered in itself is not disproportionately harmful.²⁶

The question is why Betty in *Dangerous Party* does not have the same justificatory burden vis-à-vis Albert when deciding what to do at t_1 . To simplify a bit, suppose that she knows she is followed by a bodyguard who will kill Albert without hesitation once he tries to attack her at the party. Alternatively, in a case like *Party Coercion*, suppose she knows this bodyguard will kill Albert if she refuses to leave and he tries to harm her. In both cases, she knows that her decision at t_1 will cause Albert's later death. If this knowledge is sufficient to make her decision to insert the substance into his neck at t_1 in *Party Thief* an instance of wronging him, then why is it not in these two cases?

The answer is that the causal chain between Betty's act and the harm to Albert is "broken" by Albert's own wrongful agency. The syringe case and the bodyguard cases have the following structures, respectively:

- (i) V xs at time $t_1 \rightarrow X$ dies at t_2 .
- (ii) V xs at time $t_1 \rightarrow X$ ys at $t_2 \rightarrow Y$ zs (defends V) at $t_3 = X$ dies.

Although V in both cases might owe it to X not to x at t_1 unless the reasons for doing so are sufficient, how the sufficiency of the reasons is determined differ. In (i), V needs to justify the choice also with respect to the subsequent consequences she knows this will have for X . By contrast, in (ii), it seems to me that V 's justification for x ing at t_1 does *not* need to take into account those downstream consequences it will have for X . The difference, as suggested, has

²⁶ Though see Uniacke (2017, pp. 270-71) for a view according to which a disproportionate defensive response wrongs the aggressors if the act itself is disproportionate (for instance, stabbing someone who wrongfully tries to step on your toes), while others are merely wrongful if the act itself is proportionate but the foreseeable consequences are not (for instance, pushing the toe-stepper away even if one knows that it will cause them to fall into the road and be killed by a bus). In my view, both acts here wrong the aggressor.

to do with whether the consequences that V's xing will have for X is beyond his control or a result of his wrongful intervening agency. That is why Betty, in *Party Thief*, cannot justify scaring away Albert with the syringe, knowing that it would cause his death absent any intervening agency. But she can justify refusing to leave in *Party Coercion* because Albert's death will only follow if he subsequently wrongfully attacks her. Her duty to Albert, at best, only demands good enough reasons for deciding to stay at the party, *disregarding* those consequences. She satisfies this since she is an invited guest, and he has no right to kill her. Similarly, in *Dangerous Party*, Betty does not owe it to Albert to justify her decision to go to the party, factoring in the defensive harm which she will impose once he tries to kill her. The fact that she is invited to the party suffices to satisfy her duty to him.²⁷

This also coheres with one way in which Frick motivates the Interpersonal Test. He describes a case in which X threatens to kill himself unless V pays him \$20 now. Here, X's demanding that V pay up *because* of the harm that V's refusing will cause, fails because "even if [V] refuse to pay, [X] *remains entirely free to avoid any ill consequences for himself*, simply by deciding not to shoot himself" (2016, p. 252; emphasis added).

It also fits with the spirit of Lang's view on escalation cases. He argues that the proportionality condition (and necessity condition) vis-à-vis a particular threat – say, a lethal threat – from X does not apply to V as soon as he can foresee that his choices, like resisting another immediate threat, will lead to X posing it in the future. Instead, it only applies to V when X is in fact posing the threat in question. The reason is that, once X poses said threat, the existence of the threat does not depend on any *further* decisions from X. By posing a threat, X has successfully created "an unavoidable distributive problem *for [the victim]*" – the threat is facing V

²⁷ White (2017b, p. 333) talks about a kind of "burden of responsibility" whenever one has to justify one's behaviour because it will contribute to harming or failing to rescue someone. Freedom from this burden means not having to justify one's behaviour with respect to those downstream outcomes. My view here is effectively that Betty does not owe it to Albert to suffer this burden when justifying her decision to go to the party, given his intervening agency.

and “it is now [V’s] task to manage it” (2022, p. 285). Now, on one interpretation, this is misleading for the same reasons as given in the earlier section. There is not really an unavoidable problem for V insofar as X could *cease* to threaten but refuse to do so.²⁸

However, in another sense, the claim is correct. When posing a lethal threat, X has managed to make it so that V *has* to choose between imposing some harm on him or not. He has made it so that V cannot distance himself from the harm that will happen to X because V has to impose it. This is significant according to the view suggested here because it means that the interpersonal burden only to cause harm if there are sufficient reasons for doing so now applies to V. In other words, he now owes it to X not to impose more harm than is proportionate to what he gains by imposing said harm. By contrast, the burden does not apply to V if he merely foresees that his decisions now will cause X to decide to attack, or to increase the threat. In that case, X has not yet made it so that V cannot distance himself from the harm that will happen to X.

To return to our main focus, we can now also appreciate why Betty’s defensive rights, once attacked in *Dangerous Party*, are not compromised either. (At least not in terms of what she owes Albert). The Submission Argument errs in holding that the prospective harm to Albert at t_2 bears on what she owes him vis-à-vis her decision at t_1 . Thus, once Betty shows up at the party and has to defend herself, Albert cannot claim that she wronged him by not staying home initially. She therefore does not owe it to him to disvalue her own life when determining what counts as a proportionate response at t_2 . In other words, we end up with a view like the Lenient View.

Importantly, my view differs from the general view that wrongful intervening agency breaks causal chains and moral responsibility. For instance, in discussions of war, it is

²⁸ At least, the problem for V is not more unavoidable than the problem is if X sincerely threatens to kill him if he resists the initial threat.

sometimes suggested that the just side, when determining whether an act or operation is proportionate, can disregard or discount the death and destruction which will be brought about at the hands of the unjust side in response to the act or operation.²⁹ However, I share in the scepticism that others have concerning this proposal.³⁰ If, for instance, Betty knew that her attendance at the party would lead Albert to kill an innocent bystander instead, then it seems obviously impermissible for Betty to go. Indeed, it seems to me that her action would be just as wrong as if her attendance would predictably cause the person's death via some "non-agential" intervening processes, for instance, if it would trigger a robot assassin to kill them. In other words, whether there is *third party* intervening agency between Betty's action and this person's death is morally immaterial. What makes *Dangerous Party* and similar cases different is that the wrongful intervening actions are committed by those who will also ultimately be harmed by the initial action. That is why the initial agent does not *owe* it to the prospectively harmed person not to act because of the future harm.³¹

That is why the view also captures the earlier suggestion regarding *Bystander Theft*. Although there is wrongful intervening agency between the group's refusing to comply at t_1 and them having to substantially harm the bystander Vera at t_2 , this is third party intervening agency by the mugger, not Vera herself. Thus, Vera can demand at t_1 that the group take her well-being and the risk of harm to her at t_2 into account when deciding what to do. Once they fail to do what they should, and resist the threat, there is then also an argument to be made for them owing it to Vera not to impose what would otherwise be proportionate harm on her.

²⁹ For discussion, see, for instance, Hurka (2005, p. 47), McMahan (2014, pp. 150-51; 2017, p. 153) and Rodin (2014, pp. 82-83).

³⁰ See, for instance, Tadros (2016b, pp. 122-28), Howard (2019a, pp. 216-17), and Frowe (2022, pp. 188-91) for scepticism of the general intervening agency thesis.

³¹ See also Nino (1986, pp. 184-86) for reference to claims about intervening agency to avoid some worries about proportionality stemming from somewhat similar to the ones I have focused on. However, he seems to suggest that we embrace the more general intervening agency thesis. I think he is wrong, but a narrower view, such as the one I outline in this chapter, seems plausible.

8 Clarifications and Implications

To end, let me clarify and draw out some further implications of the view. First, clarifying what we mean by intervening agency here is important. Initially, both this proposal and the denying one's own agency suggestion above seem vulnerable to this kind of counterexamples:

Accidental Murder: As in *Dangerous Party*, but Albert will only try to steal Betty's purse. However, Albert is unaware that her purse contains medicine which she must take every hour to survive. Betty also knows this about him, but she cannot keep her medicine in any other place and cannot warn Albert that his theft will constitute a lethal threat to her.

In my view, Betty owes it to Albert not to attend the party and impose lethal defensive harm on him. Moreover, if she goes, she may owe it to him not to defend herself or, at the very least, to seriously restrain her defensive actions, even at great costs to herself. However, the two views considered so far might suggest something different. After all, Albert's demand that Betty stay away from the party for his sake still involves his relying on an empirical premise that he has no justification for making true (that is, trying to steal her purse). Similarly, Albert's wrongful agency breaks the causal chain between Betty's decision to go at t_1 and his being defensively killed.

The confusion is due to the fact that we can describe Albert's agency in different ways. Although there is wrongful intervening agency by Albert in this case, we cannot hold him responsible vis-à-vis the *lethal threat*. Put differently, we might say that there is an element or aspect to his actions – that they will pose a lethal threat – which he cannot really be said to be responsible for. Of course, he can be responsible for it insofar as he is free (and required) not to try to steal her purse. Nevertheless, given the facts he has access to, he cannot properly make his decision with sufficient appreciation of the lethal threat that his wrongful action will, in fact, constitute. She may therefore owe it to Albert to factor in the harm she will have to impose on

him to avert his lethal threat if she goes, in which case she seems to owe it to him to stay home. Intuitively, wanting to go to the party is not sufficient to justify the decision. In Chapter 5, I return to questions about how Betty's duties towards Albert are affected by how responsible he will be for a prospective threat.

There are also other ways in which Albert's beliefs can make a difference to what Betty owes him at t_1 . Consider:

Party Thief Murder: As in *Party Thief*, but Betty knows that unless she gives up her invitation, Albert will try to steal it from someone else, Cruella, who will likely shoot him dead.

If Albert is unaware that Cruella is likely just to shoot him dead, I am inclined to think Betty owes it to Albert to give up her purse to protect him from being disproportionately killed by Cruella. Here, we can say that Albert's agency vis-à-vis his own death is insufficient to break the causal chain between Betty's act and his death. But suppose that he knows Cruella is likely to kill him. The view considered here suggests that Betty does not owe it to Albert to give up her invitation, despite the harm that Albert will then likely suffer. That seems plausible since he knows it is likely to befall him as a result of his wrongdoing, and freely proceeds to act.

Lastly, it is worth highlighting how my view departs from many of the more standard views on *Dangerous Party*, escalation cases and conditional threats that are in other ways similar in spirit. One thing often highlighted as morally important is that the victim is, or would be, doing someone that she has (or ought to have) a right to do: say, going to the party or staying at the party. For instance, Finlay, in discussing escalation cases, talks of the significance of asking the victims to give up something which is "theirs by right" (2015, p. 64) as important to why they are not forbidden from escalating.³² But consider now:

³² This is also an important part of the arguments given in this context by, for instance, Alexander (1999, p. 1502), Hurd (2016, p. 265) and Quong (2020, p. 139). I come back to this idea, and related ones, in the next chapter.

Albert's Dangerous Party: As in *Dangerous Party*, but it is Albert's party, and he has not invited Betty because he does not want her to attend.

By asking her to stay home here, we are not asking Betty to give up something which is hers by right. On the contrary, she wrongs Albert by sneaking into the party uninvited. But, on my view, this does not necessarily affect her right to defend herself against Albert if he attacks her. The reason is that her decision to sneak into the party only wrongs Albert because she was not invited. It does not wrong him for reasons related to the harm she foresees having to impose on him. Again, this is because between her decision at t_1 and Albert being harmed, Albert himself wrongfully intervenes by trying to kill her. Thus, there are no grounds – at least not in terms of what owes him – for holding that her right to defend against the attack is compromised either. I return to this more in Chapters 5 and 6 as well.

9 Conclusion

I began this chapter by laying out different views that one may have concerning an imperfect victim's defensive rights – the Lenient View, the Moderate View, and the Draconian View. Using Betty in *Dangerous Party* as our example of an imperfect victim, I have argued that at least many such victims do not generally owe it to prospective, culpable aggressors to modify their behaviour because of the risk that they will be attacked and have to use defensive force against them. The aggressors' prospective intervening agency prevents prospective victims from owing them duties of this kind. Thus, at least one argument for why imperfect victims' defensive rights should be compromised is rejected.

3

A Duty to Avoid Trouble and Its Limits

1 Introduction

Although I have argued that Betty does not owe it to Albert not to attend the party in *Dangerous Party*, I will argue in this chapter that it is nevertheless morally wrong for her to attend the party, foreseeing that she will have to defensively harm him. In other words, there is a “duty to avoid trouble” grounded in the badness of the defensive harm one foresees having to impose.¹ The chapter starts, in Section 2, by providing a *prima facie* case for the existence of such a duty.

Next, in Section 3, I argue that this gives rise to a challenge for Betty’s defensive rights. In short, if her duty not to attend is grounded in the badness of the foreseeable harm she has to do, she may be required *not* to harm him once she is attacked. I explain how the so-called Continuity Thesis supports this type of argument. Thus, I do not believe we can avoid some version of the Moderate View after all. However, I explain why Betty’s rights will only be minimally compromised if she wrongfully attends the party.

Lastly, in Section 4, I consider some objections to the initial *prima facie* case I made for the duty to avoid trouble. I argue that they do not rule out the duty completely, but some of the considerations affect the scope and stringency of it.

¹ To clarify, it is a duty to avoid trouble grounded in the badness of the prospective defensive harm. As far as the argument of this chapter is concerned, Betty does not violate the duty if she goes to the party but allows Albert to harm her or manages to avoid it without harming him.

2 Unnecessary Harm

To start, imagine a version of *Dangerous Party* in which Betty decides to go to the party but does not bother to pay much attention in her deliberation, at t_1 , to the fact that she will then have to kill or seriously injure Albert. When her friend points it out to her, she shrugs it off and goes because she wants to party. We would all be justifiably abhorred to learn this. Treating her decision to go as morally trivial seems deeply problematic. It reveals a too cavalier or nonchalant attitude towards the prospect of killing or substantially harming another human being.²

This observation does not decisively prove that Betty is morally required to stay home. However, it suggests that she must take the prospect of harm seriously when deliberating about what to do. Part of taking this seriously, moreover, seems to be to weigh it as a consideration against going and to go only if the reasons in favour of it are comparatively good enough. I do not think the fact that she wants to party is sufficient. Albert's life and interests, even though he is (or will be) a culpable aggressor, are not so devoid of value that the desire to party outweighs it.³

A second reason to think that Betty has a duty not to go to the party comes from reconsidering *Two Options* and the Retreat Argument, which I briefly mentioned in the previous chapter:

Two Options: Betty is at a party when Albert wrongfully tries to kill her. To avert the threat, she either has to retreat home (leaving Albert unharmed) or defensively kill him (Quong 2020, p. 134).

The Retreat Argument:

- (i) Betty is required to retreat in *Two Options*.

² See also Hereth (2017, p. 618) who argues, more radically, for a “non-cavalier” *constraint* on specific actions, like killing, suggesting that killing someone as if it is “no big deal” itself makes it wrong.

³ In Section 4, I look at some other reasons in favour of going, or against staying home, that she may appeal to in order to justify her decision.

- (ii) If Betty is required to retreat in *Two Options*, then Betty is required to stay home in *Dangerous Party*.
- (iii) Therefore, Betty is required to stay home in *Dangerous Party*.

More precisely, the argument is that, just as the necessity condition for permissible self-defence demands retreat in *Two Options*, it demands staying home in *Dangerous Party*.⁴

Naturally, one might refer to the previous chapter for arguments as to why the two cases are different.⁵ That is, there is a difference in intervening agency. But this difference is mainly relevant to whether Betty *wrongs* Albert and, as already suggested, I am less inclined to think proportionate but unnecessary harm wrongs aggressors. Instead, I agree with Helen Frowe when she claims that:

The existence of [the necessity condition] must be grounded in a more general moral requirement not to cause unnecessary harm to people and other sentient beings [...] Even if killing an attacker does not wrong her and she has no special complaint against being unnecessarily harmed, it can still be all-things-considered impermissible to harm her (2014, p. 117).

If we just look at Betty's relation to the harmful consequences in *Two Options* and *Dangerous Party*, bracketing off the question of what she *owes* Albert, I think the two cases are very similar.⁶ Imagine, for instance, that Betty is accompanied by a robot bodyguard that she foresees will kill Albert once he tries to kill her. On one level, at least, her decision to go at t_1 *knowing* that the robot will kill Albert seems as wrong as it is to kill him instead of retreating in *Two*

⁴ Of course, one way to avoid the conclusion is to reject the premise that Betty must retreat in *Two Options*. Indeed, Hurd (2016) – whom I return to later – seems to reject the retreat requirement partly because she thinks it entails a duty to avoid trouble. My argument is directed at those who agree that one should retreat in *Two Options*.

⁵ As we have seen, Quong (2020) and Lang (2022) will also disagree on similar grounds.

⁶ See also Fabre who would “absolve Betty of wrongdoing vis-à-vis Albert – in both cases – if she kills him” (2022, p. 534).

Options. More precisely, from the perspective of t_1 , Albert's prospective death in *Dangerous Party* seems as unnecessary as his death seems when Betty has to choose in *Two Options*.⁷

Now, one might wonder what the nature of the duty not to go and kill Albert is supposed to be if not correlated with his *rights*. One option is to argue that it is an impersonal, undirected duty. Betty's duty is just one instance of a more general duty to minimise or decrease overall suffering in the world because it is inherently bad. In my view, however, that suggestion is too consequentialist in spirit. It does not properly account for the fact that Betty should not go due to how it will affect Albert in particular. It is not wrong to go and kill Albert just because it detracts from the universe's overall value or contributes to an impersonally bad state of affairs. Instead, it seems wrong in a more *personal* way – it is wrong because of how it affects *Albert*.⁸

Instead, I think the duty is grounded more directly in Albert's interests. Just because he does not have a right that Betty avoids going to the party and killing him, this does not prevent his interests from having any normative force. To borrow an example for Draper (2016b, p. 174), suppose that X freely consents to be struck by V in an attempt to show how tough he is. Even if we think that X waives his right not to be struck, it seems wrong for V to strike him absent a good reason for doing so. Putting it differently, the value of X's welfare is not extinguished by his consent, even if the consent can make it easier to justify harming him and

⁷ Strictly speaking, one need not agree with me that Betty's decision to knowingly cause Albert's death in *Dangerous Party* and *Two Options* are equally wrong, and do not *wrong* Albert. As Theron Pummer suggested, in conversation, perhaps Betty's actions in both cases are wrong for the same reason (they will cause Albert's death) but only in the latter cases does the wrong constitute a *wronging* (perhaps, indeed, due to the difference in intervening agency). All that is strictly necessary for the arguments to come is that the prospect of Albert's death is sufficiently morally bad to have some normative force on Betty's decision-making at t_1 , and thus that there is a *prima facie* case for her having a duty not to go.

⁸ However, see, for instance, Rodin (2017, pp. 36-40) for the suggestion that one may sometimes have a "lesser evil obligation" not to harm even fully culpable aggressors. Arneson (2018, p. 257) and Hecht (2022, p. 22) also mention the possibility of grounding the necessity condition in a consequentialist principle. Frick (2016, p. 262) also appeals to consequentialist reasoning to explain why he thinks it is ultimately wrong not to give \$20 to the person who threatens to kill himself if not. Lastly, Karhu (2019, p. 228) also argues that, in addition to directed duties, we always have (*pro tanto*) undirected duties not to harm people because it detracts from the impartial value in the world.

perhaps less wrong to wrongfully harm him. There is still some inherent disvalue in his being harmed.⁹

The same is true for Albert in *Dangerous Party*. The inherent value of his life makes it wrong for Betty to set in motion a sequence of events which will kill or substantially harm him when she can reasonably cheaply and easily avoid doing so, even if he lacks a *right* that she not do so. Thus understood, we might call the duty to avoid trouble a duty of “charity”.¹⁰

One might object, at this stage, that if Betty’s duty not to go to the party in *Dangerous Party* is grounded in the value of Albert’s interests, then, according to the interest-based theory of rights with which we began in the introductory chapter, it seems Albert has a *right* that Betty not go to the party and kill him after all. However, I believe cases like this provide exceptions to this. There are two ways of arguing for this. One is to point out that it appears that few, if any, of the facts that standardly hold when people’s rights are about to be violated seems to be hold in Albert’s case. For instance, intuitively, Albert cannot use force to stop Betty from going to the party, he cannot demand compensation and is not owed an apology if he is harmed, and he does not seem to have a special standing to demand that she not go or to complain if she goes.¹¹

⁹ See also Lazar (2016, p. 217) for the view that consenting to or not wanting to be saved from some harm does not remove the agent-neutral value of one’s not being harmed, which is relevant to how third parties should act.

¹⁰ In the past, Draper (2009, pp. 79-81) also grounded both proportionality and necessity in a duty of charity. See also Hecht (2022, p. 23) for the suggestion that the necessity condition is a corollary of a general duty of charity. Arneson also seems to ground proportionality and necessity in a “humanitarian” duty that we owe to everyone because of the “inherent value of life” (2018, p. 261).

¹¹ This is inspired by how Firth and Quong (2012, pp. 689) and Frowe (2014, p. 92) in part argue that aggressors do not have rights against unnecessary force since it is counterintuitive that they may counter-defend or demand compensation. However, some disagree that this shows that the person lacks a right against harm. For responses relating to counter-defence and compensation, see McMahan (2016, pp. 197-204), Draper (2018, pp. 93-96), and Clark (2023, p. 50n14). Additionally, as I noted earlier, people like Ferzan (2020, p. 126), Cornell (2016, p. 1149), Kamm (2022, pp. 349-51), and Clark (2023, pp. 46-50) believe that one can forfeit the standing to complain about a rights violation, especially if the complaints are hypocritical, without losing the underlying rights. Thus, not everyone would agree with the argument in the text. Nevertheless, as I go on to argue below, the most essential part of my argument concerns the stringency of the duty not go and harm Albert, which I believe people can agree with even if they want to insist that Albert has some kind of right against Betty in these cases.

Another argument for why Albert does not have a right against Betty here focuses more on Betty and how the value of his interests should figure in her deliberation. Intuitively, it operates very differently from how people's interests should figure in our deliberation when they are protected by a right. In discussions of the necessity condition, for instance, the most popular suggestions, from those who think aggressors are wronged by unnecessary defensive harm, is that aggressors have a right not to be made to suffer excessive harm or that they have a right to be rescued from suffering said harm.¹² However, the duty not to go and harm Albert appears to me very different from duties that correlate with those rights.

One attractive way of thinking about rights is, borrowing from Raz, to understand them as providing rights-holders with “a protective shield against the moral demands” of others’ welfare by having “a special force” in our deliberation (1986, pp. 249-50). On a Raz-inspired view, they function in this way by grounding duties with *peremptory* or *exclusionary* force, which is to say that they provide not only reasons for action but also reason to exclude certain other considerations from one’s deliberation about what to do.¹³ To illustrate, suppose A promises to help his friend B move flats the next day. The promise does not simply provide A with a strong reason to help B. It also seems to make it so that A should *exclude* certain other reasons from his deliberation about what to do when he wakes up the next day. He should not, for instance, lie in bed a few extra minutes weighing his reason to help the friend up against his reason to spend the day relaxing at home. His desire to stay home should not be seriously entertained in this way. He should just get up and help the friend as promised. Indeed, as David Owens (2012, p. 91) emphasises, even if A ultimately helps the friend, he seems to have failed

¹² See, for instance, McMahan (2016, p. 196) and Quong (2020, p. 131), respectively.

¹³ See, in particular, Raz (1999, pp. 35-48), as well as Scanlon (1998, pp. 156-57). I am following Owens’ (2008, pp. 413-14) and Renzo’s (2019, p. 347) formulations of Raz’s view on exclusionary reasons as excluding considerations from one’s *deliberating* with the aim of settling what to do. Additionally, note that talk of exclusion is not uncontroversial, especially the details. I do not intend to tie my views here to any particular theory, but the basic idea that (at least many) of our obligations, particularly those correlating with rights, have some exclusionary profile seems to me very attractive.

to adequately respect the promise if he seriously contemplated staying home instead. B has some grounds for complaint in that case even if A does end up helping as promised.

Rights against being killed and rights to be rescued from death function in the same way. They also exclude various considerations for being relevant to others' deliberation. In particular, considerations having to do with others' well-being or with one's own well-being. To see this, consider two other versions of *Dangerous Party* in which Albert is out of the picture. In version (1), Betty knows that her attendance will cause the innocent Lena's death through some complex chain of events. In version (2), Betty knows that her attendance requires letting the innocent Sara die because (say) it will take a minute to save her and doing so means missing the last train to the party.

Surely, Betty's interest in going to the party seems to be excluded by the respective rights here. Not only is it wrong to attend the party because it is fun in these circumstances, but it is also wrong to seriously consider attending it. Suppose Lena and Sara discovered that Betty had seriously deliberated about whether to go to the party, despite knowing what it would mean for them. In that case, they could reasonably feel disrespected and wronged by her.¹⁴

Compare now the original *Dangerous Party*. Although I have suggested that it is wrong for Betty to go to the party, Albert's interests seem to be relevant to her deliberation in a very different way. Importantly, it does not seem to me to *exclude* her reasons to attend the party. The fact that not causing the aggressor Albert harm requires not going to the party that she is invited to and wants to go to is a *relevant* consideration. It is just *outweighed* by the value of Albert's life in these circumstances. Her duty to stay home is a duty "despite" her reasons to go. It is not a duty that "excludes" those reasons. That is why Albert has no grounds for

¹⁴ In neither case, as well as in the promise case, is the suggestion that all contrary considerations are excluded. There can be justifiable reasons to break promises, kill someone (lesser evil scenarios, for instance), and allow someone to die. The point is that these kinds of rights tend to exclude many contrary considerations. But, of course, determining exactly what considerations are not excluded, and when, is difficult.

complaint if Betty takes those reasons into consideration when deliberating about whether to go. It is not a failure to respect the normative force that Albert's interests have on her, contrary to what is true of variations (1) and (2).¹⁵

Thus, it seems that Albert's interests in the original case do not have the "special force" and "protective shield" that paradigmatic instances of rights against being killed and rights to be rescued from death have. They do not stand out as special, or protected, as Lena's and Sara's are. His interests are just one consideration among several that are relevant to Betty's decision, and they only happen to outweigh the competition given the circumstances.

Now, some may disagree. They might argue that the difference between Albert, on the one hand, and Lena and Sara, on the other, is not a difference in kind but in *degree*. Albert's interests still exclude various considerations (perhaps, for instance, the reason to go because not showing up would be a *faux pas*); it just excludes fewer, but that is sufficient for his having a *right* against Betty. Perhaps we should instead hold that his interests ground a weaker or "watered down" version of the right not to be killed or the right to be rescued from death compared to Lena's and Sara's.¹⁶

At this point, however, I agree with Victor Tadros' contention about consent and rights. He describes a case (somewhat similar to Draper's case above) in which A consents to a radically life-shortening medical procedure to save his son from a very minor injury. Should the doctor go ahead with the procedure? Intuitively not, as the reason to seriously harm A, despite the consent, is not good enough. But does that mean that the doctor *wrongs* A if he operates on

¹⁵ Similarly, I would argue, *we* (third parties) would not have grounds for blaming her for deliberating in this way either. (Though we may blame her for *going* if she does so, something which is not true of Albert, as I suggested above). In the (1) and (2) variations, by contrast, we could blame Betty for not excluding her reasons for going to the party when deliberating about what to do.

¹⁶ McMahan (2016, p. 198) suggests that aggressors' rights against unnecessary harm are weaker or less stringent than innocent people's rights against similar harm. It is also popular to hold that one's right to be rescued from harm can differ in stringency based on one's responsibility for being in danger. See, for instance, Lazar (2017b, p. 21).

him? As Tadros notes, it does not seem that A could complain about it. At the same time, the doctor is doing something wrong because it will harm A in particular. In the end, however, as long as we agree it is wrong, Tadros claims, “we gain nothing by deciding whether the doctor has violated [A’s] rights” (2016a, p. 274).

Something similar seems true concerning *Dangerous Party*. Although I am persuaded by the two arguments above, that does not make it impossible for someone else to insist that Albert has a *right* against Betty’s decision to go and harm him. But, at that point, I believe our conceptions of rights (both their function and what it takes to have one) are so fundamentally different that I doubt we gain much by debating whether Albert has a right against Betty.

However, what is crucially important (as we will see), regardless of semantic disagreements, is the *stringency* of Betty’s duty not to go and harm Albert in *Dangerous Party*. Given the above view of how Albert’s interests should function in Betty’s deliberation, it is perhaps not surprising that I believe that the duty is not very stringent. But even someone who insists that Albert has a right can agree. Compare the duty not to go to the party if it meant that Lena would be killed or that one would have to let Sara die. Plausibly, this duty is very stringent. In the first case, Betty could only go if doing so was necessary to save a large number of lives or (perhaps) if doing so was necessary to save herself from being killed by someone. It might be somewhat easier to justify going and allowing Sara to die. Plausibly, avoiding a substantial but non-lethal cost could suffice, as would the opportunity to save another life by going. But it is intuitively still very difficult to justify not saving her.

In the original case, by contrast, it seems much easier for Betty to justify going. In my view, the mere fact that it would be an especially significant party could suffice. For instance, suppose it is Betty’s last chance to see a friend who will move abroad for several years. Alternatively, imagine that she needs to go to the party to collect some medicine for her broken legs, without which they will not heal for several months. Again, I am inclined to think that she can

be permitted to go to the party in these cases even predicting that she will have to kill Albert in self-defence. Even if one believes that the bar should be set higher, it seems intuitive that it is nevertheless much lower than in Lena's and Sara's case.¹⁷ In general, I believe, Betty is required not to go and seriously defensively harm Albert only if the cost to her of doing so is low or moderate.

To summarise, I have sought to motivate a *prima facie* case for a duty to avoid trouble in cases like *Dangerous Party*. Whether we say that it is correlated with a right might not be too important as long as we agree that the stringency of the duty is much weaker than the duty operating in cases in which non-aggressors would foreseeably be harmed. Next, I look at what this duty means for Betty's right to defend herself *if* she decides to go to the party anyway.

3 The Draconian View

If it is wrong for Betty to go to the party at t_1 , it is possible that her defensive rights once at the party are compromised as well. One radical option is the Draconian View, which would prohibit Betty from using any defensive force against Albert.

Daniel Schwartz (2020) effectively makes an argument for this when discussing the necessity condition over time. He argues that, in cases similar to *Dangerous Party*, it will be unnecessary for Betty to impose any defensive harm on Albert at t_2 . His argument is similar to the Retreat Argument: the necessity constraint for permissible self-defence against Albert's threat starts applying to Betty already at t_1 . At that time, Betty has two ways of avoiding Albert's threat: staying home or going to the party and defensively killing him. From that perspective, the latter defensive harm is clearly unnecessary. Indeed, even if her options were to

¹⁷ One way to think about the case is, again, to compare *Two Options*. Ask yourself if Betty could be required to suffer a particular harm instead of defensively killing Albert. If not, then I think the interest in avoiding similar harm in *Dangerous Party* is sufficient to justify going to the party, foreseeing that she will have to kill Albert. So, we might here ask whether we think that (say) Betty would be required to jump out a window and injure her leg for a few months in order to avoid defensively killing Albert. I am inclined to say no.

stay home or go and impose some serious but non-lethal harm, the latter also seems to count as unnecessary compared to averting the threat by staying home. Thus, any defensive harm downstream of her initial refusal to stay home violates the necessity condition. Consequently, it is wrong for Betty to defensively harm Albert at t_2 .¹⁸ As I argued above that the wrongness of her going to the party is similar to the wrongness of using unnecessary harm in *Two Options*, I am sympathetic to the spirit of this argument.¹⁹ But I strongly disagree with its conclusion.

Here is an initial, but ultimately unsuccessful, response. The problem is that Schwartz argues with little mention of *why* he thinks unnecessary defensive force is wrong. But what I have been suggesting this far is that we should understand unnecessary defensive harm as wrong because it violates a duty, grounded in the inherent value of the aggressor's interests, to avoid seriously harming them. Furthermore, this duty applies to Betty in *Dangerous Party* at t_1 which is why she acts wrongfully in deciding to go to the party and defensively harming Albert. So far, Schwartz's argument seems correct. However, I also argued that the duty in question is not very stringent – it can only require avoiding harming the aggressor if the cost to the victim is low or moderate. For that reason, the duty does not apply to Betty at t_2 , when she is attacked at the party. At that point, the cost to Betty of avoiding harming Albert will be tremendous – her own life. Indeed, even if the required sacrifice is somewhat lesser – say, jumping out of a window and breaking her legs, thus ending up unable to walk for months –

¹⁸ Strictly speaking, Schwartz (2020, p. 600) concludes it is only *pro tanto* wrong. Ultimately, on his view, even in the case in which she has to kill or be killed, Betty might be able to justify defending herself based on a sophisticated lesser evil calculation that weighs “wanton” harm more heavily than “excessive” harm. I am not so convinced by his response. First, I doubt we should weigh the *badness* of these harms differently. Second, we would still have something close to *The Draconian View*. If all that protects Betty from having to sacrifice herself is the fact that her defensive harm is a (morally weighted) lesser evil than Albert's threat, it is only if the two are very similar in magnitude that Betty may defend instead of allowing herself to be harmed. But I think this is too harsh as well. Alternatively, Schwartz could argue that wanton harm weighs radically more than excessive harm in lesser evil calculations, but that seems implausible.

¹⁹ Those who disagree that Betty's decision would violate the necessity condition but agree that she should not go to the party can see the force of Schwartz's argument in another way. Since any harm done to Albert at t_2 will be downstream of her violating her t_1 duty not to go and harm him, that defensive harm will be wrongful as well since it merely completes the duty violation.

she would not be required to do so instead of killing Albert. At least, she would not be required to do so if (as I am inclined to think) this is more than what she could have been required to suffer to avoid defensively harming Albert in the same way already at t_1 . In both cases, the response goes, the duty not to cause Albert harm does not exist anymore since the cost of compliance is higher than what could originally have been demanded of her.

In this way, we end up with a view more like the Lenient View. Betty should not go to the party, given the harm she foresees having to impose on Albert. But *if* she goes, her right to defend against him is *not* compromised. Her defensive permissions are identical to those that a perfect victim would have in her situation.

3.1 Continuity and Breach

The problem with this response to Schwartz's argument is that it does not take into account that violating a duty does not tend to let one off the moral hook. On the contrary, the violation of a duty tends to create duties to undo or mitigate the violation. These remedial duties also tend to be more stringent than the original duties. To see this, consider the following case:

Burning House: Albert's house catches fire and is about to burn down, causing him to lose all his possessions. At time t_1 , Betty can easily prevent this by turning on his sprinklers. She refuses, and Albert's house burns down. Later, at t_2 , Betty can help Albert offset the harm he suffered.²⁰

Assuming that Betty is required to help at t_1 , the fact that she chose not to seems very significant for what is required of her at t_2 . Intuitively, her reasons to help offset the harm differ from those that someone who happens to walk by has. For instance, if only one of them is needed to help offset the harm, Betty should do it. Moreover, Betty's duty to help now seems more stringent than her initial duty to help and, likewise, more stringent than the duty that a stranger has to

²⁰ Tadros (2020b, p. 190) discusses a similar case.

help now. That is, we can require more substantial sacrifices of her for the sake of Albert's well-being now than we could have initially.

What is sometimes called the Continuity Thesis helps account for some of this intuition. According to the thesis, the moral reasons which ground primary duties (say, to aid someone or not to harm them) do not simply vanish if we fail to comply with them. Instead, they – or at least very similar reasons – continue to exist post-violation and ground *secondary*, or “next-best”, duties to compensate or otherwise make up for the initial wrongdoing.²¹ This can help explain why Betty's duty to help offset the harm seems different from the duty to help which other strangers have. Her duty is remedial and the result of her violating a primary duty owed to Albert. That is why it is more appropriate that she rather than a stranger help him now.

As Tadros (2020b) explains, however, the Continuity Thesis as described cannot explain the intuition that secondary duties are *more stringent* than primary ones. The thesis only holds that the same (or very similar) reasons that ground the primary duty continue to exist and ground a secondary duty post-breach. If those reasons initially only grounded a duty of a particular stringency, how could they ground an even more stringent duty later on? To account for the fact that secondary duties can be more burdensome, Tadros suggests, there needs to be some special significance attached to the fact that the secondary duty is the result of the agent wrongfully *breaching* the initial duty.

I am sympathetic to this, but requiring an actual *breach* cannot be correct. Consider, for instance, the following:

Lunch: At t_1 , Betty sees Albert in danger of drowning in the ocean. She can easily rescue him, but refuses and goes to lunch instead. At t_2 , on her way back,

²¹ For more, see, in particular, Gardner (2011). For an excellent overview of different versions of *The Continuity Thesis*, see Steel (2020).

she is astonished to discover that Albert is still alive. Betty can still save him, but it is much more dangerous because he is further away from shore.²²

Intuitively, Betty's duty to help Albert at t_2 is similar to her duty to help in *Burning House*. In particular, her duty to save Albert at that time seems more stringent. Even if the cost of saving Albert now – say, taking on a substantial risk to her own safety – is so high that she could not have been required to suffer it to save Albert initially, she can be more easily required to suffer it to save him now (within limits – more on that below). But is this a secondary duty that has resulted from her breaching the primary duty to save him at t_1 ? Not obviously. Her primary duty is to save Albert from drowning in the ocean. As we have seen, she can still achieve this at t_2 . Consider also the following example modified from Stephen White (2015, p. 19):

Milk: Betty promises to get milk on her way home so that Albert can have it for breakfast the next morning. However, she decides to stay out drinking with her colleagues. She stays out so late that their local shop closes. The only way to get milk in time for breakfast now is to take a long detour to another shop.

Within certain limits, Betty could be required to take the detour even if it is so long that she would have been released from the duty had she not been at fault for not making it to the local shop in time. Yet it is not obvious that this results from having *breached* the initial duty if the promise was to get milk on the table in time for the morning. She can still achieve that even after the local shop closes. So, in both cases, it seems somewhat inaccurate to say that the duty bearers have *secondary* duties to save someone or get milk. Strictly speaking, they have yet to breach their primary duties. Nevertheless, that does not change the sense that the cases seem

²² See also, for instance, Barry and Øverland (2014, p. 570) and Burri and Christie (2019, p. 70) for similar cases, all of whom agree that Betty's subsequent duty to act is more stringent than it was initially. To avoid confusion, let us also stipulate that Albert is not suffering in the meantime. For instance, suppose he is just sitting happily on a paddle board but will eventually drown unless he is brought to safety in time.

very similar to the *Burning House* case: in all cases, there is a duty to act at t_2 which is more stringent than the duty at t_1 *because* the duty-bearer refused to comply with their duty at t_1 .²³

Furthermore, Tadros' explanation of why breach is morally significant in *Burning House* also translates easily into an explanation for why the t_2 -duties are more stringent in *Lunch* and *Milk*.²⁴ According to this, Betty has a duty in *Burning House* to stop the fire at t_1 which is grounded in Albert's interest. Thus, the value of his interests constrains what she is permitted to do at t_1 . Morally, she is not free to choose anything besides stopping the fire. However, Betty fails to appropriately respond to that value by choosing not to stop the fire. She acts *as if* she is not constrained by it. That is why she ought to constrain herself even more in light of that value later. As Tadros puts it in the first-person perspective: "given my failure to respond to the fact that these values constrain me, it is appropriate for me to give them special weight in my practical deliberations, and to constrain myself more strongly in responding to them now" (2020b, p. 197). That is why the resulting duty to act is more stringent post-breach.

Although there are no actual *breaches* of primary duties in *Lunch* and *Milk*, essentially the same story applies. Take *Lunch*. Much like in *Burning House*, Betty does not respond appropriately to the value of Albert's interests when she walks by him at t_1 . More precisely, Betty fails to constrain her decision-making appropriately by walking past Albert, believing that she will not be able to save him later or not caring about whether he is saved. Albert's interests do not play the constraining role they ought to have in her deliberation. That is why she also ought

²³ Borrowing Slavny's (2014, p. 144) terminology, we might call the duties at t_2 in *Lunch* and *Milk* "negatory duties", as they negate the bad effects that their decisions at t_1 (going to lunch or staying too long at the pub) would otherwise have.

²⁴ There are other possible explanations, as well. For instance, Barry and Øverland (2014, p. 574) highlight that, in a case like *Lunch*, Betty would be *responsible* for Albert's death if it is the result of her not rescuing him initially. In somewhat of a similar spirit, Tadros (2016a, p. 81; 2020c, p. 157) argues that people have interests in bad outcomes (like deaths) not being attributable to one's culpable choices. Another view, inspired by Scanlon (1998), is emphasised in both Tadros (2011, p. 56) and Burri and Christie (2019, p. 74) and holds that it is easier to impose extra burdens on Betty at t_2 because she could have *easily avoided* that cost by doing what she was required to do at t_1 . For reasons of space, I skip over these. I think they all have problems, and prefer the explanation suggested by Tadros in the text here. However, I think the challenge I ultimately raise for Betty's right to defend herself in *Dangerous Party* would work regardless of our particular explanation.

to give them extra weight in her deliberation and to constrain herself even more when she realises there is still a chance to save him. And that is why her duty to save him is more stringent at t_2 . In short, it is more stringent because she failed to appropriately respect the underlying value initially, even if she did not actually *breach* the duty.²⁵

Returning to *Dangerous Party*, we can now appreciate the real force of the Retreat Argument. Betty, in that case, is similar to Betty in *Lunch* and *Milk*. Her duty not to go to the party at t_1 is grounded in the inherent value of Albert's interests. By going to the party, foreseeing the need to harm or kill him, she fails to appropriately weigh those interests and constrain herself in response to them. Once she is attacked at the party, she can still bring her actions in line with respecting that value by *not* defending herself. Of course, this is very costly, but as we have just seen, Betty's duty not to harm those interests has plausibly become more stringent because she disrespected that value initially. At the very least, this pushes us towards a version of the Moderate View, according to which Betty's defensive rights are somewhat compromised when she is attacked at the party. Thus, the important question is how much more Betty is required to sacrifice for the sake of the value of Albert's interests. The more she is required to sacrifice, the more compromised her defensive rights will be.

Below, I consider two kinds of responses to this argument. The first tries to reject it by suggesting that the Continuity Thesis does not apply to *Dangerous Party* after all.²⁶ The second accepts it but tries to argue that the increased cost Betty could be required to suffer at t_2 in

²⁵ This is also compatible with holding that Betty's duty to act at t_2 is *not* more stringent in certain cases. For instance, suppose she justifiably believed that she could go to lunch and return to save him without any increased risk or cost, that Albert would be just as happy relaxing on his paddle board in the meantime, and that the lunch is quite important to her. Then, (arguably) she has given Albert's interests sufficient weight and appropriately constrained her decision-making around saving him. In that case, if saving him is (unforeseeably) made more difficult at t_2 , she might not be required to sacrifice extra to save him. Barry and Øverland (2014, p. 585) also seem to agree with this.

²⁶ In the remainder of the thesis, whenever I talk about *The Continuity Thesis*, I also refer to the additional claim that duties tend to become more stringent when disrespected.

Dangerous Party is only *slightly* increased. Thus, her defensive rights will, at best, be minimally compromised.

3.2 Responsibility and Rights

Echoing the previous chapter, some might argue that there is no continuity between Betty's duty not to attend at t_1 and her normative position at t_2 because there is wrongful, *intervening* agency by Albert himself. In other words, Albert's intervening agency stops Betty's initial disrespect of her duty from resulting in increased stringency at t_2 .

The response fails because it does not take into account that when we say that Betty has a duty at t_1 not to go and harm Albert at t_2 , we are already factoring in Albert's intervening agency. The duty is grounded in the inherent value of his interests regardless of his wrongful intervening agency. Thus, I do not think that merely pointing to his wrongful and primary role in creating the need for defensive harm suffices to explain why her duty should not exist and be more stringent at t_2 .

The situation between Betty and Albert with respect to the value of Albert's life is similar to the relationship between a primary duty bearer and someone required to pick up the slack due to the primary's unwillingness to comply with their duty. Suppose A is primarily responsible for saving his daughter from drowning but refuses. To save her, B now has to make a sacrifice. Although it is unfair to demand this of B, it seems plausible that he can be required to sacrifice, at least if the cost is not higher than what he could have been required to sacrifice if he was in fact the only one able to save her.²⁷ Suppose now that also B refuses. The longer he refuses, the more he has to sacrifice to save the child. In my view, his refusal to do what is required of him can increase the stringency of his duty here as well, even if A is the one with the primary responsibility for saving her.

²⁷ For more arguments in favour of a duty to pick up slack, see, Karnein (2014, pp. 598-604), Tadros (2016b, pp. 105-16), Enoch (2018, p. 6), and Stemplowska (2016, pp. 593-605).

Similarly, even though it is Albert who bears the primary responsibility to ensure that Betty does not have to harm his interests (by choosing not to attack her at the party), that does not prevent her refusing to constrain herself correctly in response to said value from making her duty more stringent.²⁸

Another argument for why the Continuity Thesis does not apply to *Dangerous Party* is that it only applies to duties correlating with *rights*.²⁹ After all, the examples considered so far – *Burning House*, *Lunch* and *Milk* – all involve Betty disrespecting Albert’s rights by not discharging her duty at t_1 . However, I have suggested that Albert’s rights are not at stake in *Dangerous Party*. So, perhaps that is a way to avoid the conclusion. Again, however, I think the response is mistaken. First, as Tadros makes clear, the explanation for why duties become more stringent given above is very general and “explains the stringent duty to respond to values that we set back in breaching our primary duty *whatever* those values are” (2020b, p. 197; emphasis added) – not just if a right has been violated.

Second, the Continuity Thesis also intuitively applies in cases where rights are not obviously violated. To use a mundane example, consider littering. Littering is wrong, but it is unclear whether you violate any rights by doing it. Still, if you wrongfully litter, you can have a duty to go back and pick it up. Indeed, the duty seems more stringent than the duty not to litter initially. You can be required to do more to undo the wrongful littering than you could be required to do not to litter in the first place.

Thus, I am sceptical of the arguments that the Continuity Thesis does not apply to *Dangerous Party* at all. The more promising route is to explain why the stringency of Betty’s duty not to cause harm to Albert is only slightly increased at t_2 .

²⁸ One might suggest that, in the A and B scenario, B can permissibly offload the additional costs onto A, who has the primary duty by forcing him to do what he should. But that is not possible in Betty’s case. Forcing Albert involves doing precisely what she was supposed not to do: harm Albert’s interests.

²⁹ For instance, Ripstein’s (2016) version of *Continuity Thesis* focuses on the claim that *rights* continue existing after they have been violated.

3.3 Value and Initial Costs

So far, I have said little about *how* much more stringent the duties become after they have been disrespected. But there is always a limit. In *Milk*, for instance, Betty cannot be required to sell her kidney to make sure Albert has milk in the morning. The question is how to determine the limit in different cases. One suggestion is that what matters is the *moral value* at stake.³⁰ However, if that is the case, one might think Betty's duty not to cause Albert harm at t_2 in *Dangerous Party* is very stringent. The value at stake here is a human life, which is also what is at stake in *Lunch*. And in the latter case, it is plausible that Betty could be required to suffer very severe harm to save Albert at t_2 . That should give us some reason to doubt that the suggestion is correct, for it does not seem that Betty is required to suffer similar levels of harm to avoid killing Albert in *Dangerous Party*. Alternatively, one might argue that the moral value of Albert's life is lesser in *Dangerous Party* than in *Lunch*.³¹ In that case, Betty's duty not to injure Albert's interests at t_2 in the former case is less stringent than her duty to save Albert at t_2 in the latter case.

However, avoiding some of the contentiousness concerning whether Albert's life has less value because of his culpability, we might instead say that how much more stringent a duty can become depends on its *initial* stringency. Consider the following case:

Parent: Betty can save her son, Albert, from drowning now, and so can Vera, who is a stranger. They both refuse. Shortly after, they both have another chance to save Albert, but it is now much costlier.

It seems intuitive that Betty's subsequent duty is more stringent than Vera's. One natural explanation is that Betty's *initial* duty was more stringent than Vera's. This is compatible with

³⁰ This is suggested by Burri and Christie (2019, p. 72).

³¹ See, for instance, Rodin, who claims that "[t]he responsibility or culpability of an agent for an unjust harm may play a role in determining the impersonal badness of harming that agent [...]" (2017, p. 39).

different ways of explaining this underlying difference. One might say that they both are required to act because of the same underlying value – Albert’s life – but that this value provides reasons with different strength depending on a person’s relation to that value (for instance, being the parent of the child versus being a stranger). Alternatively, one might claim that Albert’s life has some unique agent-relative value that only applies to the mother.³²

Either way, the initial stringency of the duty helps determine the stringency of the duty after it has been disrespected. This picture also fits with the view above. Following Tadros, I said that duties become more stringent because of one’s initial failing to give the value or consideration appropriate weight in one’s deliberation. One therefore ought to give it special weight subsequently. But, plausibly, how much more weight one should give it depends on how much weight it should have been originally given.

In that case, Betty’s duty not to cause Albert harm at t_2 in *Dangerous Party* is not very stringent because, as I argued above, her duty is not very stringent initially. By contrast, the duty in *Lunch* is intuitively much more stringent initially, which is why the subsequent stringency can also be significantly higher.

3.4 Culpability

Another fact about Betty in *Dangerous Party* also helps explain why her subsequent duty is only slightly more stringent. Not only does the initial stringency of the duty matter; *how* one initially disrespected the duty also matters. Compare the following two cases, for instance:

Recital (Work): Betty is invited to her son Albert’s ballet recital but skips it because she is behind on her work and could really use that time to finish some tasks so that her boss does not punish her.

³² For more on issues relating to agent-neutral versus agent-relative values and reasons, see Löschke (2021).

Recital (Bored): As in *Recital (Work)*, but Betty skips the recital because she finds ballet incredibly boring.³³

Suppose that it is wrong to skip the recital in both cases, and that the harm Betty will suffer in in them is comparable – for instance, the boss might make her do work as tedious as it is to watch children’s ballet. Still, I believe there is a difference in Betty’s secondary duties towards Albert because the wrongdoing in *Recital (Bored)* is more culpable. If Albert finds out about her reasons for skipping the recital, it seems to me that Betty ought to be more repentant and feel and display more remorse in the latter case than in the former. The reason is that the duty to attend one’s children’s recitals has a particular exclusionary profile. In particular, that one finds their recital boring seems excluded. It is not a reason for action that a parent should seriously entertain. By contrast, the reason not to fall further behind on work and avoid being punished by one’s boss is a relevant consideration. It might be outweighed in many cases, such as in *Recital (Work)*, but it is not excluded from being relevant to one’s deliberation. That is why skipping the recital in *Recital (Work)* is less culpable than in *Recital (Bored)*, and her resulting duties towards the child differ.

Someone sceptical of this distinction between excluded and outweighed considerations might instead capture the same verdicts by talking about how reasons for violating a duty can be better or worse, and that this is what matters to the culpability of the act.³⁴ On that view, fear of being punished at work is a better reason to skip out on a recital than fear of being bored by the recital, and therefore less culpable.

In either case, the point here helps to further reduce the stringency of Betty’s duty at t_2 in *Dangerous Party* as well. I have argued that going to the party because she is invited and wants to go, although wrong, does not involve acting on considerations she ought to have

³³ These cases are inspired by Martin (2021, p. 65).

³⁴ For instance, this fits with Alexander and Ferzan’s (2009, pp. 41-68) picture of degrees of culpability.

excluded from her deliberation. Alternatively, we might describe her as not acting on a very bad reason. This contrasts with versions of the case in which Betty seems significantly more culpable for her choice. Imagine, for instance, that someone who dislikes Albert encourages Betty to go to the party and kill him once he attacks her just because they dislike him. This is not the kind of consideration that ought to factor in her deliberation, I believe. (Alternatively, it is a very bad reason to act on). Thus, if she goes for that reason, her duty not to harm Albert should also be more stringent. However, the extent to which it is more stringent is still limited by the argument above.

Treating culpability as an additional factor affecting how stringent a duty can become also fits with the underlying explanation for the Continuity Thesis. It seems natural that how much one should weigh some value after disrespecting it also depends on how inappropriately one initially weighed it. This, in turn, depends on what considerations one let outweigh it.

To summarise the argument so far, I have suggested that Betty has a duty at t_1 not to go to the party, given the inherent value of Albert's interests. If we accept this, it is difficult to avoid the conclusion that her defensive rights will be *somewhat* compromised if she goes to the party. However, I have also suggested that the compromising effect will be minimal since the stringency of her duty at t_2 is only slightly increased.

4 Objections to the Duty to Avoid Trouble

Those who lean more towards the Lenient View (that is, the view that Betty's right to defend herself once attacked at the party is not compromised at all) might instead try to resist the suggestion that Betty has a duty to avoid trouble in the first place. If it is not wrong for her to attend the party (foreseeing that she will have to harm Albert), then at least the argument from the Continuity Thesis does not apply. Heidi Hurd spells out one version of this view in the following:

One need not alter one's otherwise legitimate conduct so as to avert another's wrongdoing. One does not do wrong if one does what is (otherwise) within one's rights, even if one can predict that by doing so, another will be moved to do wrong [...] (2016, p. 264).³⁵

There is something intuitively compelling about this. Indeed, the previous chapter can be understood as a narrower version of Hurd's claim here. The core idea, recall, was that Betty does not *owe* it to Albert at t_1 to avoid a situation in which she has to harm him at t_2 because there is wrongful intervening action by Albert himself. Hurd's claim is more radical because it suggests that we do not owe it to innocent third parties to constrain our otherwise permissible behaviour as long as another person will wrongfully harm them as a result. Still, we might consider a weaker version, which only applies to instances in which a person's otherwise permissible behaviour will only risk harm to the intervening wrongdoer himself. On this view, Betty not only lacks a duty owed to Albert not to trigger the defensive situation, but she also lacks the less stringent ("charitable") duty not to harm the inherent value of his interests.

I think even this narrower version of the claim is mistaken. We cannot so quickly completely rule out the existence of a duty to avoid trouble. In the remainder of the subsection, I aim to defend the duty against objections, though I will grant that some so far overlooked considerations might make the duty weaker and less prevalent than it may have seemed initially.³⁶

³⁵ See also, for instance, Bublick who argues that we should be free to "proceed on the assumption that others would not intentionally harm us" even when we know this is unlikely, and Howard's claims that "if the performance of my duty or exercise of my right happens to increase the likelihood of others' wrongdoing, that does not void the existence of the duty or right" (2016, p. 33). Doggett (2020) also holds a similar position regarding *allowing* others to do wrong, except when their wrongdoing would harm another person whom we owe a duty to protect from said wrong. Although he does not ultimately condone the view, White also admits that it is tempting to think that "in deliberating about what to do, one should not have to consider the possibility that others will behave impermissible as a result of the decision one makes" (2017, p. 218). See also Alexander's (seemingly even stronger) claim that "the innocent need not cede *any* interest to the culpable [...] the innocent should never be forced to retreat rather than use force, including deadly force, to resist culpable aggressors. Nor should they be restricted to proportionate force [...]" (1999, p. 1502).

³⁶ I admit it is difficult to persuade someone already committed to Hurd's initial claim or even the weaker version. They have such a radically different fundamental view on the value of culpable aggressors' interests that probably nothing I say will convince them. My claims are aimed chiefly at those on the fence about the duty to avoid trouble or those who are not already wholeheartedly committed to Hurd's claim.

To start, let us go back to my earlier claim that we would all be justifiably abhorred if we learned that Betty had decided to go to the party, knowing she would likely have to kill Albert, without taking much time to think about it – treating the decision as if it was not a moral “big deal”. The fault here is that Betty fails to show the kind of minimum reverence for the inherent value of Albert’s life that we can expect. To show it due respect, she must treat it seriously as a consideration against going to the party, and only go to the party if her reasons for going outweigh that value. I then suggested that the interest in attending a party does not outweigh that value.

One response is to deny that Albert’s life has value, or insist that the value is irrelevant to Betty, but this option seems too radical. An alternative response is to accept that Albert’s life has value, is relevant to Betty’s deliberation, and indeed that her reason to go to the party does not outweigh it, but deny that this grounds a *duty* not to go. Hurd herself has something like this in mind. In discussing cases similar to *Dangerous Party*, she appears open to the suggestion that staying home would be *supererogatory* (not required, but praiseworthy) or that going to the party would be *suberogatory* (not forbidden, but blameworthy).³⁷

At least the latter suggestion might explain the sense that Betty would be blameworthy if she went to the party as if it was no big deal, and indeed even if she went there despite having taken the value of Albert’s life seriously (but not seriously enough) into consideration. However, it does not go far enough. Intuitively, we could not only blame her decision but also demand that she not go and try to prevent her from going (within limits). Suppose that everyone (except Albert) knows that her attendance is likely to lead to her having to harm him seriously, but that (for some reason) we cannot stop Albert from getting into the party. Imagine then

³⁷ See, for instance, Hurd (2001, pp. 33539). For more on the notion of the suberogatory, see Driver (1992). A similar suggestion is found in Finlay (2015). In his discussion of conditional threat cases, he also suggests that it may be “virtuous but non-obligatory” (2015, p. 70) not to resist the initial, minor wrongdoing if one knows it will lead to more defensive harm being necessary. But he seems to go even further than this, suggesting that there is an element of obligation here – an (admittedly oxymoronic) “supererogatory duty” (2015, p. 71).

Betty's flatmate who insists that she stay home with her instead, or the guard outside the party who refuses to let Betty in. It does not seem to me that they are acting wrongfully here. However, if that is correct, her conduct seems more than blameworthy. She must be trying to violate a duty in order for demanding and enforcing to be appropriate.

Thus, critics should look elsewhere for objections to the proposed duty to avoid trouble in *Dangerous Party*. Instead of rejecting it out of hand based on a claim about intervening agency, they might agree that it is *prima facie* wrong for Betty to go and harm Albert but argue that I have so far overlooked some reasons *in favour* of Betty going to the party or additional reasons *against* her staying home (or both). Once we appreciate those reasons, moreover, we should conclude that there is ultimately no duty to avoid trouble in cases like *Dangerous Party*. Below, I canvass and respond to a few suggestions along these lines.

4.1 Costs Over Time

One argument is, again, inspired by Hurd:

To say that persons must take precautions against the known risks of others' wrongdoing when they can easily do so is to say that it is wrong to go for a causal Sunday drive, for surely it is predictable that one will find oneself in the path of a negligent driver. It is to say that it is culpable to walk home at night, for one certainly knows that one is at risk of being mugged. [...] If our liberty is justifiably limited by others' foreseeable wrongdoing, then our liberty proportionately decreases as crime increases and as the ripple effects of others' wrongdoing expand in reach (2016, p. 272).³⁸

As illustrated here, the impact of the kind of duty I have motivated above, if applied to the real world, is immense. Even if one finds the reasoning behind the duty attractive when applied to

³⁸ See Bublick (1999, p. 1445) for a similar worry regarding rape and compensation. If prospective rape victims would only be compensated if they took precautions against being raped, the *aggregate* cost of this requirement would be too high. One would never go out after dark, never walk alone, never drink alcohol among strangers, and so on.

Dangerous Party considered as a one-off, isolated event, the worry is that if we are *always* required to minimize or reduce the risk of having to defensively harm aggressors, then the total cost would be too high. We could never take casual Sunday drives, walk alone in the evening, and so on. More precisely, it would require that we live (overall) miserable lives only to avoid defending against those who would harm us. That seems both counterintuitive and absurd.

To respond, it is first important to highlight that the duty to avoid trouble should be understood as sensitive to the probability that a course of action will end in a threat situation. In reality, one seldom knows that one will encounter a threat. For instance, in most instances, there is only a *risk* (and perhaps not a great one in most cases) that walking home at night will end with an attempted mugging. This should also affect the stringency of the duty to avoid the potential trouble. The less likely, the less one can be required to sacrifice to further reduce the risk. Below a certain threshold of likelihood, it might not be required to incur even minimal sacrifices in order to completely minimise the risk of having to defensively harm an aggressor.³⁹ *Dangerous Party* is special (partly) because Betty *knows* that her attendance will end in a threat.

Second, even if the continuous application of the duty to avoid trouble would be too burdensome, the correct conclusion is not to reject the duty outright. Notice that the same problem Hurd seems to point to above is also found elsewhere in the literature, for instance, in discussions of our duties to aid those in need. Almost everyone agrees with Singer's (1972) claim that one should incur a small sacrifice – say, ruining one's \$50 trousers – to save a child if one happens upon a drowning child. However, the real world is full of people in need who could be saved by similarly small sacrifices (like donating some money to charity). If we apply Singer's reasoning every time there is an opportunity to make a small sacrifice to rescue a life,

³⁹ Similarly, many would say that even though driving a car always poses a risk of harm to pedestrians, it is permissible (at least partly) because the risk is so low.

the total cost would be immense. We would be required to donate almost all of our money and property.

Many – rightly, in my view – doubt that we must sacrifice so much to save strangers, even a great number of them. But the right conclusion to draw from this is not that we cannot be required to sacrifice *anything* to save someone in need. Instead, we should limit the *total* amount of costs that we can be required to suffer over the course of our lives to save strangers from death and other serious harms.⁴⁰

We should draw the same conclusion for the duty to avoid trouble. Even if Betty cannot be required to incur small (one-off) sacrifices to reduce the likelihood of a defensive harming situation whenever possible, that does not entail that she is not required to make *any* sacrifices for that sake over the course of her life. In this sense, the natural conclusion is that the duty to avoid trouble is an imperfect one.⁴¹

Now, this might open up for a problematic kind of moral offsetting. Could Betty go to the party and kill Albert and “make up” for this later by (say) taking a taxi instead of walking home more often over the next two months (thus reducing the chance of other defensive harming situations)? That seems counterintuitive. However, it seems plausible that, even though the duty to avoid trouble is generally imperfect, there can be certain “golden opportunities” to discharge it, which make it required to act on the duty on a particular occasion. For instance, if one *knows* that a particular action will give rise to a threat situation – as opposed to merely believing there is a chance of it happening – then it might be required to discharge one’s duty

⁴⁰ For similar views, see, for instance, Timmerman (2015, pp. 208-10), Woollard (2015, p. 158), Thomson (2021, pp. 1227-231), and Pummer (2023, pp. 130-55). This is not to say that determining *how* much total costs one can be required to suffer, or how to deal with the possibility of rescuing people easily after the threshold is met, is a simple matter.

⁴¹ See also Finlay (2015, pp. 71-2), Frick (2016, pp. 255-56), and Schwartz (2020, pp. 601-4). In discussing cases similar to ours, they all argue that there is a limit to the total amount of sacrifices over time that one could be required to suffer to avoid harming, or causing harm to, the other agent (either an aggressor or, in Frick’s case, someone who threatens to harm themselves unless the other agent does as he is told).

to avoid trouble by not performing that action. Furthermore, if one knows that a particular action will give rise to a situation in which one has to impose very substantial defensive harm, it might also be required to discharge one's duty to avoid trouble by not performing that particular action. *Dangerous Party* stands out as a golden opportunity in both ways.

In sum, we should not take the worry about the aggregate cost of complying with the duty to avoid trouble as a reason to reject the duty completely, nor as a reason to think it does not apply in *Dangerous Party*.

4.2 Rights and Dignitary Interests

Another objection to the duty to avoid trouble is that it does not fully appreciate the moral cost to Betty of staying home. So far, I have claimed that Betty's interest in being at the party does not seem sufficient to outweigh the inherent value of Albert's interest in not being killed or seriously harmed. But, as briefly highlighted at the end of the previous chapter, we are not simply asking Betty to forgo something that she *wants*. We are asking her to "concede what is [hers] by right" (Finlay 2015, p. 64). The moral cost to her of staying home is not equivalent to simply missing out on the party because she missed the last train there. It is worse because she would be prohibited from exercising her *right* to go to the party. Once we factor in that, the argument goes, we should weigh the cost to her of not going to the party more than the value of Albert's interest in not being killed or seriously harmed.

I am unconvinced. First, saying that Betty has a *right* to go to the party is question-begging if we take it to mean that she is permitted to do, of course. Instead, let me outline two alternative ways of understanding this proposal. One is that Betty, in an ideal sense, *ought* to be free to go to the party and *ought* not to be required to stay home, and we should therefore weigh the cost to her of not going more heavily than we should if her freedom to go to the party was jeopardised in some other way (say, because she would have to steal someone's car to get there on time). Thus, the cost to her of staying home should outweigh Albert's interests'

value.⁴² Another slightly different way of understanding the situation here comes from Margaret Raymond (2010). Applying her argument concerning cases similar to *Dangerous Party*, the idea is Betty has two different interests in going to the party. The first is her interest in going to the party and having fun. The second is a “dignitary interest”, which is a person’s “interest in being permitted to move about freely and to pursue those activities fundamental to a free society, *without being subjugated to the unlawful demands of another actor*” (2017, p. 323; emphasis added).

The dignitary interest seems clearly involved in *conditional threat* cases, where there is a deliberate attempt to coerce and restrict the victim’s freedom by threatening to do something worse.⁴³ It is less obvious to me that it is present in *Dangerous Party* since Betty only predicts that Albert will choose to decide to attack her if she goes to the party. However, Raymond would say that it is. She claims that the interest is at stake whenever a person foresees that their behaviour would cause another agent to do something wrong, because requiring her not to act for that reason would still mean that it is “the anticipated conduct of the wrongful actor that controls her behavior” (2010, p. 321). Thus, the interest is also present in *Dangerous Party*. Moreover, once we factor in that interest, the thought goes, Betty’s reasons for going to the party ultimately outweigh the value of Albert’s interests.

Both suggestions have problems. To use a different spin on an earlier case, suppose that Betty knows that her attendance at the party will motivate Albert to steal Cruella’s purse (perhaps seeing Betty will remind him that he wants to steal Cruella’s purse). But suppose also that

⁴² For an even stronger version of this kind of idea, see also Quong (2020, p. 150) who argues that the fact that a person could have easily taken some precautions to avoid a threat situation “is irrelevant” to the question of whether they may defend themselves against the aggressor, when and because these are precautions that the victim “*should not have to take*” (2020, p. 150).

⁴³ In fact, Venezia suggests an analysis of conditional threats along these lines. He says that someone threatened with a choice between giving up their wallet or being killed can defensively kill the mugger because they are defending an autonomy interest: “the victim’s freedom to act as she thinks fit, *without having to depend on the arbitrary will of another person*” (2022, p. 70). See also Steinhoff (2013)

she expects Cruella to shoot him dead when he tries to take it. Here, Betty ought to be free to go, and requiring her to stay home means allowing Albert's anticipated wrongdoing to restrict her freedom. Surely, however, it is still wrong for her to go knowing that Cruella is likely to kill him.

A possible reply is that, given my arguments in the previous chapter, Albert's agency vis-à-vis the lethal harm is not sufficiently voluntary if he does not expect the lethal defensive harm. Thus, Betty might *owe* it to him not to go after all. But, the reply continues, the two considerations offered here are only significant (or significant enough) to liberate the victim from a duty to stay home when not in competition with *rights* against harm. Thus, they only apply forcefully enough in *Dangerous Party*.

However, suppose we stipulate that Albert knows Cruella will likely try to kill him but is willing to take the risk. Then, as I suggested earlier, Betty does not *owe* it to Albert to stay home because of the harm he will otherwise suffer. Still, it seems to me that Betty should not go, knowing that Albert is likely to be disproportionately killed by Cruella, even factoring in the fact that she ought to be free to go and that it would amount to letting someone else's expected wrongdoing control her freedom.

Relatedly, two suggestions here miss out on arguably the most morally salient fact about *Dangerous Party*. In my view, the most promising argument for why Betty should not be required to stay home focuses on the fact that Albert will try to *kill her* specifically. But this is not captured by these two suggestions. They focus more generally on whether Betty ought to be free to go or whether her not going would mean that her freedom has been restricted by predicable wrongdoing. Yet for this to be true, it is immaterial whether the predicted wrongdoing is particularly wrongful or has anything to do with Betty in particular.

4.3 Degradation

So, let me outline a different suggestion which focuses on what seems more distinctive about *Dangerous Party*, namely that what would prohibit her from going is a predicable egregious assault on Betty herself.

Many wrongs and threats are not only wrong and (potentially) harmful but also *degrading* to the victim. According to one view, a degrading wrong is one that expresses or manifests a wrongful *attitude* towards the victim. It does so, moreover, when the wrong is “motivated or otherwise explained by [the] wrongful attitude” (Bazargan-Forward 2018, p. 149). More precisely, as Helen Frowe and Jonathan Parry put it, a wrong is degrading when it is motivated or explained by an attitude to the effect that the victim is morally inferior or “lacks equal moral status (e.g., full rights over her body or equal citizenship)” (2019, p. 108).⁴⁴ Examples include rape, slavery, and relentless bullying, which all (at least typically) tend to express that the victim lacks equal moral status or is morally inferior. By contrast, breaking a promise to a friend or stealing from a stranger need not express that the victim lacks equal moral status.⁴⁵

Some also believe that a threat’s degrading nature provides reasons for harming the aggressor which are not tied to its usefulness in averting the physical threat. For instance, many believe that it can be permissible to seriously harm a rapist even if one knows that doing so will not avert the physical assault. On one view, this is because, when faced with degrading threats, there is a reason to resist which is grounded in the value of standing up for one’s “honour” or dignity. More precisely, in forcibly resisting a degrading wrong, one is expressing or “asserting oneself as a person worthy of better treatment” (Frowe 2014, p. 113). Just as an act can be degrading in virtue of expressing the attitude that the victim has inferior status or value,

⁴⁴ In a similar spirit, Lazar argues that disrespectful wrongs are those which “communicate [...] one’s indifference to [the victim’s] status as a right-bearer” (2009, p. 297) or, in the worst cases, “rejects [the victim’s] moral status” (2009, p. 301).

⁴⁵ They might occasionally be degrading. For instance, if one is motivated to steal from a person because they are black and the wrongdoer believes they lack property rights in virtue of that.

an act of forceful resistance can be valuable in virtue of expressing the opposite attitude – “*a reassertion of oneself and one’s value*” (Ferzan 2018, p. 694).⁴⁶

If correct, considerations of honour might also tip the balance of reason in favour of permitting Betty to go to the party. It seems reasonable to think that Albert’s prospective *lethal* threat to her is degrading. In that case, Betty might have an additional reason to go to the party and harm Albert – it is an act of standing up for herself and asserting her moral worth. Although I think this improves on the arguments above, I believe it proves too much. Suppose that Betty has no intention or interest in going to the party but then learns that Albert will try to kill her if she goes. The argument suggests that she now has a reason to reconsider her decision and go to the party and kill Albert to assert her honour. Alternatively, consider again a version of TWO BUTTONS in which Albert is already trying to kill Betty, and she can disarm him by pressing Button A or kill him by pressing Button B. The present argument suggests that she has a reason to press B instead of A to assert her honour. Both implications seem counterintuitive to me.⁴⁷

However, there is another way that concerns about honour and degradation might apply to *Dangerous Party*. Instead of thinking that the degrading nature of a possible threat provides an additional (honour-based) reason to go and harm the aggressor, we might think that it affects how difficult it is to demand that victims suffer certain costs. In that case, the argument does

⁴⁶ The view is controversial. People inclined to think considerations of honour can (in different ways) justify violence include Benbaji (2005, p. 609), Statman (2008, p. 669), Frowe (2014, pp. 109-115), Kamm (2014, p. 81), Ferzan (2018, p. 694), Robillard (2018, p. 111), and Statman and Benbaji (2021, p. 6). For sceptics, see Bowen (2016) and Hereth (2022). I return to the suggestion again in Chapter 5.

⁴⁷ See also Ferzan (2018, p. 696; 2019, p. 738) who, though sympathetic to an honour-based justification in the rape case, does not think it helps justify standing one’s ground instead of retreating in cases like *Two Options*. In her view, the relevant notion of honour is not at stake when the victim only stands to lose so little (for instance, the right to remain in a specific location) if they are required to retreat. Admittedly, some have expressed, in conversation, that they think there *is* an honour-based reason to go to the party and harm Albert in the first scenario, but that it is outweighed by the tremendous harm involved. In that case, the main conclusion of the subsection likely still stands:

it is wrong for Betty to decide to go and harm Albert in *Dangerous Party*. However, I remain sceptical that there is even *a* reason to cause harm here. In my view, “honour” does not provide any reason to do what is *wrong*. Thus, if one agrees that causing harm is wrong in these cases, I do not think considerations of honour count in favour of it either. Put differently, there is no honour in deliberately causing wrongful harm.

not apply to the two examples just considered as there are no independent costs to Betty in avoiding harming the aggressor in those cases. In the original case, by contrast, Betty wants to go to the party. On one view, the degrading nature of the prospective threat can then make a moral difference by making it more difficult to demand that she suffer the cost of not going.

The reason is that, in a sense, the demand that she incurs this cost is also degrading. The explanation for why Betty has to suffer the cost of staying home would make essential reference to a degrading attitude towards her, since it is true only because of Albert's (prospective) degrading attitude. By requiring her to suffer the cost in this case, there is a sense in which morality demands that she give into this degrading attitude towards herself, or implicates her in the disrespecting of her moral status and worth. I believe this can make it more difficult to require her to suffer the cost even if it is otherwise relatively small.

The argument improves on the one offered by Raymond because it focuses more narrowly on the objectionability of "giving into" degrading treatment of Betty in particular, as opposed to giving into "wrongdoing" in general. At the same time, the scope of the argument is also more limited since it only counts against the duty to avoid trouble in cases where threats are degrading to the victim (not just wrongful). Plausibly, moreover, how much it counts against the duty also depends on *how* degrading the prospective threats are. Suppose, for instance, that she expects Albert to only steal her purse at the party but that it will likely escalate. But suppose the escalation is not driven by contempt or a belief that Betty is morally inferior. Instead, Albert will escalate more out of shock that Betty resists or because he thinks it is too late to give up trying to steal the purse now. Here, the claim that she has a duty to avoid trouble does not rely on the kind of severely degrading attitude that it would if the threat would be driven by contempt and a sense that Betty is morally inferior. Thus, it will be easier to demand that she also suffer the small cost in question.

Although the thought here is attractive, it does not entail that there is no duty to avoid trouble even when the prospective threats are very degrading. At best, I have suggested, the degradingness of the prospective threat can make it more difficult to demand that someone incur a particular cost to avoid the threat. In other words, it makes the cost to the prospective victim weightier in deciding what is right. But, in my view, many minor costs, such as having to forgo a party, will still not outweigh the badness of ending a prospective aggressor's life. Thus, even if we accept the argument sketched here, it does not entirely negate the duty to avoid trouble. At best, it limits it.

4.4 Socially Unjust Threats

So far, I have said little about the source of the potential threat to Betty and some may criticise the argument for the duty to avoid trouble because the examples are too abstract to generate concrete and reliable judgements. Once we add more realistic details, however, we should reject the proposed duty to avoid trouble. For instance, one of Hurd's objections to the idea that we "must take precautions against the known risks of others' wrongdoing when [we] can easily do so" is that this is:

to sign on to the claim that a woman who wears provocative clothes and behaves flirtatiously can be blamed for her own rape, for she could clearly anticipate that it would attract a man who would wrongly consider it a license to force himself on her (2016, p. 272).

Of course, the duty to avoid trouble would not blame the woman for *the rape*. At worst, it would only blame her for not having avoided a situation in which she needs to harm an aggressor when she could have easily avoided it. However, one might claim that even this is implausible. Similarly, Jeffrey Howard discusses a slightly different question from ours: whether it is disrespectful and wrong to knowingly increase the likelihood that someone else will act

culpably (regardless of questions about defensive harming). In arguing that this is not always wrong, he cites (among others) the following case:

Neighbourhood: “Jones, a black man, walks down a street frequented by white supremacists, increasing the likelihood that they will attack him” (2016, p. 32).

One might also think it is counterintuitive that Jones should be morally required to take a detour around that street just to avoid a situation where he needs to use defensive force to protect himself from white supremacists.⁴⁸

These kinds of cases, I admit, provide more powerful intuitions against the duty to avoid trouble than the less detailed *Dangerous Party* does. However, we should not take them to completely rule out the existence of a duty to avoid trouble. The reason is that these cases involve what we might call “socially unjust threats”. By that, I mean that they are motivated and explained by attitudes and beliefs that are part of more general social injustices, such as misogyny and racism. This can make a moral difference in several ways.

First, since social injustice is pervasive, those subject to one socially unjust threat on a particular occasion (such as the threat of rape or a racist assault) are likely to be victims of many similarly motivated wrongs and harms in the past and future. Insofar as they are worse off than others in virtue of this, it is more difficult to require that they suffer additional costs for the sake of avoiding trouble. Indeed, it might also be more likely that they will have already done enough to fulfil their imperfect duty to avoid trouble, for it is more likely that they will have already, voluntarily and involuntarily, incurred serious costs to avoid trouble over the course of their lives.

Second, socially unjust threats like the ones we have focused on are also those which are most clearly degrading. They are wrongs, or threats, that express, about the potential victim,

⁴⁸ In his discussion of *Dangerous Party*, Quong (2020, p. 149-50) also mentions cases involving domestic abuse and police brutality against African Americans.

“that he or she lacks equal moral status [...] in virtue of some morally irrelevant group-based characteristic, such as race, gender or disability” (Frowe and Parry 2019, p. 108). Thus, the degradation argument from above will tend to apply forcefully in these cases. Potentially, the socially unjust nature of a prospective threat can also make it even more difficult to demand that one incur a cost to avoid triggering it. To understand how degrading being required to give into a particular socially unjust threat is, it might not be sufficient to focus only on the threat itself. In asking Jones in *Neighbourhood* to avoid triggering a specific racist threat, we are also asking him to give into a *type* of attitude – a racist one – which is much more pervasive in society and implicated in a variety of wrongs he has suffered and will suffer. In that sense, the threat, or the degrading aspect of it, cannot be separated from the more general social injustice that operates in the background. This might make it even more difficult to demand that Jones incur a cost to avoid the particular unjust threat (especially if he is unlikely to avoid many of these other racially motivated threats and wrongs).

Lastly, although I was sceptical of honour-based reasons *in favour* of going to the party in Betty’s case, it might be more plausible in the case of socially unjust threats. By walking down the street, prepared to resist socially unjust threats, Jones is not necessarily just asserting *his* rights or standing up for *his* moral status. Refusing to take the detour might be seen (at least in part) as also asserting the rights and moral status of *everyone* disrespected by the wrongful attitudes in question and who could easily have been the victim of a similarly motivated threat. Jones is expressing not only that “I” am worthy of better treatment but also that “we” are worthy of better treatment. Insofar as his actions can be understood as standing up for the rights of a bigger group, the honour-based argument might also have more force here.

In sum, although the reasons cited here help further limit the scope of the duty to avoid trouble, they do not rule it out completely. There are still grounds for thinking that it is present in some cases, at least those not involving socially unjust threats.⁴⁹

5 Conclusion

I have argued that there is a duty to avoid trouble (grounded in the value of prospective aggressors' interests), and that this pushes us towards some version of the Moderate View, namely that imperfect victims' defensive rights can be somewhat compromised. At the same time, I also argued that we should only consider their rights *minimally* compromised, and that the scope of the duty to avoid trouble is further limited by some of the considerations we looked at in the last section.

What does it mean in practice? First, regarding scope, the duty to avoid trouble will be most clearly present when (i) the victim knows or judges it very likely that an attack will occur and they will have to use defensive force, (ii) the victim has not already done enough to fulfil the imperfect duty, (iii) the prospective threat is not of the most degrading kind, and so on. Second, regarding the degree of compromise, we can start by comparing of a perfect victim analogue of the victim, such as Betty in *Two Options*. Then, we should determine what sacrifices to avoid harming Albert that she can be required to suffer. For instance, would she be required to jump out of a window and break her own legs instead of *killing* Albert; would she be required to do so instead of *paralysing* him; or would she be required to bruise her ankle

⁴⁹ Massimo Renzo, in conversation, offered a different objection to the duty to avoid wrongdoing. Suppose that, in *Dangerous Party*, Betty alerted the police that Albert was likely to attack her at the party. Then, she goes to the party, and when Albert tries to attack, the police arrest him and send him to prison. Given that prison time for attempted murder can be very harmful, Betty could have easily spared Albert from serious harm by staying home. However, Renzo suggests that Betty has not violated a duty to avoid trouble. Indeed, there is nothing problematic about her actions. I am inclined to agree, but I think it is because we involve the state and punitive reasons. It is the job of the police to arrest criminals. But I do not think such reasons justify Betty's behaviour in the original case, for I am sceptical that Betty is permitted (or able) to act on the punitive reasons that the state has in this case.

instead of killing him? And so on. Whatever we conclude is appropriate to require of this victim, the Moderate View commits us to hold that Betty in *Dangerous Party* can be required to sacrifice somewhat more than this. But, as argued, we are only committed to holding that she can be required to sacrifice *a bit* more. Thus, the duty to avoid trouble only entails minimally compromised rights for those who violate it.

The Irrelevance of Risk-Taking

1 Introduction

Until now, we have focused primarily on the rights and interests of aggressors as potential reasons why an imperfect victim's right to self-defence is compromised. Here, I want to focus instead on facts about the *victim* as potential reasons to think so.

According to what I call the Risk Argument, people who knowingly and voluntarily expose themselves to risks of being harmed ("risk-takers", for short) have less stringent rights not to be harmed than those who do not expose themselves to such risks ("non-riskers", for short).¹ They are therefore permitted to use less defensive force to protect those rights compared to what non-riskers would be permitted to do to protect their rights. More precisely:

- (1) The stringency of the right about to be violated determines how much defensive force it is proportionate to use.
- (2) Risk-takers have less stringent rights not to be harmed than non-riskers.
- (3) Therefore, risk-takers are not permitted to use as much defensive force to defend themselves as non-riskers are permitted to use.

If correct, the Risk Argument provides a further reason for thinking that imperfect victims have compromised rights to self-defence. By walking through dangerous parks and neighbourhoods, failing to retreat from dangerous situations, and generally failing to take precautions, imperfect

¹ Note, for simplicity, I will often simply talk of "risk-taking" instead of specifying "knowingly and voluntarily" each time.

victims expose themselves to increased risks of harm. Thus, they negatively affect the stringency of their rights against harm and, so, the right to protect those rights.

The chapter argues against premise (2), but doing so is easier said than done, as we will see. Before starting, however, I want to mention one caveat. I am setting to the side those who voluntarily expose themselves to the risk of harm by performing morally required or supererogatory acts. A person who helps another person escape a wrongful abduction might expose themselves to a risk of being harmed, but they are acting heroically. A police officer who does the same might act on an obligation. It is plausible that these risk-takers do not negatively affect the stringency of their rights in part because of the deontic status of their risky actions.²

Instead, I focus on people whose risky behaviour is best characterised as morally neutral or wrong. There is nothing heroic or obligatory about (say) taking a shortcut through a more dangerous park at night instead of walking through a safer but slightly longer one. At best, it seems morally neutral.³ In other cases, the risk-taking might also be wrongful for different reasons. For instance, if the risk-taker knew that he significantly increased the chances of him being violently mugged, thereby causing his mother serious emotional distress. Alternatively, the behaviour might be wrong because the person is simultaneously trespassing on private property when taking the risk. The question is what, if any, moral consequences this kind of risk-taking should have for the stringency of their rights against harm.

² Similarly, see, for instance, Eyal (2007, p. 3) who claims that risk-takers do not forfeit claims to state compensation if they took the relevant risks as part of discharging their duties or doing something supererogatory. For an argument that heroic or supererogatory rescuers should in fact be prioritised over others when it comes to compensation and aid, see Eskens (ms.). Howard (2018, p. 678) also mentions the example of people whom the state incentivises to go to dangerous places abroad to help provide foreign aid and who end up kidnapped. One reason for thinking that they have not forfeited the right to be saved by the state is that they are doing something supererogatory by helping us and the state discharge our duties to provide foreign aid. (Another reason, more in line with what Howard seems to think, is that the state has *put* them in danger. This is similar to the arguments in the next chapter).

³ Though, of course, we might also doubt this for the reasons laid out in the previous chapter.

The chapter proceeds as follows. Section 2 clarifies premise (2) – the Less Wrong Claim – and mentions some of the support it has received in the literature. Section 3 clarifies premise (1). Section 4 argues that our reasons to reject premise (2) are similar to some of the reasons some are sceptical of *Luck Egalitarianism*. Sections 5 and 6 outline another principle – the Easier to Justify Claim – which is attractive but seems intimately connected to the Less Wrong Claim. Section 7 outlines and criticises three possible ways of rejecting the latter claim without giving up on the former. Section 8 outlines a holistic approach to risk-taking, which helps see what is wrong with the Less Wrong Claim without entailing a rejection of the Easier to Justify Claim. Section 9 considers and responds to a potential objection.

2 The Less Wrong Claim

Let us start with premise (2): risk-takers have less stringent rights not to be harmed than non-riskers. We can flesh this idea out in more detail with the following claim:

The Less Wrong Claim: All else equal, someone who has voluntarily exposed themselves to a risk of a particular harm is *less wronged* by being made to suffer that harm than they would be if they had not taken such a risk or did so involuntarily.

The idea is that the wrongfulness, or the severity or seriousness, of violating a risk-taker's right is, all else equal, lesser than violating a non-riskier's (same) right. This is because voluntarily exposing oneself to a risk of being harmed is supposed to decrease the weight or strength of the risk-taker's right not to be harmed, thus making the seriousness or wrongfulness of violating it lesser. That is what is meant by saying risk-takers have less stringent rights against harm than non-riskers.⁴

⁴ I am following Quong's terminology here. By "the stringency of a right", he also means "the strength or weight of the right-holder's claim" (2020, p. 107). He also connects degrees of stringency with "degrees of wrongfulness" (2020, p. 106) and "how wrongful it would be for someone to violate [a] right" (2020, p. 119-20).

Many find the Less Wrong Claim in some form or another attractive.⁵ McMahan, for instance, considers a case in which two innocent people – let us call them Adam and Bill – are killed by reckless drivers. The only difference between them is that Adam took a huge risk by taking a nap on the side of the road while Bill was walking quite some distance from the road. McMahan says that “because [Adam] had assumed a risk that [Bill] had not, he is *wronged to a lesser degree* than [Bill], even though neither is liable to be harmed in any way” (2016, p. 198; emphasis added).⁶

Similarly, Lazar claims that “when you voluntarily expose yourself to the risk of wrongful harm, your claim against being killed as a result is somewhat reduced” (2015, p. 138). He follows this up with the following thought experiment and comment:

Suppose I am a mugger. I can mug the guy who chose to walk into the dodgy neighbourhood at night or head across town and mug somebody there. Obviously both acts are impermissible. But it is worse to harm the people who have done nothing to expose themselves to risk (2015, p. 139).⁷

Next, consider Adil Haque. In the contexts of war and voluntary human shields (that is, civilians who voluntarily place themselves in specific places to make it more difficult for the enemy to avoid harming civilians), he claims that:

⁵ For several of these citations, I am indebted to Hecht (2019b, pp. 65-69).

⁶ We also find similar sentiments in relation to compensation and punishment. Concerning compensation, many jurisdictions operate with doctrines according to which injured parties who assumed the risk of being injured (or contributed to their own injury by being negligent) may lose the claim to be compensated by the wrongdoer, or only have a claim to partial compensation. However, it is usually restricted to negligent injuries. Concerning criminal law, Bergelson (2005; 2009) has argued that a victim’s risk-taking or failing to take due care of themselves can make their rights less stringent, thus making the criminal agent liable to less punishment. For instance, if X punches V after V punched him first, X may be liable to less punishment than if he punched someone else, because V voluntarily assumed the risk of being punched in return (Bergelson 2005, p. 484). Though, see Katz (1990) for scepticism about the use of assumption-of-risk arguments in criminal law.

⁷ See his claim that “[i]f you voluntarily expose yourself to the risk of harm, then your vulnerability gives others weaker reasons for action than if you are vulnerable through no fault or choice of your own [...]” (Lazar 2015, p. 115).

[T]he fact that you can easily avoid some harm does not entail that you are morally liable to suffer that harm, let alone that you deserve to suffer that harm, only that inflicting that harm upon you is *less wrongful* (that is, easier to justify) than it would be if you could not easily avoid that harm (2017, p. 232; emphasis added).

He also continues to say that, even when killing a voluntary human shield is wrongful, it is not *as wrongful* as wrongfully killing an involuntary human shield because “some *impermissible killings* are *worse* than others” (2017, p. 232n61; emphasis added).⁸

Lastly, consider Draper who also focuses on noncombatant risk-takers in war, such as medics accompanying combatants, civilians refusing to leave dangerous areas despite being warned, etc. Without citing anyone, Draper argues that “several just war theorists” have suggested the (quite radical) view that an innocent person’s voluntary assumption of risk can obviate (i.e., remove) rather than just weaken their right not to be harmed or killed:

[I]n some wars one can find innocent bystanders who, in virtue of putting themselves in harm’s way, are liable to being harmed or even killed. [...] In war it does not infringe upon rights to injure or kill those persons who, by freely and knowingly entering, or even failing to leave, an area that contains military targets, ‘assume the risk’ of being injured or killed (2016a, p. 154).

Draper’s own position also seems to lean towards this conclusion. For instance, in cases in which combatants are mistaken in thinking that they have to kill a risk-taker to stop a wrongful threat, Draper asserts that, although “assumed risk will not provide a reasonable justification for the infliction of harm, [...] it will nevertheless show that there is no infringement upon rights” (2016a, p. 159). To illustrate, consider a case in which a journalist disguises herself as an insurgent, joins a group of insurgents, and is killed because she was mistaken for a genuine

⁸ Brown (2020, pp. 1371-373) also argues that the gravity of a wronging can depend on the magnitude of risk of being wronged that the victim voluntarily accepted. His argument focuses on risk-taking and deception on the dating market.

insurgent. Draper claims that the killing was unjustified (insofar as the journalist was not an aggressor) but probably excused. However, “unlike most excusable killings, killing her *does not infringe upon her right* because she assumed the risk of being mistaken for an insurgent by choosing to be embedded with insurgents” (2016a, p. 159; emphasis added). Thus, Draper’s view also seems to be that assumption of risk can indeed remove one’s right not to be killed.⁹ This is a very radical version of the Less Wrong Claim.

Before I continue, let me note that there is a related but different claim that one might embrace:

The More Wrong Claim: All else equal, someone who has taken steps to minimize the risk of a particular harm is *more wronged* by being made to suffer it than they would be if they have not taken such steps.

According to the More Wrong Claim, it is extra or especially wrong to harm someone if they have taken steps to avoid it. Imagine, for instance, in Lazar’s mugging case above that the non-risker was in fact so reluctant to be mugged that he drove across town to cross through a safer park there. Perhaps it is extra wrong to mug him in that case because he has already taken steps and incurred costs to avoid that harm. Maybe this also affects proportionality. But I will leave this to the side here.

The position I and those above are interested in is the Less Wrong Claim. This is not the claim that taking steps to avoid a risk of harm increases how wronged one will be by being harmed, but rather that someone’s voluntarily exposing themselves to a risk of harm *reduces* the reason not to impose it and thus makes it less wrong to harm them.

⁹ Though, see Section 7.2.

3 The Proportionality Claim

The first premise of the Risk Argument is the Proportionality Claim, which ties the amount of harm that it is proportionate to use in self-defence to the stringency of the right that will otherwise be violated. This is, for instance, precisely what Quong (2015, p. 166; 2020, p. 112) claims. He argues for the Stringency Principle, according to which *only* the stringency of the right to be violated matters for proportionality. Contra suggestions made by many others, Quong does not believe that degrees of culpability matter for proportionality.¹⁰ The most crucial point here, however, is that the stringency of the right matters for proportionality. For the purposes of this chapter, it is possible that degrees of culpability also matter.

The view that the stringency of the right determines (in full or in part) how much defensive harm is proportionate is compelling. The proportionality constraint is meant to ensure that defensive harm does not exceed “the value of what is being defended” (Quong 2020, p. 109). For a certain amount of defensive force to be proportionate, the thing it aims to defend must be morally important enough compared to that defensive harm. Furthermore, how morally important it is to defend something also seems naturally tied in important ways to how wrongful it is to violate it. The stringency of a right, then, connects the ideas of how wrongful a rights violation is and how much harm it is proportionate to impose to prevent that violation.

Of course, I have said little about what determines the stringency of a right – apart from noting that, according to the Less Wrong Claim, whether one voluntarily exposes oneself to risks is relevant. For Quong, one of the essential elements to determining a right’s stringency is the *severity of harm* that will befall the victim if the right is violated. Others disagree. Imagine a case of an overdetermined killing. Suppose that X is about to kill V, and if V defends himself with lethal force, another person, Y, whom V cannot stop, will kill V immediately afterwards.

¹⁰ For this view, see, for instance, McMahan (2005b, p. 776), Rodin (2011, pp. 82-84) and Øverland (2011b, pp. 126-27).

On a counterfactual account of harm, X's violating V's right to life will not be very harmful. It only hastens her death by a few more seconds. Is it *proportionate* to defensively kill X? Some say no because there is essentially no harm prevented.¹¹ Others say yes because V nevertheless prevents X from violating his right to life, even though he cannot prevent Y from violating the same right moments later.¹²

My view leans more towards Quong's view, as it coheres better with the interest-based theory of rights from the introductory chapter. But the details are not that important here, as we are mainly interested in whether a person's voluntary risk-taking affects the stringency of a right, in addition to whatever else seems relevant, such as the severity of harm. However, some may interject that if severity of harm is relevant to the stringency of a right, then the Less Wrong Claim is clearly true. People who voluntarily expose themselves to risks of being harmed disvalue the prospect of being harmed less than people who are not willing to take the same risks, and are thus *less harmed* if these things happen to them. That is why it is *less wrong* to impose such harm on risk-takers as well.

Some have indeed suggested that the stringency of a person's rights should be understood in a fine-grained manner that is sensitive to how much the particular person values not suffering the rights violation on a particular occasion.¹³ But I set this way of grounding the Less Wrong Claim to the side here. First, it is not evident that differences in people's willingness to take certain risks reveal differences in how much they (dis)value certain outcomes. It may reveal differences in their attitudes towards risk itself.¹⁴ Second, at least concerning the most serious rights, such as rights to life, bodily integrity, and so on, we should be sceptical of

¹¹ See McMahan (2016, p. 209).

¹² See Frowe (2016, p. 165), and Ferzan (2018, p. 686n4). Note that at least Ferzan seems to accept *The Proportionality Claim* for she seems to tie her judgement here to the idea that it is not *less wrong* for X to violate V's right to life, even though the harm of death in the circumstances is minor.

¹³ See, for instance, Thomson (1977) and Brennan (1995).

¹⁴ See, for instance, Buchak (2014) and Stefánsson and Bradley (2019). Hecht (2019b, p. 76) also makes this point.

making stringency vary according to subjective differences in how outcomes are (dis)valued. The right of a depressed person not to be killed should not be considered less stringent just because they may disvalue the prospect of their own death less than the extent to which others disvalue their own death.¹⁵ A more objective measure of interest seems plausible here. Third, even within the interest theory of rights, it is often held that the existence and stringency of individual rights depend on more than just the harm that would occur if a particular right was violated on a particular occasion. V's right that X refrain from stealing his copy of *The Hobbit* is not determined simply by how badly off V would be if it was stolen. He might not be very badly off since he has no intention of rereading it. The right is grounded partly also in V's more general interests in having the freedom to determine what happens to his property.¹⁶ Thus, I will set aside this harm-focused way of arguing for the Less Wrong Claim. I return to other reasons in favour of it below.

To summarise, the Risk Argument consists of the Less Wrong Claim and the Proportionality Claim, which together suggest that imperfect victims, insofar as they are risk-takers, have compromised defensive rights. Naturally, one could avoid its conclusion by rejecting the Proportionality Claim and hold that only the amount of wrongful harm about to be imposed matters for proportionality. Together with a rejection of the claim that risk-taking (necessarily) reduces how badly off someone will be if the risked harm occurs, one may avoid the Risk Argument. However, this is not the best way to avoid the conclusion. First, I believe the Proportionality Claim has merits over the alternatives. Second, and more importantly, we have independent reason to reject the Less Wrong Claim because the claim is morally objectionable in itself, considered apart from questions of proportionality.

¹⁵ Hecht (2019b, p. 77) and Lazar (2009, p. 298) use this kind of example.

¹⁶ I modify this example from Raz (1994, p. 48). As he also says, rights can also be contrary to our (narrow) interests. I have a right to keep ownership of my cabin even if it is more of a burden for me (there is much upkeep required, and so on).

4 Luck Egalitarian Worries

To appreciate how the Less Wrong Claim is objectionable alone, we can begin by noting its similarity with (at least some versions of) Luck Egalitarianism. According to one way of putting it, Luck Egalitarianism distinguishes “between permissible inequalities which reflect voluntary behaviour, and impermissible inequalities which reflect involuntary behaviour” (Lang 2021, p. 169).¹⁷ Its key insight is that some inequalities seem especially objectionable because they did not result from people’s free *choice*, were not within their *control*, or are outcomes for which the people cannot be held *responsible*; while other inequalities, which are due to people’s voluntary choices, seem much less morally problematic.¹⁸ To illustrate, consider an example in which you bet some money on a horse race and lose:

You have no control over which horse comes in first, but you do have control – barring the possibility that you are a gambling addict – over the decision whether to have placed the bet in the first place. Since you have no control over which horse comes in first, placing a bet involves luck, but this is an example of option luck. Luck egalitarianism standardly considers brute bad luck, but not bad option luck, to warrant compensation (Lang 2021, p. 169).

The terms “option luck” and “brute luck” come from Ronald Dworkin (2000, p. 73). The latter has to do with parts of one’s circumstances one has no control over, while the former concerns outcomes that result from one’s free choices, including gambles. That is why, in the gambling case, it does not seem unfair or unjust to allow you to suffer the losses and not to impose a demand on others or society to ensure you are compensated. If you were a gambling addict, however, placing the bet might not be properly described as a free and voluntary choice, in

¹⁷ For classic versions of Luck Egalitarianism (or Luck Egalitarian-ish views), see, for instance, Cohen (1989), Arneson (1989; 2004), and Dworkin (2000).

¹⁸ As Lang (2021, p. 168) says, precisely what we focus on here (choice, control, responsibility) is not necessarily super important. The main point is that some outcomes that follow our choices are, in an important way, attributable to our agency and luck-free in a way that other aspects of our lives are not (such as our genetic make-up).

which case the loss seems more like bad brute luck. It is then more objectionable that you should suffer the loss alone.

The Less Wrong Claim and Luck Egalitarianism are similar in spirit here. For instance, both the horse betting person and the person in the dangerous park in Lazar's mugging scenario are voluntarily taking risks – gambling with their own welfare. It is terrible for them if the risk materialises (that is, if they lose the money on the bet or get mugged, respectively). However, in both cases, we can at least point to the fact that it was an unfortunate outcome of a voluntary choice.

It is therefore not surprising that the Less Wrong Claim inherits, in some fashion, some of the problems that are often counted against Luck Egalitarianism. One such objection, my primary concern here, is that Luck Egalitarianism justifies, legitimises or otherwise “gives into” existing injustices. For existing injustices shape the kinds of options that are open to us, and the risks attached to them. Consider:

Workplace Discrimination: Violet knows that coming out as gay to her colleagues significantly increases the risk that they will discriminate against her. She then suffers discrimination as a result of coming out to them.¹⁹

Is the discrimination that Violet suffers just? Surely not. She can legitimately complain, and we ought to help stop it. But the worry is that a Luck Egalitarian principle would say it is because it is the result of a gamble that she voluntarily took. Like the horse bettor, she gambled and lost – tough luck.²⁰ The same seems true of this mugging scenario:

Mugging: Vincent knows that walking through the dark park on his way home increases the chances that someone will mug him. He could take a slight detour

¹⁹ I modify this case from Go (2021, pp. 545-46) who in turn is inspired by Anderson's (1999) seminal critique of *Luck Egalitarianism*.

²⁰ Some might say that Violet's disadvantage is unjust because, though she can choose whether to come out, she cannot decide whether to be gay. However, I think this response does not get to the heart of what is problematic about the case. Even if she could have chosen that aspect of her identity, the objection stands.

around the park and avoid the risk. Still, he goes through the park, and Baxter then steals his possessions.

Even though it is just the result of a voluntary risky action, it seems unjust that Vincent is left without his possessions.

The Less Wrong Claim is concerned with the deontic status of the relevant actions – discrimination and mugging – as opposed to the justness of those actions’ outcomes. But it seems to suffer from a similar objection. Of course, it is, in one sense, *less* worrying than the objection facing Luck Egalitarianism. The Less Wrong Claim does not (necessarily) hold that discrimination and mugging do not *wrong* the victims. It only suggests that the victims are *less wronged* by it compared to how wronged non-riskers would be by having their rights violated in similar ways.

Yet this implication also seems objectionable. We should not think that Violet’s right not to be discriminated against and Vincent’s right not to be mugged are less stringent, and less wrong to violate, just because they voluntarily exposed themselves to these risks. This seems true regardless of what we think about the potential implications for proportionality. For instance, the Less Wrong Claim would also suggest that Ben and Baxter would be less blameworthy and liable to less punishment just because they targeted risk-taking victims. However, that also seems counterintuitive.

Hurd puts the worry I have in mind succinctly when she responds to Vera Bergelson’s (2005) suggestion that the criminal law should take a victim’s “assumption of risk” or “contributory negligence” to reduce their legal rights, and therefore the liabilities of the criminals:

[W]e should then find that a woman who wore a low-cut red dress to a rough bar reduced her rights against rape if she knew that in dressing provocatively she might incite the unwanted attentions of a drunken aggressor. And we should find that the jogger who entered Central Park at dusk knowing of the risk of being mugged was complicit in his own mugging; his voluntary

assumption of a known risk properly reduces the penalty imposed on the predictable assailant. And we should find that the woman who ran to the store for a jug of orange juice on New Year's Eve voluntarily reduced her rights against being struck by a drunk driver, for she knowingly and voluntarily invited those risks when she ventured onto the roads on that treacherous night. And we should find that a person who knowingly left her keys in her car invited its theft, thus reducing the penalty justifiably imposed on the car thief (Hurd 2005, p. 510).

Not everyone will agree that this is objectionable. Evidence suggests that many people support the claims that Hurd mentions here. This is (perhaps unsurprisingly) clearly illustrated in cases involving rape and sexual assault against women. As Jessica Wolfendale explains, numerous studies find that "psychiatrists, judges, jurors, and college students believe that [...] if [a woman wearing a revealing outfit] is attacked, she is partially responsible for her assault and her attacked less culpable" (2016, p. 600).

However, this way of thinking about these cases reveals, in my view, a misguided fixation on risk-taking or "personal responsibility". But the fixation is somewhat natural because risk-taking, or personal responsibility, seems morally important in many cases. For instance, Luck Egalitarianism gets things intuitively correct in many cases, such as the horse betting case. Thus, it is not entirely unreasonable to think that risk-taking also matters and has the same effect in other cases, such as *Workplace Discrimination*.

Similarly, I will argue below that a person's voluntary risk-taking can affect the stringency of a person's right for the purposes of *justifying* harmful actions. This is what I call the Easier to Justify Claim. I suspect that many within and outside philosophy may be attracted to something like the Less Wrong Claim because they view it as a natural correlate of the Easier to Justify Claim. Nevertheless, we should resist this view. The challenge, however, is to explain how they come apart. I will do that in the following sections after first outlining the Easier to Justify Claim.

5 The Easier to Justify Claim

The Easier to Justify Claim: All else equal, it is easier to justify imposing a harm on someone who has voluntarily accepted a risk of suffering that harm than it is to justify imposing it on someone who has not taken such a risk or who has done so involuntarily.

There are at least two ways in which it seems easier to justify harming a risk-taker than a non-riskier. One has to do with *selecting* whom should suffer some (wrongful) harm. Consider the following two cases:

Horrible Choice: An evil villain, Dylan, has kidnapped Veronica and Winnie and is forcing Claire to decide whom he will kill. The former was walking alone outside at night despite knowing that an evil villain was roaming around in the area. The latter was just sleeping at home.

Driving Choice: Julia is a reckless driver whose car is veering out of control. The vehicle will hit and kill one of two pedestrians – Oliver or Peter – and Julia has to decide whom to divert it towards. The former is lying beside the busy road because he carelessly decided to nap there. The latter is standing on the other side of the road because he has been sleepwalking and ended up there, something which has never happened to him before.

In both cases, it seems to me easier to justify harming the risk-takers, Veronica and Oliver, in the sense that Claire should pick the former, and Julia should divert her car towards the former. They should not choose the other person or flip a coin between them.²¹ Some may disagree that it is *impermissible* to flip a coin. All involved parties are innocent people, after all. But even someone who thinks this should, I believe, be open to holding that at least using a coin *weighted*

²¹ Lazar also uses the mugging scenario above to motivate a view like this. His primary concern is unjust combatants who are choosing whether to kill just combatants or innocent civilians. He argues that a difference in voluntary risk-taking helps explain why they should target the combatants rather than civilians, even though both options are wrong.

against Veronica and Oliver should be permissible or that, even though a coin flip is permissible, simply directing the harm to Veronica and Oliver is also permissible.²² In either case, the risk-taking factor makes a moral difference to whom should or may be harmed.

The second way someone's risk-taking seems to make it easier to justify harming them is by affecting what we might call the *justificatory threshold* for lesser evil justifications. For instance, by affecting how much good one needs to do (or how much evil one needs to prevent) to justifiably infringe someone's rights as a side-effect of pursuing that goal. Consider the following case inspired by Hecht (2019b, p. 60):

Collateral: A group of innocent civilians are about to be ambushed and killed by enemy combatants. We can bomb the combatants and prevent it, but doing so will foreseeably kill an innocent person, Tim, who voluntarily entered the dangerous battlefield because he wanted to take some dramatic photos for a news article.²³

Many believe that killing an innocent person as a foreseeable side-effect of saving some substantially greater number of lives can be permissible. The number of lives one needs to save to justify killing someone is the *justificatory threshold*. To many, including myself, it seems intuitive that Tim's voluntarily having risked being killed makes a difference to this threshold. For instance, suppose that to justifiably kill a non-risker one would have to save at least ten innocent civilians. Plausibly, it might be permissible to kill someone like Tim even if it would only save (say) nine lives (or perhaps even fewer) because of his voluntary risk-taking.²⁴

²² I am grateful to Theron Pummer for suggesting these slightly weaker conclusions here.

²³ Howard (2018, p. 686n7) notes that journalists in dangerous areas might be helping us discharge our duties to stay informed about what happens around the world. In that case, letting Tim's risk-taking count against him here may be inappropriate. For simplicity, I stipulate that he is not helping us this way. For instance, because his article is not very important or because the photos are not essential to it.

²⁴ This is the position that Draper and Haque above also seem interested in. They both think that risk-takers in war (for instance, voluntary human shields) should count less in the just side's (wide) proportionality calculations. See also Hurka's claim that "by placing themselves near [military targets, voluntary] shields arguably take upon themselves some responsibility for their own deaths and remove it from us, so their deaths count less against our attack's proportionality" (2005, p. 47).

Lastly, although not about harming, we can also appreciate the intuitive moral relevance of risk-taking in the context of *rescuing* someone from harm:

DROWNING: Sam is surfing despite being an amateur and knowing that a storm is closing in. The storm comes and he ends up in danger of drowning. Only Anita can save him now.

It is intuitively easier to justify not rescuing Sam. Because of his risk-taking, the maximum amount of harm that Anita would have to suffer to save him before it becomes optional seems lower than what the threshold is in cases involving non-riskers.²⁵

6 The Challenge

The Less Wrong Claim is a claim about degrees of *wrongdoing*, while the Easier to Justify Claim is a claim about how difficult or easy it is to *justify* an action. Can we not just reject the former without rejecting the latter and move on? The problem is that the two claims seem intimately connected. More precisely, according to what we may call the Justification-Wrongdoing Connection, the moral facts which make the Easier to Justify Claim also make the Less Wrong Claim true.²⁶

We see this idea in Quong's view on proportionality. As said, he connects the stringency of a right with how wrongful its violation would be but also, in the same breath, connects it with how difficult or easy "it is to justify infringing the right" (2020, p. 107). Indeed, to determine how much harm it is proportionate to use to defend a certain right, he thinks we should ask how much harm to an innocent bystander one would have to avert in order to justifiably infringe that right. Thus, on his view, how wrongful it is to violate a right and how

²⁵ See, for instance, Lazar (2017b, p. 21) for the view that those who are responsible for needing to be rescued have less stringent rights to be rescued. Though, see Miller (2020, p. 319) for some scepticism about this.

²⁶ In fact, this seems embraced by Lazar, Haque, and Draper. As I have highted in footnotes 21 and 24, they seem to support versions of *The Less Wrong Claim* in the course of defending the conclusion of *The Easier to Justify Claim* highlighted in the previous section.

difficult it is to justifiably infringe it go together. The easier it is to justifiably infringe a right, the less wrongful it is to violate it.²⁷ In that way, if risk-taking, in general, makes it easier to infringe someone's rights (the Easier to Justify Claim), it also makes it less wrongful to wrongfully harm them (the Less Wrong Claim).

Beyond Quong's claims, the Justification-Wrongdoing Connection also looks natural. When some action x is said to be easier to justify than another action y , one plausible reading of this statement is as saying that the reason against x ing weighs less or is weaker than the reason against y ing. That is why one can justifiably x by doing less good than one would need to do to justifiably y . One needs less weighty reasons in favour of x ing to outweigh the reason against doing it than one needs in favour of y ing in order to outweigh the reason against y ing. Similarly, if forced to choose between x ing and y ing, and all else equal, one ought to x because the reason against it is less weighty than the reason against y ing.

On this view, the Easier to Justify Claim holds that the interests of a risk-taker provide weaker or lighter reasons against harming them than the interests of non-riskers provide against harming them. That is why the same consideration (such as saving nine lives in *Collateral*) can justify harming the risk-taker but not a non-risker. It is also why, if forced to choose between harming one of them (such as in *Driving Choice* or *Horrible Choice*), one should choose the action that harms the risk-taker rather than the non-risker.

Such an understanding of how risk-taking makes the Easier to Justify Claim true naturally leads to the Less Wrong Claim, at least on certain views of degrees of wrongdoing. Mitch Berman and Ian Farrell (2011) suggest the following picture.²⁸ How wrongful an action is depends on the balance of reasons for and against the action. More precisely, it depends on the respective weight or strength of the reasons against it compared to the weight or strength of the

²⁷ This connection is also found in Haque. In the quote above, he seems to equate the two claims when he says, "less wrongful (that is, easier to justify)" (2017, p. 232).

²⁸ See also Hurka (2019, p. 44) for some similar views on degrees of wrongdoing.

reasons in favour of doing it (if there are any). This determines the extent to which the reasons against it *outweigh* (or outdistance) the reasons in favour of it. The less the reasons against it outweigh the reasons in favour of doing it, the *less wrong* it is to do it.

To illustrate, suppose you have promised to attend a friend's party, but you skip it because a different friend has invited you to dinner. In one case, it is another friend whom you see often. In another case, it is an old friend you last saw a long ago and who is only in town for a day. However, suppose also that your friend's party is important enough to make it wrong to break the promise in either case. Still, there is an intuitive difference in *how wrong* it is to break the promise. It seems more wrong to break the promise in the first case than it is to do so in the latter case. A natural explanation is the one just given: the balance of reasons leans heavier or further away from breaking the promise in the first case than in the latter. In both cases, the reason against breaking the promise is the same but, in the first case, the reason in favour of doing it is weak while, in the second case, the reason seems comparatively strong (it is, after all, a friend you rarely get to hang out with).

If risk-taking makes it easier to justify harming someone in the way outlined above – making the reason not to harm them weigh less so that it is easier for other considerations to outweigh it – then it seems to be less wrong to wrongfully harm risk-takers as well. All else equal, the overall balance of reasons will weigh less heavily against harming them since the reason not to harm them is less weighty or strong. Thus, the very moral mechanism underpinning the Easier to Justify Claim appears to support the Less Wrong Claim.

The Justification-Wrongdoing Connection also seems attractive if we look at *why* risk-taking is supposed to make it easier to justify harming someone. In my view, the most plausible explanations here mirror the ones offered by Luck Egalitarians. There are different suggestions here, as we saw earlier. If A takes a massive gamble at the horse racing track and loses, the outcome can be justified by A's voluntary risk-taking in different ways. However, a helpful

quote comes from Cohen when he says we should eliminate involuntary disadvantages because this is “disadvantage for which the sufferer cannot be held responsible, since it does not appropriately reflect choices that he has made or is making or would make” (1989, p. 916). The flip side is that voluntary disadvantage is less problematic because it reflects choices, and one can be held (at least partly) responsible for the outcome.²⁹

This also helps explain why the easier to justify claim seems true. It is easier to justify harming Tim as he wanders onto the battlefield because we can attribute the outcome, or the harm, to a large extent, to his own free agency. By contrast, it is more difficult to justify imposing it on a non-risker because the outcome is entirely alien to their own agency. But if that is true, the Less Wrong Claim should also be true. If Baxter wrongfully mugs Vincent in the park, it is also true that the outcome is something that Vincent is also partly responsible for, and which reflects his choices. After all, he knew it was likely to happen and could have acted to avoid or minimize the risk earlier. In other words, it is not something that is completely alien to his agency, as it would be if he was just sound asleep at home and Baxter broke into his house and took his money. Again, then, the two claims seem to go hand in hand. How can we reject one without the other?

²⁹ Alternatively, one might borrow from Scanlon’s (1998, pp. 256-67) value-of-choice view here and hold that it is not A’s having made a choice and being partly responsible for the outcome that matters, but instead that A had an adequate opportunity to choose whether to take the gamble or not. The same argument stands. If this underpins the Easier to Justify Claim, it also supports the Less Wrong Claim. By stipulation, the risk-takers in our examples also have adequate opportunities to avoid the harmful outcomes.

7 Some Suggestions

7.1 Luck Egalitarianism Revisited

One natural starting point in searching for a solution to this challenge is to look at *Luck Egalitarian* solutions to similar problems that cases like *Workplace Discrimination* and others create for them.³⁰

One approach is suggested by Zofia Stemplowska's (2009) discussion of similar problems. Although she finds it intuitive that a person's responsibility for a negative outcome could help justify that outcome, she highlights the same kind of worry we have focused on here:

Imagine a woman who, uncoerced, puts on a miniskirt knowing that this may expose her to a higher threat of sexual violence. Even if she knew that putting on a miniskirt would lead to a sexual assault – and thus bore agent responsibility for her disadvantage to some degree – it would not follow, for the same reason as above, that her responsibility justified the resulting disadvantage (2009, p. 245).

The “same reason as above” refers to Stemplowska's claim that “disadvantage is not justified by the presence of responsibility *if rights are being violated*” (2009, p. 246; emphasis added). More precisely, she argues that personal responsibility only has justifying force (that is, goes towards justifying a disadvantage) if the relevant choice was made against a background in which the person's rights are upheld and enforced.³¹ The reason is that if we consider personal responsibility to justify disadvantage (or at least counting towards justifying it) in rights

³⁰ I highlight only one here, but see Go (2021) for an overview, and critiques, of other possible solutions from within the *Luck Egalitarian* camp. One other view merits mentioning, however. Both Roemer (1995) and Eyal (2007), for instance, seem to be much less worried about disadvantages due to culpable or otherwise wrongful risks. Roemer mentions the example of getting injured after jaywalking which is “a gamble one need (and perhaps should) not have taken” (1995, p. 3). However, wrongful risks of this kind do not seem to make it *less wrong* to choose to harm them. I return to this in the next chapter.

³¹ According to Stemplowska (2013), it is also a mistake to think that early *Luck Egalitarians* (like Dworkin, Cohen and Arneson) held that responsibility for one's own misfortune *always* counts in favour of it being justified.

violation cases, then we fail “to acknowledge that [the outcome] is the result of an injustice in the first place” (2009, p. 245). In both Stemplowska’s sexual assault case and *Workplace Discrimination*, then, we should not consider the resulting disadvantages in any way justified by the agent’s responsibility because sexual assault and discrimination based on sexual orientation violate rights.

The same could be said against the Less Wrong Claim. It is not *less wrong* to mug Vincent or to discriminate against Violet because doing so violates their rights. Thus, questions about their own choices and responsibility are beside the point. However, the problem with this response is that it creates problems for the Easier to Justify Claim. In cases like *Driving Choice* and *Horrible Choice*, for instance, I suggested we should consider the risk-taker’s voluntary agency as a consideration in favour of harming them rather than the other potential victim. But in both cases, the risk-taker would suffer a rights violation. In *Driving Choice*, for instance, Julia, the reckless driver, will violate the rights of whomever she crashes into. Her job is to decide whose rights to violate.

Second, Stemplowska’s view relies on treating whether someone’s right is violated as established independently of facts about their risk-taking. But that is partly what the Easier to Justify Claim rejects. For instance, in *Collateral*, we are asking whether killing Tim to save (only) nine lives would violate or permissibly infringe his right not to be killed. According to the Easier to Justify Claim view, the answer to this relies on taking risk-taking as relevant to determining the stringency of his right. Thus, we cannot so neatly separate questions about rights violations from questions about risk-taking if we want to keep *The Easier to Justify View*.

7.2 Cost Imposition

Another possible response comes from Hecht (2019b). She focuses on cases like *Collateral* but argues that focusing on risk-taking is a red herring. On her view, we can give up both the Less Wrong Claim and the Easier to Justify Claim because another principle better accounts for the

intuitions we want the latter claim to account for. Hecht argues that what makes it easier to justify harming people like Tim in *Collateral* is that his risky behaviour (potentially) *imposes* a cost on others (the civilians). That is why it is appropriate to weigh his interests less when deciding what to do.³²

More precisely, the idea is this. Had Tim not ventured onto the battlefield, there would be no moral obstacle to saving the group of innocent civilians. The choice is easy: we should bomb the liable aggressors to save the innocent civilians. By entering the battlefield, Tim effectively makes the group of people worse off. For now, we need to consider that saving the nine civilians will also involve killing Tim. If there are only nine civilians and Tim is given as much weight as a non-riskier, then it is impermissible to save the nine. According to Hecht, we should weigh his interest less when deciding what to do, to neutralise or lessen the extent to which Tim can make the civilians worse off in this way. Once we do that, killing him in the course of saving the nine is (or might be) permissible. This also explains why it is not *less wrong* to harm in *Mugging* and *Workplace Discrimination*. The victims there do not impose costs on innocent people by taking the risks.

Hecht captures an important moral feature of many cases. However, it is wrong to completely disregard the Easier to Justify Claim and its focus on voluntary risk-taking as morally relevant. To see this, consider first:

Collateral II: As in *Collateral*, but the reason we can save the civilians is that the enemy combatants have revealed their position after spotting Tim on the battlefield and started moving away from him.

Tim's behaviour here is not a *cost* to the civilians. Had he not taken the risk of going onto the battlefield, we would not be in the position of being able to save the civilians at all. On Hecht's

³² Hecht (2019b, p. 68) notes that, despite all the talk of assumption of risk, Draper's view mentioned above might come closer to Hecht's view here.

view, we should not discount Tim's interests and instead conclude that it is wrong to save the civilians. Yet, it seems to me still easier to justify harming Tim in the course of saving the civilians than it would be if he had not voluntarily exposed himself to the risk of being killed. That is true even though it might be *even* easier to justify doing so when Tim's behaviour also imposes a cost on others.

Second, Tim's cost imposition seems most clearly relevant when he can be held responsible for it. But even when he is responsible for exposing himself to a risk, and this imposes a cost on others, he may not necessarily be responsible for this cost *qua* cost. Suppose, for instance, that in *Collateral*, Tim is aware of the risk he is exposing himself to by entering the battlefield, but he does not know that there are any civilians around (perhaps he has good evidence suggesting that everyone is evacuated). If he cannot be held responsible for the cost that he is imposing on others, then he is, in that respect, equivalent to someone who enters the battlefield utterly unaware that it is a danger or a cost to *anyone*, including himself.³³ Still, it seems to me easier to justify harming Tim than it is to justify harming this latter person. The natural explanation is that, although neither is responsible for the *cost imposition*, Tim is at least voluntarily exposing himself to the risk of being harmed. Thus, contrary to Hecht's view, the risk-taking factor itself also makes it easier to justify harming someone.

7.3 Unequal Weighing but Equal Rights

The challenge we are concerned with is grounded in the fact that the Easier to Justify Claim and the Less Wrong Claim seem to go together. But perhaps I was mistaken in thinking they

³³ Here is another, perhaps clearer example, inspired by one of Thomson's (1990, p. 229) thought experiments. Suppose X's light switch is faulty, and he knows there is always a risk that he will get a shock from turning it on. One day, an unpredictable set of circumstances makes it so that flipping the switch will cause both him and his next-door neighbour, V, to suffer a shock. Here, he seems responsible for the shock to himself but *not* the shock to the neighbour – even though the same responsible act (flipping the switch) causes both outcomes.

are connected in this way. In fact, rejecting *The Justification-Wrongdoing Connection* is quite popular in other contexts. Here is Frowe, for instance:

[T]he fact that it is harder to justify some modes of harmful agency compared to others does not entail that *unjustified* instances of those modes of harming are, other things being equal, morally worse than *unjustified* instances of other modes of harming. On the contrary: other things being equal, unjustifiably killing someone as a side-effect is plausibly as morally bad as unjustifiably killing her eliminatively, or in ways that make use of her (2021, p. 3470).³⁴

David Sussman also argues that torture has “an especially high burden of justification” (2005, p. 4) – even greater than killing – but says that this “is not to say that when torture is wrong, it is *worse* than wrongful acts of maiming and killing” (2005, p. 4n10). In fact, he claims that *nothing* in the view that it is harder to justify torture than it is to justify killing suggests that it is “more important for us to prevent torture than it is to prevent murder, or that torturers should be condemned or punished more severely than murderers” (2005, p. 4n10). Thus, two actions can differ in how easy or difficult they are to justify while being equally wrong when wrongful.

Perhaps this also applies to risk-taking: the Easier to Justify Claim does not give us any reason to adopt the Less Wrong Claim. One way of making the idea more precise goes as follows. A person’s risk-taking is relevant at the deliberative stage, and we should weigh their interests less heavily when determining right and wrong. However, if we nevertheless conclude it is wrong to harm them, having factored in their risk-taking, then this person’s right against being harmed is just as morally important and stringent as the right a non-risker has. It is equally wrong to violate it, and it provides the same degree of protection in terms of how much one

³⁴ Tadros claims something similar: from saying that manipulative killing is harder to justify than eliminative killing, it “does not follow that *wrongful* manipulative killing is worse than *wrongful* eliminative killing” (2018, p. 291). Focusing on blame, Quinn (1989, p. 289) also argues that it is consistent to hold that it is more difficult to justify *doing* harm than it is to justify *allowing* harm, and to hold that wrongful instances of either can be equally blameworthy.

can do to prevent others from violating it. In short, even if it is easier to outweigh a risk-taker's right, the right is just as serious and important when it is not outweighed.

However, a proponent of this view still seems to have the burden on them to explain why this is so, contrary to the challenge raised in Section 5.³⁵ As said, it seems very natural to suppose that the Easier to Justify Claim and the Less Wrong Claim go together. Furthermore, the view outlined here also does not go far enough in rejecting the Less Wrong Claim. According to it, a person's risk-taking is always morally relevant when figuring out what to do – at the deliberative stage. The view therefore seems to *legitimise* a certain way of thinking about risk-takers that I believe we ought to reject. Consider, again, Stemplowska's example of V who "puts on a miniskirt knowing that this may expose her to a higher threat of sexual violence" (2009, p. 245). Not only is it wrong to conclude that it is less wrong to sexually assault V, but it is also wrong to consider the reason not to assault her less weighty already at the deliberative stage. The potential sexual assaulter who takes V's risk-taking as grounds for weighing the reason not to assault her less than he would weigh the reason not to assault a non-risker is guilty of a certain kind of disrespect towards V *even* if he ends up doing the right thing. He has already given her risk-taking a normative significance it should not have. The problem with the current view is that it legitimises or justifies this way of deliberating.

8 Holism about Risk-Taking

To respond to the challenge created by the Justification-Wrongdoing Connection, let us go back to how it was set up initially. Once we understand the general underlying thinking more accurately, we can better pinpoint the mistake. I already went over a natural way of setting up the challenge earlier, but to make it more precise, let me borrow from Jonathan Dancy (2004, p. 41-2). In relation to reasons for action, Dancy talks about "modifiers" which are

³⁵ See also Steinhoff's (2006, p. 339) reply to Sussman.

considerations that are not themselves reasons but affect the weight or strength of reasons. There are two kinds: “intensifiers” and “attenuators”. The former increases the weight of a reason. As an example, Dancy uses a case in which you see someone drowning. This fact – they are in need – is a reason to help them (if you can). If you learn that you are also the only person who can save them, this is not an additional reason to help, but it can *intensify* the existing reason to help. The other kind of modifier decreases the weight or strength of a reason. As an example, Dancy incidentally uses a case like DROWNING. Here, he says that the fact that the person is at fault for being in need does not remove the reason to help them (or count as a reason not to help them) but it makes the reason to help less weighty or strong.³⁶

What underlies the Easier to Justify Claim, I believe, is the view that fact about someone’s risk-taking functions as an attenuator. It is easier to justify harming a risk-taker than a non-risker because the reason not to do so is *attenuated* and thus easier to outweigh. This, of course, naturally leads to the Less Wrong Claim as well. At least if we understand degrees of wrongdoing as Berman and Farrell (2011) suggested above. In cases like *Mugging* and *Workplace Discrimination*, since the risk-taking factor is present, the reasons not to harm the victims are also attenuated, thus making it *less wrong* to wrongfully do it.

The mistake here, however, lies in thinking about risk-taking atomistically – that is, as *always* attenuating the reason not to harm someone. Instead, we should understand the modifier more holistically. I borrow the terms holistic and atomistic, again, from Dancy. Atomists about *reasons* hold that “a feature that is a reason in one case must remain a reason, and retain the same polarity, in any other” (Dancy 2004, p. 7). Holists, by contrast, hold that “a feature that is a reason in one case may be no reason at all, or an opposite reason, in another” (Dancy 2004, p. 7). To illustrate holism, consider Dancy’s (2004, p. 74) non-moral example. The fact that something *looks* red to me is ordinarily reason to believe that there is something red in front of

³⁶ For more on modifiers, see, for instance, Bader (2016) and Kernohan (2022).

me. But if I know I have taken a drug which makes all blue things appear red and all red things appear blue, then the perception of something red seems to provide *no* reason for thinking that there is something red in front of me. On the contrary, I have reason to believe there is something blue in front of me.

The Less Wrong Claim and the Easier to Justify Claim come apart because we should consider the risk-taking modifier holistically. It attenuates the reason not to harm someone only in some contexts, not all. In fact, one can understand Stemplowska's view above partly as embodying this basic idea as well. She focuses on risk-taking as a *reason* to hold an outcome to be just as opposed to as a *modifier*, but the insight is similar. Risk-taking, or personal responsibility, only has justifying force – that is, only counts in favour of an outcome being just – in certain contexts: those in which no rights violations are involved in bringing about the outcome. Outside that context, facts about the harmed party's voluntary risk-taking have no justifying force. As she says of a case similar to *Workplace Discrimination*:

[I]t looks like claims about responsibility are irrelevant. That one is responsible in any of the above contexts (for the specific outcomes) *does not even suggest a weak reason* for why one's situation should be seen as just (2009, p. 243; emphasis added).

I agree, also when it comes to the Less Wrong Claim and risk-taking as a modifier. In these cases, claims about the prospective victim's risk-taking provide no attenuating force to the reason not to harm them. It just as wrongful to violate their right as it would be if they had not voluntarily accepted the risk in question. However, as also explained when discussing Stemplowska's view above, I do not think the relevant difference between cases in which risk-taking attenuates reasons and those in which it does not has to do with whether rights are being violated.

Thus, we need to look elsewhere for answers to the following crucial questions: what is it about the context in cases like *Mugging* which makes it such that risk-taking is not an attenuator, and what is it about the context in cases like *Collateral* and *Driving Choice* which makes risk-taking an attenuator? An answer to the first starts to reveal itself if we look closer at what exactly goes wrong if Baxter, in *Mugging*, considers Vincent's risk-taking an attenuating factor. Suppose he mugs Vincent and then tries to argue that Vincent should not be so angry and upset with him. He might say, "Fine, I should not have done it, but you cannot be *that* angry at me. You knew this was likely to happen and voluntarily accepted that risk". Vincent can reasonably shut him down, I believe, by saying, "That does not matter. You had *no reason* to mug me in the first place". In shutting down Baxter in this way, Vincent is pointing to the fact that the context was not one in which there were any reasons *in favour* of harming him, and that this is why the risk-taking factor does not attenuate the reason not to mug him either.

This leads to a natural answer to the second question as well. Hecht was on the right track in pointing to *costs* as key to explaining why the Easier to Justify Claim seems true. In my view, the risk-taking factor is an attenuator when there is a certain cost to *not* harming the risk-taker – that is, when there is also reason in favour of harming them. (Contrary to Hecht's view, though, that the risk-taker is responsible for this cost is not necessary, as I argued above). More precisely, someone's risk-taking has attenuating force in contexts involving *forced choices*. In *Collateral* and *Driving Choice*, the agent is facing a forced choice. Either allow nine people to be killed or kill one, or kill Oliver or kill Peter, respectively. In these forced choices, the risk-taking factor can be considered a relevant factor that weakens the reason not to harm one of the parties involved. That is why one may be able to justify killing Tim as a side-effect of saving only nine lives, and why it is permissible to divert one's car towards Oliver (the risk-taker) instead of Peter, respectively.

Admittedly, what counts as a forced choice in this respect is not obvious. Take *Mugging* again. It may be a forced choice of sorts. Suppose, for instance, that Baxter really wants to mug Vincent because it is fun or that Baxter needs to do it to afford an ice cream he really wants. In that case, Baxter is forced to choose between mugging Vincent or going without ice cream. Is that a forced choice sufficient to make Vincent's risk-taking an attenuating consideration? Surely not. But how and where do we draw the line? It is difficult to say, but one potential suggestion borrows from the literature on *aggregation*. Compare the following:

Life v. Paralysis: We must choose between letting one person die or letting n people suffer permanent paralysis.

Life v. Headaches: We must choose between letting one person die or letting n people suffer a mild headache.

Do the numbers count? Some think they always count, and some think they never count.³⁷ But, arguably, the most commonsense view is some form of Limited Aggregation: the numbers count sometimes but not always.³⁸ For instance, it seems possible that there is some n such that we should let one die in *Life v. Paralysis*, but there seems to be no such number such that we should let someone die in *Life v. Headaches*.

There are countless kinds of harm between headaches and paralysis, and it is difficult to determine where we should draw the line between kinds of harms that can aggregate to outweigh a life and those which cannot. However, one influential way of thinking about this focuses on whether the two kinds of harm that compete are comparably serious.³⁹ Paralysis is less severe than death but very close. That is why paralysis is a *relevant harm* compared to

³⁷ See, for instance, Halstead (2016) and Horton (2018) for the former and Taurek (1977) for the latter position.

³⁸ Though commonsense, this is not to say it is uncontentious – quite the contrary. See, for instance, Tomlin (2017a) for a summary of some of the many issues and a novel critique.

³⁹ See, for instance, Scanlon (1998, pp. 229-41), Kamm (1993, p. 146), Howard (2018, p. 679), and Voorhoeve (2014, p. 67).

death, and the numbers count. A headache, by contrast, is so insignificant compared to death that it fails to be relevant, and the numbers therefore do not count.

There is yet a deeper question on how to determine “relevance” here. I do not intend to take a stand on this. My point is that whatever one would fill in as an explanation here could plausibly also be used to help us determine what counts as a genuine forced choice in the context that we are interested in. If the harms involved – say, a mugging versus an ice cream – are not of comparable seriousness, then the risk-taking factor does not function as an attenuator. The choice facing Baxter is, in this respect, equivalent to a case in which there is no reason at all to harm Vincent. In *Driving Choice* and *Collateral*, by contrast, we are dealing with a choice between harms that are clearly relevant to each other (indeed identical ones).⁴⁰

Two related points worth mentioning emerge from this view of how risk-taking functions as a modifier. First, focusing on deliberation, facts about whether someone voluntarily exposed themselves to a certain risk of harm should be considered a “secondary consideration”. One first needs to consider the relevant harms at stake and get the unmodified reasons for and against harming them “on the table”, so to speak. The risk-taking factor should only be considered if the harms involved are comparably serious, thus giving rise to a forced choice. It is on this point that I believe the seemingly natural and commonsense approach to risk-taking, which leads to the Less Wrong Claim, gets it wrong. It treats risk-taking as a primary consideration. One should factor it in as an attenuator whenever it is true of the subject, and thus always deliberate about what to do by comparing the attenuated reason to other reasons. In other words, one’s deliberation about what to do, when it involves risk-taking persons, should always be shaped by that fact. By contrast, on my view, one should first establish that the choice one faces is difficult enough to merit taking such facts into consideration.

⁴⁰ There might be other better ways to draw the lines here.

Second, and relatedly, on this view, the risk-taking modifier can be understood as affecting only the comparative weight of a reason. It does not attenuate the weight that the reason not to harm someone has in isolation from other reasons – its “noncomparative” weight.⁴¹ It weakens the weight of the reason not to harm them only *compared* to the weight of reasons for creating or allowing other, comparably harmful outcomes. In other words, in *Collateral*, it is not the case that Tim’s risk-taking has reduced the weight of the reason not to kill him as such. It has reduced the weight of the reason not to kill him *compared* to the weight of the reason not to let the other civilians be killed. Similarly, in *Driving Choice*, the reason not to kill the risk-taker as such is not weaker. Instead, the weight of the reason not to kill him *compared* to the weight of the reason not to kill the non-risker is weaker.

However, a critic might still ask *why* the risk-taking modifier should operate in the way sketched here. After all, I said that it seems plausible that the Easier to Justify Claim is true because, for those who voluntarily expose themselves to the risk of harm, the harm can be attributed partly to their own agency. Yet this fact about the risk-taker is true regardless of whether it is imposed in a forced choice or for no good reason at all. The harm would still be partly due to a free choice they made, they could have easily avoided the harm, and so on. So why should the modifier not have force in all cases in which this is true?

The answer brings us close to a foundational moral claim about the nature of the value of human life and welfare. On one view, “modifiers play a role in determining to what extent [something] is to be valued [...]” (Bader 2016, p. 49). In that sense, when Vincent shuts down Baxter’s claim in the example above by asserting that the risk-taking was irrelevant since Baxter had no reason to harm Vincent in the first place, he is expressing a more fundamental view about the kind of value that human life and human welfare has. It is just not the sort of thing

⁴¹ Some, like Scanlon think no reasons have noncomparative strength – “[t]he strength of a reason is an essentially comparative notion” (2014, p. 111). On such a view, my claim in the text can be understood instead as saying that *whether* the risk-taking modifier attenuates a reason depends on which reason it is compared to.

we should value less unless there are also relevantly strong reasons *in favour* of doing so. In other words, the fact that a harm could be partly attributed to a person's own agency does not give us reason to care less about their interests or life unless there are also relevant reasons in favour of doing so. That will be the case in *Collateral* and *Driving Choices* where there is a serious cost to others if we do not value the risk-takers' lives and interests (comparatively) less. However, it will not be true in cases, like *Mugging*, where there are no relevantly serious costs to valuing the person's interests and life fully.

In sum, I have diagnosed the mistaken reasoning that makes the Less Wrong Claim seem like a natural correlate of the Easier to Justify Claim, and offered an alternative view of the risk-taking attenuator. On this view, someone's risk-taking modifies the reason not to impose a certain harm on them only if there are also powerful reasons to impose that harm.⁴²

9 An Objection

One potential objection to my view is that it cannot account for the (arguably) intuitive claim that muggers should pick risk-takers over non-riskers if they are going to mug someone, even though they should not mug anyone. Some may say this is where the Less Wrong Claim becomes attractive again. Take Lazar's mugging scenario again. The mugger, M, is choosing between targeting someone, V1, who is voluntarily walking through a dangerous park, or someone, V2, who decided to walk through a safer park instead. Surely, some may say, M should target V1 rather than V2. Alternatively, and weaker, they may say that it at least seems permissible for M to target V1 rather than V2 (but not vice versa) if he is going to target one of them.

⁴² The reader might recall the notion of exclusion from the previous chapter. In a way, the main thrust of the argument here can be phrased in those terms as well. The view is essentially that a person's risk-taking attenuates the reason not to harm them only when we also have non-excluded, competing reasons to harm them. Put differently, the exclusionary force of a person's right against harm demands that we exclude their risk-taking as a consideration in our deliberation unless there are also non-excluded reasons in favour of harming them.

My account seems incompatible with this. In contrast to *Driving Choice* and *Horrible Choice*, M is not *actually* facing a forced choice between two comparably bad outcomes. He is perfectly capable of not mugging either of them, which is what he ought to do. Since there is no actual forced choice, the risk-taking factor is, on my view, not operating as a modifier. In other words, because M should not harm any of them, the view cannot say that he ought to harm V1 over V2.⁴³

The sense in which M should, or is permitted to, mug V1 instead of V2 is sometimes called a conditional or contrastive obligation or permission. M should not mug anyone, but *if* he is going to mug someone, he should mug V1, not V2. Another example is a person who is deciding between committing a painful gratuitous murder or a painless gratuitous murder. Obviously, he should not murder anyone. But *if* he is going to do it, he ought to choose the painless option.

Although I am inclined to agree that M should mug V1 *instead* of V2, this should not convince us to embrace the Less Wrong Claim. Although there is a sense in which M should mug V1 over V2, it is also important to highlight that there is an even more important sense in which it is equally wrong to mug either of them. If we do not accept this, we return to the worry that V1 could only do *less* to defend himself than V2 could, would be owed *less* compensation, and so on.

At the same time, as said, it is tempting to say that the asymmetry in risk-taking between the two makes a difference. How do we square this with my account? The answer, I suspect, lies in the fact that there are two kinds of comparisons at play from M's perspective. When comparing mugging V1 to not mugging anyone, the risk-taking factor is irrelevant since there is no forced choice. But compared to mugging V2, the risk-taking factor seems relevant after

⁴³ This may be a concern that Lazar would press. As noted in footnote 21, he wants to say that unjust combatants should kill just combatants instead of civilians, even though they should not kill anyone.

all. That comparison is similar to the comparison involved in *Driving Choice* and *Horrible Choice*, after all.

The first comparison is the most important one for determining proportionality, and so on. But the second can also be relevant. We can understand M as having *excluded* the best options (not mugging anyone) from his deliberation, and in that sense, he is facing a “simulated” or “conditional” forced choice between whom to mug. In such cases, it might be fitting to be guided by what would be right if the forced choice was genuine (as in *Driving Choice*). But, to reiterate, since it is wrong to exclude the best option from his deliberation, and the option is still available to him, M does not escape the fact that V1 can do just as much harm to stop him as V2 could.⁴⁴

In short, mugging V1 is in reality as wrong as mugging V2 given that M is able not to mug either of them. That fact determines V1’s defensive permissions, probably also how harshly M ought to be punished, and so on. However, if M is determined to mug someone, then it might be appropriate to be guided *as if* he faced a genuine forced choice between V1 and V2.

10 Conclusion

In this chapter, I have looked at the Risk Argument for thinking that many imperfect victims – those who voluntarily expose themselves to risks of being wrongfully attacked – have compromised their defensive rights. In response, I have argued that the argument’s crucial premise – the Less Wrong Claim – relies on a mistaken, albeit natural, view of the moral relevance of risk-taking as always attenuating the reason not to harm someone. Once we move to a more holistic understanding of risk-taking as a modifier, we can see how the Easier to Justify Claim

⁴⁴ For more on conditional/contrastive obligations and permissions see, for instance, Pummer (2019; 2022) and Muñoz and Pummer (2022). I am grateful to Theron Pummer for pressing me on the objection here and discussing various possibilities. My answer to the objection is similar to the view offered by Pummer and Muñoz to make sense of cases like the gentle-murder case. They also highlight the significance of comparing options when the best option is *excluded* but still available to the agent.

is true without entailing the Less Wrong Claim. In that way, we avoid the conclusion of the Risk Argument.

Provocation and Duties to Protect from Harm

1 Introduction

The thesis has so far focused primarily on criticising possible arguments for thinking that imperfect victims in general have their defensive rights compromised. However, in this chapter and the next, I look at cases in which a victim's defensive rights seem more obviously and severely compromised because of their own choices. Here, I focus on provocation. Consider the following:

Provocation: Billy is taunting and humiliating Ali. He makes fun of Ali's career and looks, pushes him to the ground and spits at him. As Billy walks away with his friends, laughing, Ali gets up angrily and lunges at Billy with a knife, intending to kill him.

Although Billy's treatment of Ali is despicable and wrong, it does not give Ali a license to kill him. More precisely, I assume Billy has not forfeited his right against being killed by Ali. Thus, Billy is *prima facie* permitted to defensively harm, indeed kill, Ali in self-defence. But that seems too quick. The fact that Billy has brought about this attack on himself by provoking Ali seems to count significantly against him being permitted to defensively harm Ali. To account for this, one might refer back to the duty to avoid trouble from Chapter 3 as a reason for thinking Billy's defensive rights are compromised as Billy qualifies as a very culpable violator of that duty. However, given the nature of that duty, this proposal does not capture the sense that Billy's defensive rights are significantly compromised and that he wrongs Ali if he seriously

defensively harms him.¹ This chapter aims to make this intuition more precise and provide a theory that supports it.

The chapter proceeds as follows. Section 2 clarifies what I take “provocation” and “compromised defensive rights” to mean in this context. Section 3 looks at some legal perspectives on provocation and punishment and whether they can answer the defensive harm question. I argue that their relevance to our issue is minimal. Section 4 outlines the Partial Responsibility View according to which provokers compromise their defensive rights because of their partial, or shared, responsibility for the threat. I argue that it is vulnerable to counterexamples and fails to account for an intuitively relevant interpersonal reason why provokers compromise their defensive rights. Section 5 then outlines and defends *The Protective Duty View* according to which provokers’ defensive rights are compromised because they limit aggressors’ fair opportunities to avoid (defensive) harm and thereby incur duties to protect them from harm.

2 Two Preliminaries

2.1 The Concept of Provocation

Let me begin by distinguishing two kinds of cases which are often grouped under the heading of “provocation”. Compare *Provocation* with the following:

Provocateur: Bea wants to harm Alf. She knows he is unreasonably sensitive and arrogant and likely to attack her if she makes a joke at his expense in public. Bea then jokes about him in public because it will create an opportunity for her to defensively harm him.

¹ In support of the idea that Billy would wrong him by seriously defensively harming him, notice that it seems plausible that an *apology* would be in order not only for the bullying but also for the harm. Though, as will become clearer below, this depends somewhat on how the details of the case are spelled out.

Many treat the two cases as the same or versions of the same phenomena. For instance, Ferzan's (2013) theory of provocation and self-defence provides the same treatment of both cases. Similarly, in his brief discussion of provocation, Rodin mentions a case that combines elements from both cases: "a man who cruelly torments his long-time enemy with the intention of goading him to make an attack so that he can kill him and claim innocence by reason of self-defence" (2002, p. 79). Hurd also discusses "provocation" in the context of people who "intend to provoke a dispute so as to use it as cover to murder their adversary" and who "have their adversary's death as their purpose" (2016, p. 271).

However, we should distinguish between "provokers" and "provocateurs". The latter, of which Bea is an example, is best characterised by their *intentions*. They act *for the sake* of creating a situation in which they will have to harm the other person to defend themselves. One can seriously compromise one's defensive rights by acting this way, and I will return to that in the next chapter. But we need not assume that Billy fits this description for it to be intuitive that he compromises his defensive rights. Imagine that he does not want Ali to react violently and in fact hopes Ali will not do so. For instance, Billy is actually scared of Ali and only bullies him because he wants to look tough in front of his friends. He has nevertheless compromised his right to defend himself once Ali attacks.

The reason is that Billy is still a "provoker". I take this to mean something more similar to what legal scholars have in mind when discussing provocation. What makes provokers special is not their intentions but their *effect* on the other agent's self-control. The dominant legal view on provocation is that it provides a *partial excuse* for wrongdoing.² As one might put it, the partial defence is a concession to human weakness. We are imperfect beings. Sometimes, for understandable (but not justified) reasons, we can fail to live up to the moral and legal

² See, for instance, Ashworth (1976, p. 317), Dressler (1988, p. 469), Uniacke (1994, p. 13-14), and Macklem and Gardner (2001, p. 819). It typically only works as a partial defence for murder charges, but as I use it here, it can partially excuse wrongdoing more generally.

demands we ideally should live up to. This can happen, for instance, when others severely taunt us, humiliate us, and so on. It can cause even otherwise peaceful agents to become so enraged that their ability to reason and control themselves appropriately is undermined. A murder resulting from this, it is held, may be partially excused because it is the product of provoked lack of self-control.

It is because Billy affects Ali in this way that I call him a provoker and believe he compromises his right to self-defence. At least, we can stipulate that his bullying has this self-control undermining effect on Ali. Without that stipulation, the intuition that Billy's defensive rights are severely compromised seems much less powerful.³ Suppose, for instance, that Ali did not have his self-control undermined and instead coolly and calmly decided to lethally attack Billy in retaliation. In that case, it is much less clear that Billy's defensive rights are seriously compromised and that he will wrong Ali if he seriously defensively harms him. Instead, the case seems more similar to *Dangerous Party* from Chapters 2 and 3, except that Billy is independently culpable for his actions. But as I argued at the end of Chapter 2, with *Albert's Dangerous Party*, I doubt that one's defensive rights would be seriously compromised even when that is the case. So, in the remainder of this chapter, I focus on provocation in the self-control undermining, partially excusing sense.

In the legal literature, there is also a debate about how the lack of self-control must come about for the defendant to count as having been *adequately provoked*. For instance, suppose that V jumps the queue in front of X at the local shop. This sends X into a rage, and he stabs V to death. Was X *provoked* into killing V? Most lethal theorists would say no because even though X was caused to have his self-control undermined, the act in question was not egregious enough to count as a provocation. Theorists disagree on how to draw the line here.

³ Montague seems sympathetic to my view here. In relation to provocation, he says that it only seems wrong for a provoker to kill in self-defence if the provocation caused the aggressor to act "nonvoluntarily" (1989, p. 86).

For instance, Joshua Dressler argues that an act is sufficiently egregious if “the juror or ordinary, usually law-abiding person would be expected to become enraged” by the act as well (1982, p. 464). In that case, the anger (which undermines self-control) was itself legally excused.⁴ In contrast, Timothy Macklem and John Gardner argue that the provocative act must be so egregious that it makes the feeling of anger *justified*. On their view, what makes provoked wrongdoing partially excused is that it results from justified anger.⁵

My use of the term “provoker” diverges from this part of the legal concept. In the defensive context, whether the anger that caused the loss of self-control is justified or something a “usually law-abiding person” is likely to fall prey to is less important than whether the provoker *foresees* that the target is likely to lose self-control and attack. Imagine, for instance, that in the queue-jumping example, V *knows* that X is disposed to lose control and attack him if he jumps the queue. Perhaps a series of unfortunate events that he could not control has led X to be disposed here and now to react unreasonably and unjustifiably to the queue jumping. If V, knowing this, proceeds, then even though the “usually law-abiding person” would not react similarly and the outrage is not justified, I believe V nevertheless counts as a provoker with compromised defensive rights.⁶

Thus, the more relevant requirement for provocation, in the sense I am interested in here, is foreseeability. We find similar requirements in other discussions of provocation in the self-defence literature as well.⁷ It is essential for Billy’s compromising his defensive rights that he appreciates that his behaviour will likely cause Ali to undermine his self-control and attack.

⁴ However, “[i]f the ordinary law-abiding person would not become angry by a particular provocative situation, then the defendant’s violent response as a result of his anger is subject to moral condemnation, because the anger itself is deserving of moral condemnation” (1982, p. 465).

⁵ See, for instance: “An excuse for an angry action, *qua* angry, depends on the justification of the anger itself” (Macklem and Gardner 2001, p. 819).

⁶ What if X is fully culpable for the fact that he is likely to lose self-control so easily? See Section 5.3.

⁷ For instance, as we will see, Ferzan’s (2013) account makes foreseeing the threat absolutely central to how one compromises one’s defensive rights. Similarly, Hecht’s (2019a) account focuses on provokers as partially morally responsible for the threats they face which also requires some degree of foreseeability.

This requirement will almost always be met when one engages in severe taunting, humiliation, and so on because it is common knowledge that people can become furious as a result.

In the legal sense, by contrast, provocation is possible even when the agent is unaware that their actions are provocative. Imagine, for instance, a clueless tourist, unaware that a particular hand gesture is the most severe form of disrespect and humiliation in the country he is visiting. If the tourist enrages a local by using such a hand gesture, we might claim that the enraged local is partially excused on provocation grounds. But, as said, my use of the term “provoker” here is more restricted. The intuition that the victim’s defensive rights have been compromised seems most clear in those cases.

2.2 Compromised Defensive Rights

In what sense are the defensive rights of provokers like Billy compromised? According to Ferzan (2013), the provoker *forfeits* his right to use any defensive force against any threat he foresees is likely to result from his provocation. Thus, as long as Billy could reasonably foresee that Ali would lethally attack him, he lacks the right to protect himself against being killed. Importantly, Billy only forfeits his *right to use defensive force* and retains the *right not to be killed*. In that way, Billy is wronged if Ali kills him, and Ali is *liable* to be defensively harmed (in virtue of threatening to violate his rights) and so will not be *wronged* if Billy wrongfully defends himself with force necessary and proportionate to protect the right.⁸ The moral problem for Billy is simply that he cannot justify the use of force against the liable agent because of his problematic role in creating the threat to himself. On Ferzan’s view, Billy’s defensive rights are compromised in a somewhat similar way to how one’s standing to blame wrongdoers can be undermined by one’s hypocrisy.⁹

⁸ At least, he is liable to as much defensive harm as he would be if he directed his attack at an innocent person. Some might think he is liable to less defensive harm than a fully culpable aggressor since he is partially excused.

⁹ For more on the lack of standing to blame, see, for instance, Cohen (2006) and Wallace (2010).

I will revisit Ferzan's view in the next chapter. But it at least does not seem like a satisfactory account of how provokers compromise their defensive rights. First, if Ali is liable to defensive harm, then the view does not account for the sense that Billy's use of serious defensive force would wrong Ali. Second, it amounts to a version of what I in earlier chapters called the Draconian View insofar as it removes the right to use any defensive harm against any foreseen threat. However, suppose that Billy can prevent Ali from killing him simply by shooting him in the foot. Surely, imposing this comparatively moderate harm to avert the lethal threat is permissible.¹⁰

Hecht's (2019a) account of how provokers compromise their defensive rights is more plausible. She defends a scalar version of what I have earlier called the Moderate View, according to which how compromised a provoker's defensive rights are depends on how severe their provocation was. In particular, she argues that provocation affects the proportionality condition for permissible self-defence. The more severe the provocation, the lower the maximum amount of permissible force will be. Thus, while it may be impermissible to kill Ali, it can be permissible to use some lesser force, such as shooting him in the foot, to avert the threat. Others also support similar versions of this scalar version of the Moderate View when it comes to provocation and self-defence.¹¹ I will refine it more towards the end of the chapter, but it is a version of this view I will assume when I talk of provokers *compromising* their defensive rights in the rest of the chapter. The next question is *why* the defensive rights of provokers are compromised in this way.

¹⁰ See also Hecht (2019a, pp. 169-73) for an extended critique of Ferzan's view along similar lines.

¹¹ For instance, Draper argues that the amount of defensive harm that is proportionate to impose against a given threat can become lower "the less responsible the aggressor and the more responsible his potential victim" (2009, p. 81). Rodin argues that a provoker's "ability to claim the right of self-defence is diminished to the extent of their fault [for the creation of the threat situation]" (2002, p. 79). Farrell also suggests that there might be a "higher standard of scrutiny" (2013, p. 582) for whether provokers satisfy the proportionality condition (though he talks mostly about *provocateurs*).

3 Excuses, Desert, and Honour

One place to start looking for an answer is to look to the law and how provocation is thought to affect how severely a provoker's wrongdoing should be *punished*. This may offer some insight into the defensive context, but, ultimately, not much.

First, as said, the dominant view of provocation in the legal literature is that it is a partial excuse. So, one might think that Billy's right to defend against Ali is compromised because he is facing a partially excused aggressor. In fact, as touched on earlier, many believe that how much defensive harm an aggressor is liable to depends on how culpable they are for the threats they pose.¹² Nevertheless, even accepting this, the suggestion is too broad. On this view, it would not matter to Billy's defensive rights whether he was the one who provoked Ali or whether a *third party* provoked Ali into attacking Billy. Yet Billy's defensive rights only seem compromised when *he* is the provoker. Put differently, even if we believe that victims in general are permitted to use less force against partially excused aggressors, Billy's defensive rights seem even more restricted in *Provocation* because he is the provoker.

One way to account for this is to look towards the sizeable minority of legal theorists who view provocation instead as a *partial justification*.¹³ They build on the admittedly intuitive idea that harming or killing provokers often seems *less wrong* than killing innocent people. Although Billy has not made himself liable to be killed by Ali, there might be reasons *in favour* of attacking him that make the attack partially justified.¹⁴ Inspired by this, we might say that Billy's right to self-defence is compromised because he is not facing a fully wrongful threat but a partially justified one, and the proportionality requirement should recognise this by

¹² See, for instance, McMahan (2005b, 776), Øverland (2011b, p. 126-27), and Rodin (2011, pp. 82-84). For a critique, see Quong (2020, p. 102-7).

¹³ Another way is to keep the focus on the partial excuse but argue that it is only when the victim is *responsible* for causing this partial excuse that their rights are compromised. This is more along the lines of the Protective Duty View I go on to develop below.

¹⁴ See, for instance, McAuley (1987, pp. 150-56) and Rozelle (2005, p. 229). That said, see Eldar and Laist (2014) for a critique of the notion of "partial justification".

decreasing the maximum defensive force that is permissible. More accurately, keeping in mind the narrow notion of a “provoker” above, we might say that Billy’s defensive rights are compromised (or especially compromised) because he faces a threat which is both partially justified and partially excused.¹⁵

What reasons could there be in favour of attacking Billy? One possibility is that he *deserves* some of the harm Ali is threatening him with because he is guilty of seriously bullying Ali. A potential issue with this suggestion, however, is that it makes Billy’s defensive right only *minimally* compromised. Serious bullying is wrong, but we are not justified in imposing severe, harmful punishment in response to it. Thus, only a very minor part of the threatened lethal harm would be justified.¹⁶

Instead, or in addition, one might argue that Ali has an *honour-based* reason for attacking Billy.¹⁷ I mentioned this idea in Chapter 3. Although controversial, many argue that when people threaten our very dignity or honour, even if we cannot stop the physical threats, some violent resistance might nevertheless be justified because:

Concrete acts of resistance are needed in order to communicate to the aggressor, to ourselves, and to an actual or potential audience that we are not just passive objects to be trodden upon. By carrying out such acts, we reaffirm, or protect, our honor (Statman 2008, p. 669).

¹⁵ This is also more in line with the hybrid legal views which hold that the provocation defence (at least some instances of it) involves elements of both a partial excuse and a partial justification. See, for instance, Ashworth (1976, p. 307), Bergelson (2009, p. 89), and Berman and Farrell (2011, pp. 1065-77).

¹⁶ See also Ferzan (2013, p. 613) for more discussion of the limited relevance of desert to provocation and self-defence.

¹⁷ See, for instance, Horder (1992, §2) linking the 17th-century development of the legal provocation doctrine and prevailing conceptions of honour. Though not using the term honour, Berman and Farrell also talk about provoked agents having “agent-relative reasons to avenge a grievous wrong” (2011, p. 1094).

As I explained earlier, I believe this is best understood as the idea that there is some intrinsic expressive value in violently standing up for oneself and “asserting oneself as a person worthy of better treatment” (Frowe 2014, p. 113).¹⁸

If one buys into this idea, one might think that Ali’s attack on Billy is partially justified because Ali has an honour-based reason to stand up for himself (even though it does not fully justify killing him). Perhaps alone, or in addition to the desert-based reason, this could explain why the threat is more than minimally partially justified and, thus, why Billy’s defensive rights seem more seriously compromised as well.

However, this suggestion fails for another reason. In order to get justificatory force, even only partial force, from the honour-based reason to harm Billy, Ali’s attack must achieve fully or at least partially the expressive value of reaffirming his honour. In my view, this involves expressing or communicating one’s status as a *moral agent* because we have our particular moral value and dignity (at least partly) in virtue of our agenthood – that we are beings capable of deliberating and responding appropriately to reasons.¹⁹ But Ali’s attack fails to manifest this status precisely because it results from his lack of self-control. It results from his failure to deliberate and constrain himself appropriately – his failure to operate as a moral agent should. Thus, I do not think Ali’s killing of Billy succeeds in affirming his honour. This is also supported, I believe, by the sense that Ali should not be happy or satisfied if he succeeds. Instead, he should deeply *regret* that Billy’s bullying got the best of him and undermined his ability to control himself.

¹⁸ As mentioned in Chapter 3, people disagree on what, if any, role honour plays in justifying violence. Note also that few of those inclined to think it plays a role would think it applies to *Provocation*. They hold that honour only plays such a role only when the harm done is contemporaneous with the physical threat. But that is not the case here because Billy has started walking away.

¹⁹ See also Tadros’ claim that “people have distinctive and significant moral value and moral status because they’re able to critically reflect on values and reasons for action, and respond to them appropriately” (2020a, p. 233).

To conclude this section, my scepticism is not meant to suggest that the concerns raised here are entirely irrelevant to provocation and defensive harming, or the ethics of self-defence more generally. But they do not to help us adequately account for how and why Billy's defensive rights are seriously compromised.

4 The Partial Responsibility View

The more popular account of why a provoker's right to self-defence is compromised is what I call the Partial Responsibility View. There are different versions of it, but all are united in highlighting the claim that provokers are *partially responsible* for the threats or situations they face and therefore compromise their defensive rights. Hecht, for instance, claims that provokers foreseeably contribute to unjust threats (to themselves), and that this "contribution leads to forfeiture of part of her defensive rights" (2019a, p. 174). Similarly, Ferzan claims that provokers forfeit their defensive rights when they "consciously disregard the substantial and unjustifiable risk that their words or conduct will cause another person to attack them" (2013, p. 616).²⁰

The obvious counterexamples to the Partial Responsibility View involve the kinds of cases we have looked at so far in the thesis. Consider, for instance:

Sexual Assault: Britney decides to walk home alone at night rather than take a taxi, even though she knows this increases the likelihood of her being sexually assaulted. Adam then tries to sexually assault her in the park.

²⁰ See also Draper's claim that a (provoking) victim's defensive actions "may be impermissible in virtue of the potential victim's *share of responsibility*" (2009, p. 74; emphasis added) and Rodin's claim that a provoker's "ability to claim the right of self-defence is *diminished to the extent of their fault* [for the creation of the threat situation]" (2002, p. 79; emphasis added).

Park Assault: Beth decides to walk home through a more dangerous park rather than a less dangerous park, even though she knows this increases the chance of her being assaulted and mugged. Andy then tries to assault and mug Beth.

It is counterintuitive that Britney's and Beth's right to defend themselves against these attacks should be compromised, even if they could foresee that these attacks were likely to follow. Of course, if these, or at least the second one, are one-off scenarios, some compromising effect on the right to self-defence may be unavoidable given the argument in Chapter 3. However, even in that case, I argued that the compromising effect would be minimal.²¹ It is quite another thing to suggest that their defensive rights would be as (or close to as) seriously compromised as Billy's rights intuitively seem to be in *Provocation*.

Anticipating this objection, Ferzan claims that we should distinguish culpable/unjustified and non-culpable/justified behaviour from the victims. In her view, "we need an account of unjustifiable provocation that respects the liberty interests we've in walking streets, dressing immodestly [...]" (2013, 619). More specifically, Ferzan suggests that what separates victims like Britney and Beth from victims such as Billy is that the former's behaviour is justified by legitimate interests while Billy's behaviour – bullying Ali – is not. That is why only Billy is *culpably* responsible for contributing to the threat he faces, and only culpably partially responsible victims compromise their defensive rights.²²

There are other possible ways of drawing this culpable/non-culpable distinction. One might say that it is not just that Billy culpably contributes to a threat but, more specifically, that he has violated some of Ali's independent rights (for instance, the right not to be humiliated, pushed and spat on).²³ Alternatively, one might say that what protects Britney and Beth from compromising their defensive rights is that their behaviour is "otherwise legitimate": behaviour

²¹ Recall also the extra potentially justifying considerations that apply in *Sexual Assault*.

²² Rodin (2002, p. 79), Draper (2009, pp. 74-76), and Hecht (2019a, p. 168) all also restrict their attention to victims who act culpably in these kinds of ways.

²³ Bergelson (2009, p. 89) suggests something like this.

which would be morally unproblematic but-for the likelihood that someone else would react wrongfully to it.²⁴ Or because their behaviour is “morally protected”: behaviour which is important enough to be protected by rights even when others are likely to do wrong if one exercises those rights.²⁵

However, imagine now versions of *Sexual Assault* and *Park Assault* where everything is the same except that Britney and Beth are walking on Adam and Andy’s private property, respectively. In that case, the prospective victims seem culpably partially responsible as well. The behaviour is not justified by liberty interests, it violates the aggressor’s rights, and it is not “otherwise legitimate” or “morally protected”. Nevertheless, it seems that the original intuition about their defensive rights remains unchanged. Surely, their rights to defend against these aggressors do not depend on whether they are wronging, or have wronged, them in this way.

In the other direction, consider the following case:

Religious Provocation: Bernard is in a park about to give a sincere sermon on the moral superiority of his religion when he notices Alfie walking by. He knows Alfie’s family was recently murdered in a terrorist attack committed by a deluded person in the name of Bernard’s religion. He rightly suspects that overhearing a passionate sermon on the moral superiority of that religion is likely to send Alfie into a violent rage and cause him to attack Bernard.²⁶

If Bernard proceeds with the sermon and Alfie predictably attacks him because of a loss of self-control, then Bernard’s defensive rights seem to me seriously compromised. But it is not obvious that his behaviour would count as culpable according to the above ways of drawing the distinction. He does not violate any of Alfie’s independent rights, and the sermon is “otherwise permissible”. It may also count as “morally protected” insofar as he might be said to

²⁴ I borrow this term from Hurd (2001, p. 312).

²⁵ I borrow this term from Howard (2016, p. 32).

²⁶ The case is inspired by Howard’s (2019a, p. 218) *Atheist Provocation* case.

exercise the right to freedom of speech when giving the sermon and merely foresees the negative effect it is likely to have on Alfie.²⁷

Thus, the Partial Responsibility View still seems open to counterexamples. The deeper problem is that, in explaining provokers' compromised defensive rights, the view focuses primarily on the fact *that* the victim foreseeably contributed to a threat and then tries to make this principle match our intuitions by also bringing in talk about the independent *deontic* status of the victim's actions. Instead, in line with what I argued in Section 2.1., I believe the more relevant distinction here concerns *how* (not *that*) the victim foreseeably contributed to the attack. In both *Provocation* and *Religious Provocation*, the provokers compromise their defensive rights because they contribute to the threat by causing the aggressor to be *partially excused* for attacking them.²⁸ They foreseeably enrage and undermine the other person's ability for self-control. What sets the victims in *Sexual Assault* and *Park Assault* apart from Billy and Bernard, then, is that their behaviour did not contribute to the threats by making the aggressor partially excused for attacking them. That is why there are also relevantly different from Billy and Bernard in the trespassing versions of the cases. Even if their behaviour is independently culpable, trespassing does not make the aggressors partially excused for trying to sexually assault or mug them, respectively.

Proponents of the Partial Responsibility View could perhaps accommodate this by insisting that a victim's partial responsibility compromises their defensive rights only when the

²⁷ What counts as *morally protected* speech when adverse effects will predictably follow is complicated. See Howard (2019a; 2019b) for more on this. A complete analysis of the right to free speech might conclude that Bernard's speech should not count as "morally protected" after all. One argument for this that fits with Howard's general approach is that Bernard's speech is unprotected because Bernard owes Alfie a duty of non-endangerment not to engage in such speech given the circumstances. This is also the duty that my Protective Duty View below focuses on in explaining why Bernard's defensive rights are compromised if he provokes the attack.

²⁸ Note, this is compatible with thinking that there is still a difference in how severely compromised their defensive rights will be because there is a difference in how culpable Billy and Bernard are for contributing to the threats in this way.

aggressor is *also* partially (or fully) excused for attacking them.²⁹ However, it is not clear how well-motivated this restriction is. Partial responsibility operates differently in other contexts, after all. Suppose A wrongfully encourages B to attack C. He is partially responsible for the threat to C and seems potentially liable to defensive harm, even if B is not in any way excused for attacking C. So, why should partial responsibility in our context only matter when the other party is partially or fully excused for the attacking?³⁰

More fundamentally, this restricted Partial Responsibility View does not capture the intuitive significance of the fact that the provoker causes the aggressor to attack out of a loss of self-control. In my view, Billy's defensive rights are not compromised just because the following two facts happen to be true: (i) he is partially responsible for the threat, *and* (ii) Ali is partially excused for the attack. Billy's defensive rights seem compromised because he *makes* (ii) true by provoking Ali.

This is significant because it identifies an important wrong that Billy is guilty of. He *wrongs* Ali in a distinctive way when he causes him to lose self-control and attack him, thereby creating a situation where Billy has to harm him to protect himself. Moreover, I believe, it is because of this wronging that Billy also owes it to Ali to restrain himself seriously in defending against him. That is why his right to self-defence is seriously compromised.

In other words, there seems to be a fundamental interpersonal element to the explanation, which is not so obviously captured by a restricted Partial Responsibility View sketched above. I flesh this suggestion out in more detail in the next section. However, I concede it is possible for a proponent of the restricted Partial Responsibility View to incorporate what I have

²⁹ Rodin (2002, p. 79; 2011, p. 95), Draper (2009, p. 75), and Hecht (2019a, p. 176) all indeed only focus on examples in the provocation is severe enough to reduce the aggressor's responsibility as well. Only Ferzan's (2013) view explicitly holds that one may forfeit one's defensive rights also against fully culpably aggressors.

³⁰ This worry is particularly pressing for Hecht (2019a) who models the provoker's compromised defensive rights precisely on how indirect threats (like A) make themselves liable to defensive harming by being partially responsible for threats they contribute to.

just claimed. They might say that provokers are partially responsible for the threats they face in a special sense when and because they have caused the aggressor to be partially excused for attacking them. But we are still owed a deeper explanation for why this particular kind of partial responsibility, but not others, has a compromising effect on defensive rights. In one sense, the view I develop below could be understood as providing that deeper story. But I treat the views separately because mine highlights the significance of the relationship between Billy and Ali more than it does the comparative degrees of responsibility for the threat.

5 The Protective Duty View

Let us take a step back. Ideally, what we want from a complete account of the compromising effect that provocation has on defensive rights are two things. First, it should explain how the provoker's rights are compromised – the moral mechanism by which it happens. Is it, for instance, because they forfeit their right to self-defence in the way suggested by Ferzan or (say) because the fairest principle for distributing unavoidable harm would demand that part of the harm be suffered by the provoker?³¹ Second, it should explain precisely what it is about the provoker's behaviour which causes their defensive rights to be compromised in this way. For instance, is it simply that they foreseeably contributed to the threats they face, that they did so while also acting culpably or perhaps something else?

The Protective Duty View answers both questions in a novel and unified way. It holds that provokers compromise their defensive rights because they owe a stringent protective duty to the aggressor. Once we understand the general nature of this duty, we can also see why it is triggered by the victim's behaviour in cases such as *Provocation*.

³¹ The latter is what Draper (2009, p.76) would say.

5.1 Duties to Protect from Harm

Start with the first part. How is a provoker's right to defend themselves compromised by provoking the attacker? To answer this, the Protective Duty View takes inspiration from Massimo Renzo's (2021) account of manipulation and liability to defensive harming. Consider the following case:

Bomb: Brianna manipulates Aba into thinking that Vince is about to detonate a bomb, killing a group of innocent people, and can only be stopped if he is killed. Thinking it is his duty, Aba tries to kill Vince. To protect himself, Vince can either defensively kill either Brianna or Aba.

Intuitively, Vince should kill Brianna instead of Aba, although Aba is the one posing a threat and even though (in many people's views) he would be permitted to kill Aba if that was the only way of saving himself.

According to Renzo, this is not simply because Brianna is most responsible or culpable for the threat to Vince. Instead, a fuller and more accurate account invokes the idea of *secondary duties* to protect from harm. Suppose Brianna could tell Vince that he should direct the defensive force towards her instead of Aba. An argument for her having a duty to do so goes as follows. Brianna's primary duty was not to manipulate Aba in the first place. She cannot undo this, but by signalling to Vince, she could at least prevent the manipulation from resulting in Aba's death. She ought to do so since it is the next-best thing compared to not having manipulated him initially. Moreover, she could be required to do so even though it is *very* costly because her situation is the result of her culpably breaching the primary duty. That is why, in the original case, Vince should kill Brianna instead of Aba. He would effectively enforce Brianna's duty to Aba.

Let us call this a protective duty because it involves protecting the agent from harm that would otherwise befall them as a result of the primary duty violation. According to the

Protective Duty View, it is the same sort of protective duty owed to Ali by Billy which explains why Billy's right to defensively harm Ali is severely compromised. To simplify a bit, as I will return to the details in the next subsection, by provoking Ali into attacking him, Billy wrongs him and incurs a duty to ensure that Ali does not end up harmed because of the initial wrongdoing. The only way to ensure this is to refrain from using defensive force against Ali.³² Although it is very costly to comply with this demand, it is reasonable because it is the result of his culpably disrespecting a duty owed to Albert initially.

Thus, Billy's compromised defensive rights are not just grounded in the idea that he has forfeited the right to defend himself, as suggested by Ferzan, or that the fairest principle for permissible self-defence distributes harm in proportion to the comparative degrees of responsibilities among the relevant parties. It is the normative consequence of a wrong committed against Ali by provoking him initially.

This argument relies, of course, on the same kind of reasoning inspired by the Continuity Thesis that we say in Chapter 3 concerning the duty to avoid trouble. The difference is that the initial duty in *Provocation* is *owed* to Ali and more stringent than the duty to avoid trouble. That is why the resulting compromising effect on Billy's defensive rights is also more substantial. I say more about what grounds this directed, more stringent duty below.

First, let me address two possible initial objections to the Protective Duty View. One is that the account is incompatible with the claim that Billy is *wronged* if Ali succeeds in killing him. I said that Billy has not lost his right not to be killed by Ali. But, one may ask, how can that be true if Billy also *owes* it to Ali not to stop him? The answer is that Billy does not owe it to Ali to let him kill him, but owes it to him not to do what is necessary to stop this, namely

³² I write here as if Billy is not permitted to use any force against Ali, but I explain why the view is compatible with a version of the Moderate View in Section 5.3.

harm him. In that sense, he has a duty to allow his right to be violated by Ali.³³ However, he would obviously not be required to allow Ali to kill him if he could stop him without force.

Another objection goes in the other direction. If Billy has a right that Ali not kill him, then Ali is *liable* to be defensively killed. Thus, he cannot be *wronged* by Billy's defensively harming him. However, there are (at least) two responses to this. One is to claim that Ali is only *partially* liable to be defensively harmed.³⁴ He has lost his right against defensive harm *qua* being a wrongful aggressor, but he has a right not to suffer it because Billy owes him a protective duty which can only be satisfied by not imposing that harm. Another option is to claim that Billy's protective duty effectively *pre-empts* Ali's liability.³⁵ The grounds for his liability are there, but do not result in liability because of Billy's duty to protect him from harm. Both options avoid the objection.

To summarise, the Protective Duty View holds that Billy's defensive rights are compromised in the following way. After provoking Ali, Billy owes him a duty to protect him from being harmed as a result of the provocation, and this duty prohibits Billy from imposing (otherwise permissible) defensive harm on Ali.

5.2 Fair Opportunities to Avoid Harm

The Protective Duty View as described so far provides a framework compatible with different views on precisely what it is about the provoker's behaviour that compromises their subsequent defensive rights. It depends on what one thinks triggers the protective duty. Echoing the first version of the Partial Responsibility View, one could argue that it is triggered whenever one does something that foreseeably puts people, even aggressors, at risk of being harmed. By

³³ See also McMahan (2012, p. 295) and Tomlin (2017b, p. 249-50) for discussions of cases where people seem to have duties to allow their rights to be violated or infringed.

³⁴ This is inspired by Quong's (2020, pp. 145-48) view on the necessity condition.

³⁵ This is inspired by Renzo's (2021, p. 3492) view on manipulation cases like *Bomb*.

contrast, I believe that, in this context, to owe someone a stringent duty to protect them from harm, one must have made them more seriously vulnerable to harm.

To see this, we can put issues of self-defence to the side for a moment and observe how the protective duty operates elsewhere. Consider, first:

Footbridge: Amy and Bella are hiking in the mountains. Amy gets stuck on a faulty footbridge and is in danger of falling to her death. Only Bella can save her, but doing so is very costly and dangerous.

It is doubtful that Bella is required to save Amy because the cost and risk to her are very high. It is popular to think that our duties to rescue strangers cannot ordinarily demand substantial sacrifices from us. However, suppose we complicate the backstory:

Footbridge Lie: As in *Footbridge*, but Amy is stuck on the footbridge because Bella lied to her by saying it was safe to cross when Amy asked her before crossing it. Bella knew it was dangerous.

Intuitively, Bella owes it to Amy to save her even at substantial personal costs and risks. I believe this is explained by Bella owing her a stringent *protective duty* now. One way of putting the idea, which echoes Renzo's view above, goes as follows. Bella should not have lied to Amy initially. She cannot undo that, but at least she can step in to prevent that initial wrong from ending with Amy being severely harmed. That is why she has a stringent duty to help Bella. But that is not quite correct. Compare the following case:

Footbridge Arrogance: As in *Footbridge*, but Amy is stuck on the footbridge because, although Bella told her not to cross, Amy refused to listen. Bella is an experienced hiker and knew that it would be too dangerous for an amateur like Amy to cross, so she warned Amy about the danger. However, she rightly suspected that Amy was arrogant, overconfident, and jealous and that she would try to cross if Bella crossed. Nevertheless, Bella decided to cross. Amy followed and got stuck.

Although it might be all things considered wrong to cross the bridge, I do not think Bella owes Amy a stringent protective duty to save her at great personal costs now. One explanation for this is that her crossing the bridge is only wrong because of the inherent value of Amy's life. That is more similar to the duty to avoid trouble from Chapter 3. This might explain why Bella does not owe Amy a stringent protective duty now. So, what is required for a stringent, directed protective duty to arise, as it does in *Footbridge Lie*?

The answer cannot simply be what I suggested above, that in *Footbridge Lie*, Bella *wrongs* Amy by lying to her and therefore incurs a protected duty to ensure she does not end up harmed as a result of that wrong (lie). That would bring us back to the counterexamples facing the Partial Responsibility View. Suppose that, in *Footbridge Arrogance*, the mountain belongs to Amy and her family, that Bella is trespassing, and that Amy has commanded her not to cross. Still, it does not seem that wronging Amy by trespassing and disregarding her command is sufficient to ground a stringent, directed duty to protect her when she arrogantly tries to cross the footbridge and ends up in danger.

Instead, for there to be a stringent protective duty at t_2 , part of the reason why Bella owes it to Amy not to act at t_1 must concern the risk of harm at t_2 . That is when the Continuity Thesis applies and holds that Bella's initial actions result in a stringent protective duty at t_2 . By contrast, the duty owed to Amy not to (say) *trespass* on her family's mountain does not seem grounded partly in reasons having to do with the harm likely to befall Amy when she, as a result, unreasonably tries to cross in *Footbridge Arrogance*.

With this account in the background, I submit that Bella owes Amy a protective duty at t_2 in *Footbridge Lie* because her initial action (the lie) violated a duty of non-endangerment owed to Amy. Of course, Bella can foresee that her actions in both cases will result in Amy being in danger, so the critical question becomes what it takes to endanger someone in the relevant sense.

The reason Amy does not violate a duty of non-endangerment in *Footbridge Arrogance* is the same as why the Submission Argument from Chapter 2 does not work. That argument held that in (say) escalation cases, one owes it to the aggressor to hand over one's money because one knows that resistance will trigger an escalation and the victim having to kill the aggressor in the end, which is disproportionate to the value of keeping one's wallet. The argument fails, I claimed, because, between the victim's initial resistance and the victim killing him, the aggressor is entirely free, and required, not to escalate. That is why the aggressor cannot complain at t_1 that the victim's initial action will lead to his death.

The same is true of *Footbridge Arrogance*. Amy's unreasonable intervening agency between Bella's crossing and her ending up in danger prevents her from being able to complain that Bella's initial act endangers her. That is why Bella does not violate a duty of non-endangerment. On the contrary, the case is more accurately described as one in which Amy unreasonably endangers herself. Thus, Bella does not have a protective duty to help Amy at high personal costs and risks when Amy does predictably end up in danger either. This holds even if Bella's crossing is all things considered wrong, or indeed wrongs Amy for other reasons, such as being an instance of trespassing.³⁶

Although there is intervening agency in *Footbridge Lie* as well, the resulting endangerment is not freely chosen in the relevant sense. Bella has reduced Amy's *fair opportunities* to avoid danger by lying to her. The lie is not just wrong because it is dishonest but because it makes it unfairly more difficult for Amy to choose in a way that does not risk her being seriously harmed. That is why Bella's lie violates a duty of non-endangerment, and she now owes Amy a stringent duty to protect her from actually being harmed.

Here, I take inspiration from David Brink's (2021) account of responsibility – the Fair Opportunity Conception of Responsibility. According to this, to be fully responsible for an act

³⁶ In the latter case, the situation is more similar to *Albert's Dangerous Party* from Chapter 2.

or outcome one must have had fair opportunity to avoid performing the act or causing the outcome. More precisely, it requires the capacity and ability to recognise and respond to the relevant reasons for action and having adequate opportunities to exercise that capacity and ability without undue interference. Because Bella interferes with this in *Footbridge Lie*, she reduces Amy's fair opportunities to avoid danger, and thus makes her less than fully responsible for her own endangerment.

More precisely, Bella unduly interferes with Amy's opportunities to avoid danger by epistemically disadvantaging her. By answering Amy's question about whether it is safe to cross, she presents herself as giving reliable testimony about the bridge's safety. Absent clear defeaters, this provides grounds for Amy to trust Bella's testimony and rely on it in deliberating about what to do. In short, she provides Amy with seemingly good epistemic support for deciding to cross. Thus, Bella undermines Amy's fair opportunities to appreciate the real danger and choose in a way that would not expose her to it. In *Footbridge Arrogance*, by contrast, Bella's behaviour does not reduce Amy's fair opportunities to avoid danger. In fact, by warning her, she gives her an excellent opportunity to avoid it. That is why the two versions of Bella differ with respect to whether they also end up with stringent protective duties towards Amy.

Finally, this more general theory of protective duties also helps us understand why Billy's defensive rights in *Provocation* are seriously compromised. The reasons that apply to Ali demand that he not kill Billy. But by angering Ali so much that his self-control is undermined, Billy unduly interferes with Ali's fair opportunities to avoid attacking him. He undermines Ali's ability to recognise and adequately respond to those reasons in the heat of the moment.

Part of the reason this wrongs Ali is that it essentially endangers him. For it reduces Ali's fair opportunities to avoid being defensively harmed. Imagine, for a moment, that Billy knows his friend, Victor, will come to his defence if Ali attacks him. Then, Billy's provocation

endangers Ali by reducing his fair opportunities to avoid being harmed by Victor. In that case, Billy owes him a duty to protect him from Victor's force, even if it comes at great personal costs. The same is true in the original *Provocation* case, with the difference that Billy's provocation reduces Ali's fair opportunities to avoid being harmed by Billy himself. Nevertheless, the argument for the protective duty still applies. Billy owes it to Ali to ensure that he does not actually end up suffering this harm, and that requires not using defensive force against him.

This view also explains why Britney and Beth in *Sexual Assault* and *Park Assault* are fundamentally different from Billy. They have not unduly interfered with Adam's and Andy's ability to recognise the reasons that apply to them and respond appropriately. Thus, they have not reduced their fair opportunities to avoid attacking them and so have not violated a directed duty of non-endangerment. So, they do not have protective duties towards Adam and Andy when they predictably attack them either. In this sense, Britney and Berth are like Betty in the *Dangerous Party* case from Chapters 2 and 3.

5.3 The Moderate View

As presented so far, one might worry that the view entails a version of the Draconian View: no harmful defensive options are permissible for Billy. By provoking Ali, Billy reduces his fair opportunities to avoid attacking him. Therefore, any defensive harm imposed in response to that threat is harm that Billy has a duty to protect him from. In that case, he is also prohibited from (say) breaking Ali's leg, even if that was sufficient to prevent the lethal threat. But that seems too harsh. Indeed, I claimed earlier that the more compelling view of provokers' defensive rights was a scalar version of the Moderate View. Provocation essentially makes it more difficult to satisfy the proportionality condition, and it does so in proportion to the severity of the provocation.

However, the Protective Duty View does not entail the Draconian View. Remedial duties, in general, come with constraints. Consider *Footbridge Lie* again. Suppose that Amy was

just in danger of bruising her toe, but that Bella would have to throw herself off the bridge, to her death, to protect Amy from this. That is not required by the protective duty because the harm it protects Amy from is so small compared to the sacrifice Bella would have to make. The same suggests that Billy's protective duty cannot require him to sacrifice his life to avoid breaking Ali's leg.

That said, there is something attractive about the Draconian View in cases such as *Provocation*. Even if Billy is not required, in the end, to allow himself to be killed instead of (say) breaking Ali's leg, he still seems to *wrong* Ali if he imposes this minor harm. It is harm that Ali only has to suffer because Billy wrongfully provoked him and reduced his fair opportunities to avoid attacking him. Although he might be justified in imposing comparatively minor harm in the end, Ali still seems owed an apology for this harm. The Protective Duty View accounts for this insofar as it holds that Ali has a right to be protected from any defensive harm and Billy is only permitted not to do so in the circumstances because of the disproportionate cost that this would require of him. In other words, the view holds that Billy is permitted to impose the harm only as a lesser evil.

Of course, this does not account for the second, scalar part of the Moderate View mentioned above, namely that the severity of the provocation affects the extent to which the provoker's defensive rights are compromised. This is not just the idea that how easy or difficult it is to justify a defensive harm as a lesser evil depends on the severity of the provocation. It is the claim that the severity of the provocation also affects what defensive harms would *wrong* the aggressor. Contrary to what I suggested above, it is intuitive that how severe Billy's provocation was also affects whether Ali would be wronged by being made to suffer a given defensive harm.

The Protective Duty View can account for this, at least in some fashion, once we realise that the stringency of a protective duty depends in part on the extent to which one reduced the

other agent's fair opportunities to avoid being harmed, and that this can depend on facts about the other agent as well. Consider *Footbridge Lie* once again. Intuitively, how stringent Bella's duty to sacrifice to save Amy from the danger depends partly on what other reasons Amy had against crossing. Suppose that Amy had already encountered many warning signs saying that the footbridge was faulty. Bella still unduly interferes with Amy's decision-making when she provides her seemingly honest testimony to the contrary. Still, if Amy decides to cross and ends up in danger, the stringency of Bella's protective duty to help is also diminished by Amy's degree of fault for endangering herself. She had some evidence suggesting that she should not take the risk that she chose to disregard.

The same applies to *Provocation*. Richard Holton and Stephen Shute (2007) convincingly argue that loss of self-control does not come in degrees. It either happens, or it does not. But loss of self-control does not (necessarily) *remove* responsibility. At least apart from extreme cases, provoked people are still agents. Their agency is *compromised* because they fail to resist the anger and thus cannot reason and restrain themselves appropriately. However, not every loss of self-control is morally the same. People can differ regarding how easy it is for anger to undermine their self-control, what causes them to be angered, and how they act when they lose self-control. Such differences, moreover, are sometimes due to factors we can hold them responsible for.

Imagine, for instance, that X is not generally likely to lose control if he hears a nasty joke about himself. But he is also racist and so highly likely to be enraged and lose control if a black person makes such a joke. Alternatively, suppose that X is as likely to lose self-control and attack someone regardless of who tells such a joke but that he is much more likely to do severe harm if the joke teller is black. These are, of course, extreme cases, but the point generalises. Indeed, I suspect that most of us are not completely blameless for many of our various

dispositions to be angered and lose self-control. (Most of us, for instance, would likely easily benefit from more mindfulness exercises.)

These facts matter for the extent to which someone caused to attack out of a loss of self-control can be said to have had their *fair opportunities* to avoid attacking reduced. The more at fault someone is for the fact that a particular action will trigger a loss of self-control or for how violently they will react, the less the action reduces their fair opportunities to avoid posing a threat.

Take a case like *Provocation*, then. Imagine one version in which Ali loses self-control and tries to kill Billy just because Billy tells a mean joke about him. In another, suppose that he loses self-control and tries to kill Billy because Billy (perhaps falsely) tells him that he has brutally killed his mother. Normally, Ali is more at fault for his loss of self-control and how he reacts in the first case than in the latter. After all, apart from special cases, we can more reasonably expect that people not be so overly sensitive to jokes³⁷ or at least that they have taken steps to work on it, than we can expect people not to lose self-control when learning about their mother's death from the unrepentant killer himself.

This difference should also affect the stringency of Billy's subsequent duty to protect Ali from defensive harm once the attack occurs, because there is a difference in the extent to which Billy has reduced his fair opportunities to avoid suffering defensive harm. This matters for how large of a sacrifice that Ali can reasonably demand that Billy suffer to protect him from being defensively harmed in response to that threat. In that sense, the severity of the provocation matters for how compromised Billy's defensive rights become.

However, suppose that we stipulate that Ali, in the joke version of the case, is fully culpable for the fact that he will lose self-control after hearing the joke. For instance, he has

³⁷ Of course, there are possible exceptions. It could be that Ali is not at all at fault for being disposed to react this way to a joke.

had several opportunities to get his anger under control but has culpably chosen not to take advantage of them. Does that mean the provocative joke does *not* reduce his fair opportunities to avoid attacking and that we should treat him as we do (say) Adam in *Sexual Assault*? No. The fact that he will lose self-control because of the joke significantly limits his opportunities to avoid attacking *here and now*. He cannot snap his fingers now to cure his anger issues and avoid losing self-control if the joke is told. Although the *stringency* of the resulting protective duty is affected by the fact that Ali is fully culpable for his disposition to lose self-control now, it is a mistake to treat Ali as if his decision to attack Billy is voluntary and fully culpable (which Adam's decision to attack Britney in *Sexual Assault* is).³⁸

Ultimately, the Protective Duty View is compatible with a scalar Moderate View on how provocation affects one's defensive rights. There are (at least) two facts that affect a provoker's defensive rights. The first is a general proportionality constraint which asserts that one cannot be made to suffer significant sacrifices when it would only protect the aggressor from a (sufficiently) minor harm. The second, more substantial, factor concerns the aggressor's degree of responsibility for the fact that he will be successfully provoked into posing a threat by the victim's behaviour.³⁹

6 Conclusion

In this chapter, I have examined how provocation compromises one's right to defend oneself. In short, according to the Protective Duty View, provokers incur protective duties to ensure that the aggressors do not end up harmed as a result of having their fair opportunities to avoid attacking reduced.

³⁸ This is similar to how I said we ought to treat cases like *Dangerous Party Machine* from Chapter 2, in which the aggressor is culpable for setting a threat in motion but cannot undo the threat anymore.

³⁹ A third potential factor is how culpable the provoker is for reducing the aggressor's fair opportunities to avoid harm. Intuitively, there is a difference in the defensive rights of Billy in *Provocation* and Bernard in *Religious Provocation* because of this difference in culpability.

Defensive Entrapment and Manipulation

1 Introduction^{*}

I said in the last chapter that we should distinguish between “provokers” and “provocateurs”. The former are those who engage in behaviour which causes others to be (partially) excused for their aggression. The latter, by contrast, are best characterised by their malicious intentions. They try to get others to attack them *for the sake* of creating a situation in which they can defensively harm the other party.

Although the two groups likely often overlap, they can come apart. One can be a provocateur without causing the other agent to be partially excused for their violent behaviour. Provocateurs, even when they are not also provokers, intuitively compromise their defensive rights as well. Consider:

Fight Wish: Bree is a skilled self-defence instructor and wants to beat someone up. She dresses up as a fragile old lady and walks around in a dangerous park at night. She decides to approach the scariest person she sees, Alan, to ask for directions, hoping that he will assault her so that she has to use defensive force against him. As expected, he attacks her.¹

Infidelity: Bo knows that his brother-in-law, Adil, is ruthless and has antiquated beliefs about family and honour. He is likely to try to kill Bo if he is caught

^{*} Section 4 overlaps with my (2022) article “Entrapment and Manipulation”, *Res Publica* 28(4): 557-583.

¹ I borrow and modify this case from Ferzan (2013, p. 598) which is inspired by the movie “Death Wish”. Note, cases with a similar, intentional element (though it is not always specified whether the victim also partially excuses the aggressor) are also found in, for instance, Wasserman (1987, p. 368), Rodin (2002, p. 79; 2011, p. 92), Farrell (2013, p. 572), Alexander (2013a, p. 625; 2021, p. 109), and Brownlee (2019, p. 122).

cheating. Over time, Bo comes to hate Adil and wants to harm him. He plans to cheat, make sure that Adil learns of it, and then defensively harm him when he tries to kill him. Adil then tries to kill Bo.

I think both victims – Bree and Bo – have compromised their defensive rights in some sense. To borrow Alexander’s (2021, p. 111) words, their defensive actions would be “morally tainted” or “not fully justified”, given how they have ended up facing these threats.² This might be considered an objection to the Protective Duty View insofar as it cannot account for this intuition. Bree and Bo have not caused Alan and Adil to be (partially) excused for attacking them. Their ability for self-control has not been undermined (at least we can stipulate that it has not) and they do not seem to have had their fair opportunities to avoid attacking Bree and Bo undermined in other ways. However, I think it is correct to treat the cases differently from the ones covered by the Protective Duty View.³

This chapter provides a novel view of how and why provocateurs compromise their defensive rights. The plan is as follows. Section 2 considers what I call Alexander’s Challenge to the suggestion that Bo’s and Bree’s defensive rights are compromised and briefly consider some possible responses. Finding them lacking, Section 3 notes the similarity between these cases and cases of *legal entrapment* and proposes that we look at the former through the lens of legal entrapment. Section 4 then outlines and defends a novel theory of legal entrapment: first, I criticise existing accounts for focusing too little on the victims of entrapment; second, I outline the Manipulation Account, which argues that entrapment makes punishment morally

² There is an irony in borrowing Alexander’s words here since, as we shall see, he disagrees with this view.

³ One may wonder whether Bree and Bo have some justification for their actions. If Alan and Adil are so easily triggered, they are dangerous individuals and likely to pose similar threats to unsuspecting victims in the future. In that case, Bo might help others by killing Adil, and Bree might help others if her defensive actions deter Alan from trying to assault anyone in the future. However, I set these kinds of cases to the side, as my primary focus is the role that malicious intentions play in compromising an imperfect victim’s defensive rights. So, let us stipulate that they would not pose such threats to others at any other time. In fact, even if Bree’s and Bo’s actions would help others in this way, one might doubt that they can justify their actions on those grounds because, as Ferzan (2013, p. 620) says, they would be vigilantes and violate the rule of law.

tainted because of what I call manipulation-by-hidden-intentions; third, I consider and respond to some potential to this account of entrapment. With that theory of entrapment on the table, Section 5 applies it to cases like *Fight Wish* and *Infidelity*, and responds to two objections.

2 Alexander's Challenge

The view that provocateurs compromise their defensive rights in some form is popular.⁴ However, not everyone agrees. A notable exception is Alexander (2013a; 2021).⁵ In cases like the two considered above, Alexander claims that the victims may be culpable for their initial behaviour, but that this culpability has no implication for their right to defend against culpable aggression.⁶ On his view, they have culpably created the conditions for justified self-defence. As he says, commenting on similar cases, “neither [victim] loses his privilege of self-defence if the deadly attacks they provoke are culpable reactions to the provocations – even if they were hoping to provoke those culpable responses” (2013, p. 625). In other words, he supports a version of the Lenient View. The victim’s defensive options are the same as if they had played no culpable role in creating the threat.

Although I believe he is incorrect, it is worth highlighting what might be called Alexander’s Challenge to the idea that provocateurs compromise their defensive rights.⁷ Consider a variant of *Fight Wish* – call it *No Fight Wish* – in which Bree innocently asks Alan for directions (she does not know it is a dangerous park), and Alan tries to assault her. There is nothing compromising her right to defend herself against Alan. According to the view of the right to

⁴ For instance, even Hurd, who is so sceptical of duties to stay away from trouble, puts that scepticism to the side when the agent has “their adversary’s death as their purpose” and “intend to provoke a dispute so as to use it as cover to murder their adversary” (2016, p. 271).

⁵ See also Wasserman (1987, p. 369).

⁶ Take *Infidelity*. Alexander would say that Bo is culpable for cheating on his wife. Does the fact that he knowingly and intentionally increases the likelihood that he will have to (justifiably) kill Adil add to his culpability? Alexander (2013a) seems somewhat sceptical, but in (2021) he is more inclined to think so. In either case, however, he believes there are no implications for his right to self-defence against Adil.

⁷ Alexander does not frame a challenge in this way, but it builds on some of the points he makes in Alexander (2013a; 2021).

self-defence set out at the very beginning of the thesis, Bree's right to defensively harm Alan is grounded in the following facts:

(1) Bree has a right not to be assaulted by Alan.

(2) Alan is fully culpable for assaulting Bree.

Alexander's Challenge is that we should say just the same about Bree's defensive rights in *Fight Wish* as in *No Fight Wish* unless we can point to something about Bree's initial behaviour which could have made (1) or (2) false in *Fight Wish*. But, the challenge goes, it is implausible to think that (1) and (2) are made false by Bree's initial behaviour.

I agree with the second point here. First, I doubt that Bree's behaviour has made her forfeit or otherwise compromise her right not to be assaulted by Alan just because she is acting with malicious intentions. Attacking her has no value, and it is implausible that people, even wicked people, forfeit or weaken their rights not to be harmed for no good reason. Second, Bree's approaching Alan in the park, intending to get him to attack, does not obviously make his attack less culpable. In his mind, after all, he is attacking an innocent park-goer, and he has no good reason for doing so.

However, focusing on (1) and (2) is not the only way to respond to Alexander's Challenge. Another possibility, already touched upon in the previous chapter, is to agree that the right not to be assaulted in (1) is unaffected but that Bree has forfeited the right to *enforce* her right not to be harmed by Alan. As said, this is the view suggested by Ferzan (2013; 2020). I criticised it in relation to provocation, but it might have more hold concerning provocateurs. It is one thing, some might say, for V to engage in behaviour that he "merely" *knows* is likely to cause X to attack him and then claim the right to self-defence. It is vastly more objectionable for V to engage in behaviour *aimed* at getting X to attack him and then claim the right to self-

defence. Here, it may be particularly plausible that V “lose a standing to complain about being wronged (and lose the ability to prevent being wronged)” (Ferzan 2020, p. 126).⁸

The view is somewhat attractive and perhaps part of an overall view of provocateurs’ defensive rights. Nevertheless, it fails to capture an intuitive feature of the case. On Ferzan’s view, since Alan is about to violate Bree’s right not to be harmed, he is *liable* to defensive harm. Thus, he is not *wronged* if she defends herself (at least if she does not use more force than a perfect victim analogue would be permitted to use). The problem for Bree is only that she lacks the “standing” to justify her otherwise permissible defensive action against a liable aggressor. Indeed, on Ferzan’s view, it is largely immaterial whether Bree wrongs Alan at any stage. For instance, in her treatment of a case similar to *Fight Wish*, Ferzan argues that Bree’s conduct is primarily culpable because she violates the rule of law (not Alan) (2013, p. 620).

However, I believe that Bree’s behaviour *wrongs* Alan in some sense as well. The moral taintedness of her right to defend against Alan is a result of, and tied to, an initial wrong she perpetrates against him by *attempting to lure him into attacking her so that she may harm him*. In other words, echoing my previous arguments, we should view the case through the lens of the Continuity Thesis. The relevant question is therefore what the nature of the initial wrong is and what kind of duty not to defensively harm it generates. It cannot simply be the violation of a duty to avoid trouble, as that does not account for the *wronging* element. Nor can it be the violation of a duty of non-endangerment, as Bree does not reduce Alan’s fair opportunities to avoid attacking her. Instead, I will argue that Bree violates a specific kind of duty not to manipulate Alan.

⁸ As noted earlier, similar views that one can forfeit the right to complain about being wronged (and possibly to prevent being wronged) are found in Cornell (2016, p. 1149), Kamm (2022, pp. 349-51), and Clark (2023, pp. 46-50).

3 The Entrapment Solution

The motivation for my solution to Alexander’s Challenge begins with the observation that, as suggested, what makes the provocateur cases special is that their actions are motivated by the *intention* to gratuitously create a threat situation so that they may use defensive harm. This seems sufficient to establish that they are acting wrongfully and that their defensive rights are compromised. Once we learn about this intention, it becomes vastly less relevant to determine whether the victim’s actions are “otherwise legitimate”. It is not important whether Bree is walking in a public or a private park, or whether Bo has his wife’s consent to sleep with someone else. Their malicious intentions seem sufficient to morally taint those actions in a way that compromises their rights to self-defence as well.

To understand how such intentions taint their actions and subsequent defensive rights, I suggest we understand these cases as analogous to legal entrapment (hereafter, “entrapment”). On the face of it, the two phenomena seem quite similar.⁹ One popular view of entrapment holds that it occurs when police, or agents of the state, solicit, persuade, or otherwise encourage a person to commit a crime (that they would otherwise not commit) for the purpose of arresting and punishing them. More precisely, it is said to have the following elements:

- (1) The police incite a target to commit a crime.
- (2) The target would not have committed the crime absent the incitement.
- (3) The police incite the target with the intention of having them arrested and punished.

As an example, consider the following:

⁹ Alexander (2013a, p. 627) and Ferzan (2013, p. 620) also briefly mention the similarity between cases like *Fight Wish* and entrapment but do not discuss it in detail. Katz (2013) makes more of the similarity but goes in the other direction: using cases like *Fight Wish* to understand entrapment.

Drugs: A local police department has decided they need more drug-related arrests. They send an undercover officer, Paul, to pose as a drug dealer at a university party. He notices a student, Dan, and asks him to deliver some drugs down the street. Dan initially declines, but Paul is persistent, insisting that he really needs the sale. Finally, Dan agrees, and he is then arrested by Paul's fellow officers when he delivers the drugs at the other house.

Cases like *Fight Wish* and *Infidelity* seem to me to be instances of what we may call *defensive entrapment* as they satisfy elements analogous to (1) – (3) mentioned above. Take *Fight Wish*, for instance. Bree tries to incite Alan into attacking her; Alan would not have attacked her absent this (at least we can stipulate it), and Bree acts intending to use defensive force against Alan.¹⁰

Further support for the similarity is found in the fact that the two cases give rise to similar intuitive judgements. What we might call “the puzzle of entrapment” consists of two seemingly incompatible intuitions about cases like *Drugs*. On the one hand, something about how the state has made Dan commit the crime makes it problematic to punish him. At least, punishing him as harshly as would be appropriate if someone else has encouraged him seems wrong. On the other hand, Dan seems just as deserving of, or liable to, punishment as he would be if he had been persuaded into breaking the law by a genuine drug dealer. In Dan's mind, after all, he *is* persuaded by a genuine drug dealer, and that is ordinarily not enough to justify or excuse breaking the law.

The same sort of incompatible intuitions arise, I believe, in cases like *Fight Wish*. On the one hand, there is something inappropriate about Bree's using defensive force against Alan,

¹⁰ If condition (2) is not satisfied, it might be permissible to create a situation at t_1 with the *intention* of imposing defensive or punitive force at t_2 . For instance, if Bree knows that this is necessary to prevent Alan from attacking her in the future, at a point in time she cannot defend herself, or if the police know that Dan is a drug dealer but that this is the only way to stop him. I return to this issue in Section 4.4.2. For now, we are interested in cases in which the intention is clearly malicious, by which I mean that there are no such preventive grounds for forming and acting on it.

given how she has caused him to attack her. On the other hand, as Alexander's Challenge suggests, he seems as culpable for his attack as he would be if someone else had caused him to attack Bree in the same way.¹¹

I propose to take the similarity between legal entrapment and defensive entrapment seriously. In the next main section, I take a detour into the ethics of punishment and outline an account of why entrapment taints punishment in order to come back to cast new light on defensive entrapment.

4 A Theory of Entrapment

4.1 Existing Views

It would be nice if we could apply an already-existing theory of entrapment to cases of defensive entrapment. Unfortunately, existing accounts do not capture some of entrapment's distinctive and problematic aspects. My worries here are, in fact, similar to the ones mentioned concerning Ferzan's view of provocateurs above. But I will go more in-depth here, as it will help us understand what is needed for an improved account.

The puzzle of entrapment, as said, is generated by two seemingly incompatible intuitions:

- (1) It is wrong to punish (or fully punish) entrapped offenders.
- (2) Entrapped offenders are liable to (full) punishment.

¹¹ There might be a difference in how forceful these intuitions are. Dan needed more of a push, and his wrongdoing is minor. However, we can make *Fight Wish* more similar by imagining a version in which Alan needs more of a push (without his culpability being undermined) and the threat is less serious. We can also make *Drugs* more similar to *Fight Wish* by imagining versions in which Dan needs less of a push and in which his crime is more serious. There still seems to be *something* problematic about the state's punishing him. I return to some of these complications in Section 4.2.

Admittedly, some reject this puzzle. First, Howard (2016, p. 34) is somewhat sceptical of (1), for instance. However, he does not explicitly argue for its rejection, and it seems to me implausible to reject it in paradigmatic examples of entrapment, such as *Drugs*. Still, we should concede that entrapment does not necessarily make punishment *all things considered* impermissible. It only makes it *pro tanto* wrong. There can be instances in which it is ultimately right to punish the entrapped offender after all. For instance, it may be permissible to punish targets of entrapment who have been radicalised by undercover agents and now pose a danger to society if not punished for their crime. Similarly, if the crime in question was sufficiently grave or caused innocent people harm, the state may owe it to the crime victims to hold the offender accountable. Nevertheless, we should hold that the targets have suffered a wronging by being entrapped and will be wronged by being punished. It is just that the state has a *lesser evil* justification that makes it all things considered permissible to inflict this wrong. I omit the “pro tanto” qualification (unless necessary) for ease of exposition.

Second, Paul Hughes (2004, p. 45) and Hohan Kim (2019, p. 72) are sceptical of (2).¹² They claim entrapment always involves pressuring tactics severe enough to diminish the target’s responsibility. Thus, entrapped agents are not fully culpable and thus always deserve lesser punishment than their non-entrapped counterparts. Although I agree that entrapment *often* involves such severe pressuring (or (partially) exculpatory deception, for that matter), it is not *always* the case. For instance, Dan’s responsibility does not seem diminished in *Drugs*. Imagine again a version of the case in which Paul is a genuine drug dealer. The persuasion used in that case does not reduce Paul’s responsibility to a level that affords him a (partial or full) excuse for breaking the law. Thus, reduced responsibility is not essential to entrapment cases.¹³

¹² Note, Kim has one of the more comprehensive views in the literature. I return to other parts of his view below.

¹³ However, when this is true, arguments about the protective duty from the previous chapter are also relevant.

However, most philosophers working on entrapment take the puzzle at face value and seek to solve it. It is beyond the scope of the chapter to highlight and criticise all existing accounts, but I will outline three particularly instructive ones.¹⁴ Consider first:

The Standing Account: the state should not punish entrapped criminals because it has lost its standing to blame the criminal.

The suggestion borrows from the idea that one can lose one's standing to *blame* another person for their wrongdoing because (say) one is guilty of doing something similar or complicit in the other person's wrongdoing. Similarly, Hock Lai Ho (2011, p. 88-91) suggests that the state dirties its hands when instigating a crime and thus loses its standing to blame and condemn anyone else for that crime. Kim (2019, pp. 83-85) also argues that hypocrisy and complicity concerns undermine the state's standing to blame the entrapped criminal when it has intentionally created the crime and bears some responsibility for it. However, Kim also recognises a potential weakness of that argument. Hypocrisy and complicity bear most obviously on the standing to *blame*. Yet, for many, punishment necessarily involves more than blame and is (at least partly) justified by more than the value of blaming wrongdoers. On many popular views, for instance, punishment involves harsh treatment and is partly justified by how this contributes to deterrence. Our view should also explain why imposing harm for that purpose is morally problematic in entrapment cases. Thus, Kim also argues for a second solution to the puzzle:

The Legitimacy Account: the state should not punish entrapped criminals because it has violated one of the preconditions for the right to punish.

More precisely, the state lacks the legitimate authority to punish entrapped agents because (i) a precondition for the *right* to punish is a commitment to crime prevention but (ii) states that entrap people *create* crimes and so violate that precondition (Kim 2019, p. 84). In a similar

¹⁴ See Howard (2016, pp. 26-28) for a comprehensive overview of other accounts and critiques of them.

spirit, Andrew Carlon argues that a state which entraps has lost its “right of prosecution” because it “has ceased to embody the principles of justice” (2007, p. 1123). Lastly, consider:

The Incoherence Account: the state should not punish entrapped criminals because it is incoherent and undermines the integrity of the state.

According to Antony Duff, Lindsay Farmer, Sandra Marshall, and Victor Tadros (2007, p. 243), for instance, the problem with entrapment seems to be the following. In entrapping someone, the state is expressing – through its encouraging behaviour – that the criminal act is not worthy of condemnation. However, in seeking to punish the offender at the trial, the state expresses that the act is worthy of condemnation. By entrapping and punishing, the state would act incoherently, which, in turn, undermines the integrity of the trial, the criminal justice system and the state itself.¹⁵

While these accounts all have a certain appeal, they do not provide a complete explanation of what is wrong with entrapment. They are all instances of what I call Self-Compromise Theories. Each explanation offered for why it is wrong for the state to punish Dan in *Drugs* focuses primarily on the state and its having morally compromised or tainted *itself*. The state’s failure to live up to certain principles or expectations does all the explanatory work. That it has done so by *wrongfully entrapping Dan* plays little, if any, role. For instance, as said, the worry, according to the Incoherence Account, is that entrapping and punishing a person “would damage the integrity of the criminal justice system (Ashworth 1999, p. 307). Similarly, Kim says that the rationale for the Legitimacy Account “is all about [the state] and not about [the

¹⁵ Dworkin (1985, p. 32-33) also focuses on the incoherence of entrapment as the key to why it is wrong, and Ashworth (1999, pp. 307-10) also worries that the state’s incoherent behaviour threatens the justice system’s integrity. Note that the view here is similar to the Standing Account but subtly different. The issue is not that the state has done something morally wrong or is indeed complicit in the very wrong that the offender has committed and therefore lacks the moral standing, or moral high ground, required to blame the offender. Instead, the issue for the state is that it has already expressed that something is not worthy of condemnation and that it would be unprincipled to “change its mind” at the trial.

entrapped person]” (2019, p. 83). And Ho claims that a stay of proceedings in entrapment cases should *not* be “granted to protect the entrapped or uphold any of his or her rights” (2011, p. 95). It is just the result of the state lacking the clean hands required to condemn and blame them.

However, I think that part of the intuitive explanation for why it is wrong to punish Dan is the fact that the state *wrongfully entrapped him* into committing the crime in the first place. It is not just a matter of the state having compromised the standing, legitimacy, or integrity required to justifiably punish someone. Rather, an important part of the explanation is that it wrongs Dan when it entraps him and *therefore* should not to punish him for the crime.¹⁶

In support of this, we can observe the fact that it seems reasonable for Dan to resent the state if it punishes him, but Self-Compromise Theories cannot so easily account for this because the wrongness of punishing him, on those views, has little to do with what is owed to him. At best, it seems his resentment would be the same as the one we could direct at anyone seeking to punish us without the requisite integrity, standing, or legitimacy. However, the resentment seems stronger and more particular in Dan’s case.

To see this, notice that the concerns highlighted by the Self-Compromise Theories can, at least in principle, also be found in very different cases. Suppose, for instance, that the state has entrapped Dan but also wants to punish someone else, Dean, whom they have not entrapped, for a similar crime. The state may lack the right to punish him for self-compromising reasons. If the state is guilty of causing similar crimes in the past (as when it entrapped Dan), it might lack the clean hands, legitimacy, or integrity necessary to permissibly punish Dean for the crime. But it is much less clear that we should not punish Dean. Even if one thinks so, I doubt Dean should be very resentful towards the state, at least not compared to how resentful Dan

¹⁶ Howard (2016, p. 29) agrees that entrapment *wrongs* the targets (because it disrespect them), but as said, he does not think this entails that it is wrong to punish them. I return to his view in more detail below.

may legitimately be. The obvious difference is that the state has only wrongfully entrapped Dan.

This is not simply suggested as a problem for the extensional adequacy of the views. Certain versions of Self-Compromise Theories might be extensionally adequate and able to distinguish between Dan and Dean.¹⁷ However, they would seem to still be unable to account for the sense, suggested above, that the reason not to punish Dan appears intimately connected with the earlier wrongful act of entrapping him. The wrongness of punishing Dan at t_2 does not follow “incidentally” from the state’s morally compromising itself. It follows normatively from the nature of the wrong of entrapment itself. In other words, the duty not to punish an entrapped wrongdoer has a remedial flavour. The state should not have entrapped Dan in the first place, and its duty not to punish him now is grounded partly in recognition of that.

To make sense of this, we need to reorient our starting point. To understand why it is wrong to punish Dan, we should not just focus on how the act of punishing at t_2 is wrong because there is a lack of standing, legitimacy, and so on. Instead, we need to go back to the wrong of entrapment itself at t_1 and provide an account of its wrongfulness that also explains why the decision not to punish Dan at t_2 is a morally appropriate decision.

Simply pointing to the fact *that* entrapment itself is wrong will not do. For instance, Self-Compromise Theories might agree that the entrapment act itself is wrong for various reasons. Plausibly, it is wrong to create crimes, to express to others that it is fine to commit a crime, and so on. But there is nothing inherent in the nature of these kinds of wrongs that entails that it is inappropriate to punish Dan at t_2 . Nor will it do to simply claim that entrapment itself

¹⁷ For instance, Tadros’ (2009, p. 393) version of the Standing Account concerning the standing to punish socially deprived offenders focuses mostly on the state’s *complicity* in their crimes, and he argues that complicity undermines one’s standing to blame wrongdoers more seriously than merely being hypocritical does. Still, the view suggests that the difference between Dan and Dean vis-à-vis the right to punish is a matter of degree. However, I believe the contrast between Dan and Dean is more categorical. The fact that the former but not the latter was entrapped provides a way of drawing a sharper line between them.

wrongs the target. Howard's (2016) Subversion Account, for instance, holds that entrapment wrongs the targets because it is disrespectful to people's moral agency to knowingly increase the likelihood that they will act culpably. As he notes, though, this alone does not show that subsequently punishing the target would be wrong.¹⁸ There is nothing inherent in disrespecting someone in this manner which entails that it is inappropriate to hold them accountable for their wrongdoing. Indeed, he (2016, p. 44) suggests that one needs to supplement the account with another one, like the Standing Account, to get that result.

Thus, to repeat, what is needed is an account of entrapment which clarifies the nature of the wrong in a way that also makes the remedial duty not to punish an intelligible and appropriate upshot of it. To do so, the act must be understood as wrong partly for reasons having to do with the prospective punishment.

4.2 The Manipulation Account: Entrapment

What I call the Manipulation Account is inspired by the earlier definition of entrapment. According to the account, entrapment gets its distinctive wrongness from condition (3): the police incite the target with the intention of having them arrested and punished. A natural first question is what it means to intend for something. Here, I understand intentions roughly according to Gideon Yaffe's (2010) account, which, in turn, is inspired by Michael Bratman (1987). Intentions are practical mental states that "play a role in deliberation and in the motivation and guidance of action" (Yaffe 2010, p. 53). More precisely, to intend that something *p* occur comes with rational and practical commitments – for instance, to pursue courses of action one believes are necessary for *p*, not to deliberate about courses of action one believes are incompatible with *p*, and so on. And when an intention that *p* plays its proper causal role, it is what motivates one to pursue the action in question. In the context of entrapment, we can say that condition (3) is

¹⁸ Of course, Howard considers this a virtue of his account because, as said, he is somewhat sceptical that it is wrong to punish entrapped offenders.

satisfied when the police are committed to the outcome of the target being arrested and punished in the sense that this commitment explains their other commitments and actions – for instance, trying to make the target commit the crime, making arrangements to subsequently arrest them, and so on.

It is surprising that, to my knowledge, no one has suggested to solve the puzzle of entrapment by focusing on (3). It plays such a prominent role in most of the definitions of entrapment.¹⁹ Thus, I think we have good reason to suspect that it will be key to explaining what makes it wrong to entrap and punish as well.

Here is an over-simplified statement of the core idea behind the Manipulation Account:

- (i) Entrapment involves a plan aimed at punishing the target.
- (ii) This is an evil plan since the target is (presently) innocent.
- (iii) It is morally bad if evil plans are successful.
- (iv) Therefore, it is morally bad if entrapped criminals are punished.

Although somewhat attractive, premise (iii) is mysterious and not quite right. For instance, it is not clear why and in what sense it is “morally bad” that an evil plan is successful.²⁰ Moreover, it does not account for the sense that punishment *wrongs* the target. It is not just a bad state of affairs, or bad for the universe, if it happens – the targets themselves have a moral complaint

¹⁹ For instance, Ho claims that it is essential to entrapment that “what motivates the operation from the start is the desire to have the person convicted and punished (2011, p. 74); Duff, Farmer Marshall, and Tadros claim that entrapment involves inciting someone “for the purpose of arresting and prosecuting him” (2007, p. 242); Hill, McLeod and Tanyi claim that it is necessary that the entrapping agent “intends to be enabled, or intends that a third party should be enabled, to prosecute or to expose the target for having committed the act” (2018, p. 550); and Dworkin says that “the central moral concern” regarding entrapment is that it manufactures crime “in order that offenders be prosecuted and punished” (1985, p. 24). Not everyone agrees, however. Kim (2019, p. 80) and Howard (2016, pp. 36-37) think that negligent entrapment is possible. Still, the intuition that subsequent punishment is inappropriate is clearest in cases where (3) is satisfied. See also Tadros (2005, p. 318-9) for an argument that focuses on the intentional aspect as key to whether someone has been entrapped.

²⁰ I borrow the idea behind premise (iii) from Parr (2016, p. 994) and Duus-Otterström (2017, p. 1073).

against it. The obvious reply is to change the view to say that the success of evil plans involving particular people *wrongs* those people. However, this raises even more questions. In particular, how does this success wrong people; what does it mean for an evil plan to be about a particular person; and so on.

The Manipulation Account is a more sophisticated version of this simple idea. Put briefly, the intentional aspect of entrapment makes it a form of manipulation and provides a manipulation-based reason not to punish entrapped offenders.²¹ In this subsection, I will flesh out the manipulative aspect of entrapment in more detail. In the next, I explain how this also morally taints the subsequent act of punishment.

On my view, entrapment is best understood as involving a particular kind of deceptive manipulation that we may call manipulation-by-hidden-intentions or intention-manipulation for short. Consider an example from outside the legal context:

Job: Belinda has received a job offer at a great department, D1. Her colleague, Amos, encourages her to accept it, citing the genuinely good reasons for working there. Amos suspects that another letter from Belinda's dream department, D2, is coming. However, he wants Belinda to commit to D1 before it arrives because he wants to see her fare badly and secretly hopes he will receive an offer from D2 instead. Convinced by Amos, Belinda accepts the first offer.

Intuitively, Amos has manipulated Belinda into accepting the first job, but he has not deceived her about the good reasons for working there. Perhaps he has deceived her because he has withheld information about another job offering arriving soon.²² But suppose they both know there is a chance that an offer from D2 will arrive, and that Belinda is nevertheless persuaded to accept the first offer by Amos' encouragement. It still seems manipulative.

²¹ Hughes (2007, p. 58) and Kim (2019, p. 75) also claim that manipulation is involved in entrapment, but the kind of manipulation they refer to undermine the target's responsibility. As explained above, this kind of manipulation cannot be essential to entrapment.

²² Chisholm and Feehan's (1977, pp. 144-45) broad definition of deception could account for this.

The reason, I submit, is that Amos engages in intention-manipulation. He tries to steer Belinda towards a particular decision while, and partly by, hiding his true intentions for doing so. In discussing similar cases, Moti Gorin (2014) agrees and argues that one can manipulate others simply by hiding one's true intentions. The reason, according to Gorin, is that there is a Transparency Norm governing communication which "requires that an interactive partner not hide her intentions in interacting when these intentions are relevant to the intentions and interests of the person with whom she's interacting" (2014, p. 78). In other words, (at least, absent clear defeaters) people expect transparency concerning others' intentions when interacting with them.

In support of the view that violating this norm is an instance of *manipulation*, Gorin refers to Sarah Buss' claim that one of the key problematic features of deceptive manipulation is that it "prevents [the target] from governing herself with an *accurate understanding* of her situation" (2005, p. 226). Similarly, according to Gorin, by hiding their true intentions and "playing on the expectations of manipulees, expectations generated by adherence to the Transparency Norm, manipulators prevent manipulees from governing themselves with an accurate understanding of their situation" (2014, p. 78). In a case like *Job*, for instance, Amos is trying to steer Belinda towards a decision while keeping her in the dark as to why he is guiding her in this direction.

One initial worry with this view is that hiding one's true intentions can sometimes be perfectly fine and, indeed, the right thing to do.²³ If A wants to propose to his partner, B, at the place they first met, he might try to convince B to meet him there later while keeping his intention to propose hidden. This goes against Gorin's Transparency Norm, which holds that one should be transparent whenever one's true intentions are "relevant to the intentions or interests of the person with whom [we are] interacting" (2014, p. 78). For A's intentions are clearly

²³ Buss (2005, p. 225) makes a similar point about deceptive manipulation more generally.

relevant to B's intentions and interests. But surely A does not manipulate B here.²⁴ In fact, social norms seem to demand that A *not* be transparent about his intentions here.

Thus, we need a more refined account of exactly when guiding others' decisions while hiding one's true intentions is manipulative, but doing so is beyond the scope of this chapter. However, I suspect that such an account will at least entail that keeping one's true intentions hidden is manipulative when they are relevant to the other agent's intentions or interests in a *negative* sense. For instance, if the intentions are such that they aim at negatively affecting the other party's well-being.²⁵ This helps explain why Amos manipulates Belinda (he aims for her to miss out on a job she wants) and why A does not manipulate B by keeping his planned proposal secret.

Next, there is also a more fundamental question of precisely *why* hiding one's intentions in this way should be manipulative. Gorin would highlight that it violates the expectations created by his Transparency Norm. However, we can still ask the deeper question of why breaking it is wrong. More fundamentally: why should there be such a norm in the first place? In fact, we may wonder whether the norm and the expectations it creates capture the most important feature of Amos' wrongdoing. Suppose Belinda has always been a bit wary of Amos and always had a suspicion that he is not fully transparent about his reasons for acting as he does. It still seems manipulative to guide her decision in the way he does.

Instead, I believe Raz points us to what is more fundamentally problematic about intention-manipulation. He notes that manipulation and coercion generally tend to affect our decision-making. Often, they get us to act in ways we would otherwise not have acted by perverting our decision-making procedure or by limiting our options, respectively. That is usually a

²⁴ I am using "manipulation" here in a moral sense as implying wrongdoing. If one prefers a non-moral version of the term, one might say the case is not an instance of *wrongful* manipulation.

²⁵ The full account must be broader than this, I believe, because it seems possible to paternalistically intention-manipulate someone.

source of complaint against the manipulator and coercer. However, Raz also contends that the wrongness of manipulation and coercion “transcends the severity of the actual consequences” that they have on the victim’s decision-making (1988, p. 379). They are also wrong in virtue of the *kinds* of treatment that they are. They “subject the will of one person to that of another” (1988, p. 378), which is wrong even if the actual effects on one’s decision-making are negligible.

This is true of *Job* and intention-manipulation cases more generally. When Amos directs Belinda’s decision-making by hiding his intentions, he subjects her will to his in a disrespectful manner. He makes her a “mere” pawn in *his* plan.²⁶ More precisely, because he hides his true intentions in guiding her behaviour, he prevents her from governing herself with an accurate understanding of her situation, namely that she is being encouraged to make a certain choice because Amos wants a certain (secret) outcome to occur.²⁷ In that way, Amos treats Belinda, or her agency or decision-making capacity, as a mere means to achieving that outcome. This wrongs her even if she would have made the same decision in the end had he not manipulated her (say, because she received other information supporting her decision to go). This is a key reason why intention-manipulation is wrongful.

Importantly, however, this kind of manipulation does not necessarily affect the culpability of the target’s actions in cases where they are wrongful. Suppose, for instance, that Belinda knows that accepting the offer from D1 means having to leave her ill mother behind, who is dependent on Belinda staying, and that this makes it wrong to leave. This might make Amos’ manipulating her into accepting the offer even more wrongful.²⁸ However, the fact that

²⁶ See also Buss’ claim that manipulation is problematic partly because the manipulator treats the target as “a character in his plot, rather than as someone with whom he shares the world” (2005, p. 229).

²⁷ See also Garnett (2018, pp. 558-63) who argues, in the context of coercion, that one subjects someone to a “foreign will” when the target’s aim in doing as the coercer wants is different from the coercer’s aim in getting them to comply.

²⁸ Especially if we also accept Howard’s (2016) Subversion Account.

he is hiding his true intentions does not affect how culpable she is for wrongfully choosing to leave her mother behind. The reason is that whether Amos hides his intention does not matter for how Belinda ought to respond to the reasons in front of her. Suppose that Amos encourages her to go by pointing out the great interdisciplinary work she could partake in at D1. Whether this is pointed out to Belinda as part of a manipulative effort, as in the original case, or as part of a good faith effort to convince her which is motivated by a desire to see her thrive academically, it does not change how she ought to weigh that reason compared to the harm that leaving her mother will cause. So, it does not change how culpable she is for leaving. The difference only matters for whether Belinda is also a victim of wrongdoing herself.

Now, in my view, cases like *Drugs* are wrongful because they also involve this kind of intention-manipulation. There may be all sorts of deception and manipulation involved in convincing Dan. Paul deceives Dan about his identity, needing Dan to deliver drugs for him, and so on. But what seems essential to why it *wrongs* Dan is that it involves intention-manipulation: Paul uses Dan as a mere means or a pawn in a secret plot against himself. After all, Paul's intention is to have Dan commit the crime and be arrested and punished. He also, naturally, keeps this intention hidden from Dan – if he did not, and Dan knew what Paul was aiming at, then the plan would presumably not be successful.

More generally, given that the intentional aspect is key to the definition of entrapment, entrapment always involves this kind of intention-manipulation. According to the Manipulation Account, that is why entrapment *wrongs* those subjected to it.

4.3 The Manipulation Account: Punishment

Still, why does intention-manipulation make the subsequent act of punishing the entrapped offender wrong? After all, the mere fact that the punishment would follow a wrongful act from the police is insufficient to make it wrong. Suppose, for instance, that an undercover officer

runs a red light and a civilian, inspired by this, runs the red light as well. That does not make it wrong to punish the civilian for his wrongdoing.²⁹

The difference is that the undercover officer's wrongdoing does not *aim* at punishment. Entrapment, by contrast, does. It is wrongful in part because of the *intention* that the target be punished, and this *morally taints* the realisation of that outcome. The act of punishing the offender is not simply the act of holding a culpable offender accountable. It is also the act of *fulfilling* or *completing* the wrongful manipulation that began earlier. Moreover, we plausibly have (*pro tanto*) duties not to complete wrongs in this way. That is why the very nature of entrapment, as an instance of intention-manipulation, provides a duty not to punish entrapped offenders.³⁰

However, a possible objection goes as follows. Although entrapping Dan at t_1 is an instance of manipulation partly because Paul hides the intention to have him punished at t_2 , the manipulation ends once Paul has successfully manipulated Dan into committing the crime at t_1 . Whether he is ultimately punished is immaterial to whether he has been manipulated. Thus, any duty not to punish Dan cannot be explained as manipulation-based or as a continuation of the duty not to manipulate him in the first place. In that case, we are back to the question above: why does wrongfully entrapping someone at t_1 make it wrong to punish them, while the undercover officer's wrongfully running a red light does not similarly make punishment wrong?

In response, although it is true that whether Dan has been manipulated does not depend on whether he is actually punished, the reasons not to punish him are still intimately connected to the wrongful act of manipulating him at t_1 . Some wrongful acts can be aggravated by the

²⁹ But if the undercover officer ran a red light to try to get someone else to do so as well so that he could arrest them, I would say that it is a form of entrapment. It only presents an opportunity for wrongdoing, of course. However, as I argue in Section 4.4.2., this can also be entrapment.

³⁰ See also Carlon who at one point says that when we spare entrapped offenders from punishment, we aim at "the prevention of a wrong's fulfilment" (2007, p. 1116).

realisation of certain outcomes (or mitigated by the frustration of certain outcomes). Consider, first, a case modified from Frowe and Parry (2019, p. 125):

Revenge: After their break-up, Andreas shares nude photos of Blair without her consent on a “revenge porn” website.³¹ Carl and others later view the photos on the website.

As Frowe and Parry argue, the wrongful act of *sharing Blair’s nude photos* is normatively dependent on the realisation of the outcome in which Carl and the others (or some other people) view the photos. This outcome is not just one of the many negative causal consequences that Andreas’ uploading them to the website may have. Rather, the realisation of this outcome partially constitutes the wrong of *sharing* the photos. Moreover, the extent to which it is realised (that is, how many view them) seems to affect the gravity of the wrong he does to Blair.

Of course, as I said above, the intention-manipulation involved in entrapment does not constitutively depend on the realisation of the outcome in which the target is punished. However, I believe that the intimate relationship between the act and the outcome here makes it such that the realisation of the outcome can aggravate the wrongfulness of the manipulation. As already suggested, the reason is that the manipulation is wrong partly in virtue of the intention that this outcome occur.

This seems to me true more generally for actions that are wrong partially in virtue of intending for certain outcomes to be realised. Take, for instance, the notion of “manipulative harming”. The classic example is X pushing V over a bridge such that his body will stop a runaway trolley from hitting someone else further down the track. Although controversial, one view of why this kind of harm is especially difficult to justify is precisely that it involves *using* or *treating* V as a mere means. Using someone as a means, moreover, involves intending for

³¹ Revenge porn is “[s]exually explicit images or videos of an individual, published online without their consent and with the intent to cause them distress” (Chandler and Munday 2016).

them or their bodies to play a role in furthering some goal or end.³² Imagine, then, the following instance of manipulative harming:

Enchanted Treasure: An enchanted treasure chest requires a large blood sacrifice to be opened. Abby wants the treasure inside, so she kidnaps Bob and draws his blood. Once sacrificed, she can take the treasure inside.

Although the harm element of Abby's wrong is over and done once the blood is sacrificed, the extent to which she wrongs Bob seems to me to partly depends also on the extent to which the manipulative aspect – the intended outcome – is realised. Suppose, for instance, that Bob comes to his senses again after losing all the blood and realises he can intervene before Abby opens the chest and takes the treasure. Perhaps he can kick it off the cliff they are standing on. I think he is permitted to intervene, and indeed that we should intervene if we can.³³ Although it does not prevent the harm, it does prevent Bob from being *successfully* used as a mere means. It is perhaps not much of a consolation to Bob if he is severely harmed, but I think there is some value in Bob or us, for Bob's sake, preventing Abby from achieving the outcome she tried to use him for.³⁴

In further support of this, consider Cam. She cannot do anything to stop Abby from kidnapping and harming Bob, but she has been hired to show up after Abby has drawn Bob's blood to carry the blood over to the sacrifice area. If Bob knew all of this, and he could impose

³² See, for instance, Tadros (2011, p. 243-47; 205) and Kerstein (2013, p. 58). Admittedly, there are also ways of trying to spell out what is distinctive about this mode of harming which do not focus on intentions. See Ramakrishnan (2016) for a great overview.

³³ Not necessarily all things considered, of course. If one could instead take the treasure and compensate Bob for the harm or donate it to charity, it might be the better option.

³⁴ An alternative explanation is that we should intervene to prevent Abby from *benefitting* from her wrongdoing. However, imagine that her further aim is to give it to someone else to benefit them. In that case, one may argue that we should intervene to prevent those people from benefitting from Abby's wrongdoing. But this does not necessarily capture the sense that intervention should be for *Bob's sake*. It is not just done to prevent these *others* from *benefitting*. In fact, if the treasure was cursed, these people might not benefit from the injustice overall. Still, we have reason owed to Bob to intervene to prevent Abby from achieving her aim of securing the treasure and giving it to them.

some non-trivial and proportionate harm on Cam to escape Abby's harming him, then she would seem liable to it. One explanation for this is precisely that, even though she is not responsible for the harming, she is wrongfully complicit in the *manipulative* harming because she will help realise the intended outcome of Abby's harming Bob – that is, the very outcome in virtue of which Abby's harming him is manipulative.

Thus, it seems that Abby's manipulatively harming Bob at t_1 morally taints the intended outcome at t_2 and provides reason to prevent it from occurring or, at least, not to help realise it. In this same way, the intention-manipulation involved in entrapment also morally taints the subsequent act of punishment. In a sense, the state at t_2 has the same reason not to become "complicit" in the intention-manipulation that it (or one of its agents) is guilty of at t_1 that Cam has not to be complicit in Abby's wronging of Bob by realising the intended outcome. Realising either outcome would aggravate the manipulative wrong suffered by the victim. Since one cannot undo the initial wrong, the least one can do is refrain from aggravating it.

This helps us solve the puzzle of entrapment that we began with. It is true that Dan has no justification or excuse for breaking the law at t_1 and is thus a culpable offender. *Qua* culpable offender, then, he has no complaint against being punished. However, *qua* the victim of wrongful manipulation, which the punishment would aggravate, he has a complaint against being punished. The Manipulation Account, therefore, helps account for the sense that entrapment cases push seemingly incompatible intuitions.

In this way, we also see why entrapment only makes punishment *pro tanto* wrong. As explained, there are also reasons to punish Dan *qua* culpable offender. What is right all things considered will depend on the details of a given entrapment case. In most cases, including *Drugs*, it is likely that the reasons not to aggravate the wrong weigh heaviest. There was, for instance, no risk that Dan's action would actually harm anyone – at least if we stipulate that he

was delivering fake drugs. Most entrapment scenarios also seem to occur within controlled environments where the police can neutralise or minimise risks to innocent people.

Things are different, however, if Dan was entrapped into committing a crime that either physically harmed an innocent person or was significantly disrespectful to other citizens (for instance, a “simulated” hate crime). This does not negate or diminish the fact that the police have wrongfully manipulated Dan and thus have a reason not to punish him. But in this case, there are plausibly more powerful reasons *in favour* of punishing Dan than in *Drugs*.³⁵ For instance, the state might have powerful duties owed to those harmed or disrespected by Dan’s act to hold him accountable. This is particularly plausible if one is inclined to think that part of the value of punishment lies in its communicative and expressive value.³⁶ This value is not limited to expressing condemnation of wrongdoers or disapproval of wrongful acts. There is also value in standing up for those who have been wronged – the victims – and communicating to them and everyone else that *they* have been wronged and that *they* matter. In that sense, there might also be a *pro tanto* duty owed to victims to punish criminals.³⁷

It is difficult to say when this is weightier than the *pro tanto* duty not to fulfil the manipulative wrong against Dan.³⁸ Furthermore, it is not necessarily the case that the state has to choose between wronging one of the parties. Other, less standard methods, which allow victims to be heard and respected without condemning and harming the wrongdoer, might be possible and preferable.³⁹ However, I leave some of these questions for another time. The main

³⁵ Duff (2010, p. 135) also argues that punishing can be appropriate when the entrapped crime causes harm.

³⁶ For more on the expressive and communicative function and role of punishment, see, for instance, Feinberg (1965) and Duff (2001).

³⁷ My argument here echoes some of the arguments for honour-based justification for defensive harming we have seen earlier. Tadros (2009, p. 410-11), Fritz (2019, p. 324), and Howard and Pasternak (2023, p. 49-50) also make this kind of argument in relation to issues regarding the state’s right to hold socially deprived offenders accountable. See also Alm (2019, pp. 75-80) for a different argument for duties owed to victims to punish offenders.

³⁸ Note, the duty owed to the victim in these cases might also be more stringent than in non-entrapment cases because the state, by entrapping the offender, is complicit in wronging the victim.

³⁹ For instance, in the context of socially deprived offenders, Howard and Pasternak (2023, pp. 51-55) suggest the practice of a “restorative justice conference” as a way for wrongdoers to be held accountable and for victims to get what they are owed, despite the state lacking the standing to blame the criminal.

takeaway of the Manipulation Account is that the intention-manipulation involved in entrapment provides a *pro tanto* directed duty not to punish the targets. This makes it wrong to punish offenders in a variety of cases, even though there may be exceptions when innocent third parties are harmed or disrespected.

4.4 Objections

The Manipulation Account has some potentially untraditional and revisionary implications for entrapment doctrines worth highlighting and defending.

4.4.1 *Private Entrapment*

First, it is often said that any adequate theory of entrapment must avoid “the problem of private entrapment”.⁴⁰ That is, it must not entail that citizens who are encouraged by *other citizens* to commit a crime are victims of entrapment and should not be punished. However, the Manipulation Account may entail just that. Consider, for instance:

Envy: Barry loves Clara, but she loves Adrian. Barry devises a plan to get Adrian out of the picture. He knows that Adrian is looking at serious prison time if he is arrested now because, although he is reformed, Adrian has been arrested for several crimes in the past. Barry calls the police to report a crime and then persuades Adrian to commit a crime. As Adrian commits the crime, the police arrive and arrest him.

This fits the description of entrapment insofar as Barry manipulates Adrian into committing a crime with the aim of having him arrested and punished. Thus, the Manipulation Account suggests that the act of punishing Adrian is also morally tainted.

However, if the crime in question has wronged another innocent person, then there are, as suggested above, victim-centred reasons to punish Adrian. Furthermore, it is easier for the

⁴⁰ See, for instance, Yaffe (2005, p. 4) and Tunick (2011, p. 171).

state to justify punishing Adrian because of these reasons than it is to justify punishment when the state is also the perpetrator of the initial intention-manipulation, such as in *Drugs*. This is because, although everyone has some reason not to fulfil or complete manipulative wrongs (by realising the intended outcomes), the duty not to do it is most stringent for the person or entity responsible for the initial wrong. This seems true of remedial duties in general. If A ends up in possession of a bike that B stole from C, then A ought to give it back. However, the duty to do so is less stringent than the duty B has to ensure that C gets the bike back.⁴¹ It is much easier, in other words, for the state to punish Adrian when entrapped by Barry than if the state was responsible for the entrapment.

Second, if the crime did not wrong an innocent person (say, Adrian just delivered some drugs), then I am inclined to think it might be right for the state not to punish Adrian after all.⁴² It fits with the idea that we should not want it to be easy for wrongdoers to subvert or co-opt our criminal justice system to further their wrongful plans. But that is precisely what we allow if we insist on punishing Adrian. Thus, the implication that punishing perpetrators of non-harmful, privately entrapped crimes might be wrong is not, I submit, necessarily unwelcome.⁴³

4.4.2 *Mere Opportunities*

A second objection holds that the Manipulation Account is at odds with condition (2) of the earlier definition of entrapment, according to which a target is “entrapped” only if, had the

⁴¹ I modify this case from Satz (2012, p. 137).

⁴² This is less controversial than some might think. For instance, Ho (2011, p. 92) seems open to it being wrong to punish in some instances of private entrapment. Dein and Collier (2014, p. 4) and Stark (2018, p. 8) also discuss actual cases in which a stay of prosecution was granted on the basis of private entrapment.

⁴³ Alternatively, one could argue that in private entrapment cases, the state will always have outweighing reasons to punish. Maybe “upholding the rule of law” is sufficient to outweigh the reasons not to punish privately entrapped people. This, again, would not entail that this reason always outweighs the reason not to punish if the state was responsible for entrapping Adrian. As said, the *pro tanto* duty not to fulfil *one’s own* manipulative wrongs (in this case, to punish Adrian) is much more stringent and difficult to justify.

police not incited or enticed them, they would not have committed the crime. To see why this restriction seems intuitive, consider the following:

Pickpocketing: There is a big pickpocketing problem in a certain part of town. The police send undercover officers with wallets sticking out of their pockets to lure out the pickpockets and arrest them. Someone grabs one of the wallets and is arrested when he tries to run away.

The police here just present a “mere opportunity” for someone to break the law. That is proactive policing, not wrongful entrapment. The thief should not be able to escape punishment just because he wrongfully grabbed the opportunity presented by the police. Popular legal doctrines concerning entrapment are designed to account for this. According to the Subjective Test, one cannot use the entrapment defence if one was already “predisposed” to commit the crime. According to the Objective Test, one can use the defence only if the police’s methods and efforts to get one to commit the crime were so severe that most otherwise law-abiding, reasonable citizens would have also given into it.

Neither test would support the thief avoiding punishment on the grounds of entrapment. However, since the police were acting here with a hidden intention to get someone to “bite” and then arrest them, they are engaged in intention-manipulation. Thus, per the Manipulation Account, the state should not punish the thief. In that way, the view also seems incompatible with condition (2).

But the fact that my account diverges from these popular tests should not be counted against it because we should reject them.⁴⁴ It is wrong, I believe, to intention-manipulate even the “predisposed” or easily persuaded if one has no good reason for doing so. Instead, we should agree with Grant Stitt and Gene James when they say, “[n]o one should be offered an

⁴⁴ See also Dworkin (1985, p. 24), Howard (2016, p. 36), and Lippke (2017, pp. 286-90) for scepticism about the tests.

opportunity to commit a crime unless there's probable evidence that he's engaged in ongoing criminal activity" (1984, p. 130). I would add: or unless there is probable evidence that he will commit a similar crime (or a more serious one) at some later point in time, at which point it will be impossible or more difficult for the police to intervene or arrest him.

This points us towards a different test: the Prevention Test. According to this, one cannot use the entrapment defence if the police had sufficient evidence that intention-manipulation was necessary to stop the target's ongoing criminal activities or prevent the target from committing future crimes.⁴⁵ This is because, when the test is met, the entrapment was *justified* on preventive grounds. More precisely, it was justified on liability grounds and not merely lesser evil grounds. If the person was engaged in criminal activity or likely to do so soon, then they were liable to be intention-manipulated to be stopped. They were therefore not *wronged* by the intention-manipulation, and we do not have the ordinary *pro tanto* duty not to fulfil the manipulative wrong by punishing them.⁴⁶

So, although it diverges from the two tests, the Manipulation Account is compatible with, at least the core idea behind, condition (2) and consistent with claiming that presenting a criminal opportunity to thieves and then punishing them in *Pickpocketing* is permissible. The police had a preventive justification for going there and trying to stop the pickpocketing problem.

4.4.3 *Virtue Testing*

A last objection is that the Manipulation Account does not explain why punishing the target in *all* entrapment scenarios is wrong. According to the account, entrapment will morally taint punishment as long as it is aimed at punishment. This is true in a variety of cases. For instance,

⁴⁵ Of course, determining when one has "probable evidence" for this is a difficult question in itself.

⁴⁶ See also Nathan's (2017) argument that people engaged in criminal activity can be liable to the manipulation and deception involved in sting operations.

if the officer hopes that entrapping and punishing the target will reduce crime overall; if the officer is sadistic and just wants to see people punished; or if the officer hopes he can secure a promotion by getting more people arrested and sent to prison. But consider the following:

Virtue Testing: An undercover police officer, Perry, wants to test how virtuous citizens are, so he encourages Dom to commit a crime. He really hopes that Dom will not actually do it but intends to arrest him if he commits the crime.

Although Perry encourages Dom to break the law, his actions do not aim at punishment. He only conditionally intends for it and *hopes* the condition will not be satisfied. Thus, according to the Manipulation Account, the punishment is not morally tainted either. However, some may argue that this is surely an instance of entrapment, and that it is wrong to punish Dom.⁴⁷

Still, the Manipulation Account can at least explain why Perry wrongs Dom. He encourages Dom to act while hiding his true reasons, namely that he is trying to test Dom's virtue. Thus, he is intention-manipulating Dom. That said, I grant that the account, in itself, does not entail that it is wrong to punish him. But I confess that my intuition that punishing is wrong is at least weaker than in paradigmatic cases.

Nonetheless, it might be possible to account for the claim that Dom should not be punished in a different way. It is known from the criminal law that it is sometimes appropriate to treat those who only conditionally intended for something as equivalent to someone who unconditionally intended for the same. Consider, for instance:

Carjacking: Blake approaches Violet in her car and threatens to shoot her unless she hands over her keys. Violet complies and runs away. Later, Blake is arrested and testifies that he did not intend to use the gun if Violet complied but that he would have done so if she had resisted.⁴⁸

⁴⁷ Both Tunick (2011, p. 187) and Dworkin (1985, p. 33) criticise entrapment as illegitimate government virtue testing.

⁴⁸ Inspired by *Holloway v. United States* [1999] 526 U.S. 1.

Is Blake guilty of the U.S. federal crime of carjacking *with the intent to cause death or serious bodily harm*? In one sense, no. He only *conditionally* intended to cause harm. His main plan was simply to scare Violet into handing over the keys. But suppose we add that Blake was confident that Violet would resist and so anticipated that the condition in his conditional intention would be satisfied. In that case, we might lean more towards finding him guilty.

Yaffe (2004) provides a helpful way of thinking about these cases. Recall, on his view, intentions are tied to various rational commitments that structure one's deliberations. If you intend to x soon, then you are rationally committed to various things: to pursue actions and steps essential to x ing, not to form intentions to do things incompatible with those actions, and so on. To simplify a bit, Yaffe builds on this and suggests that we look at how the conditional intention structured Blake's deliberations by comparing the structure of his deliberations to the deliberations of those who carjack with the intention to harm and of those who carjack without such an intention. For instance, if Blake had spent little time contemplating whether Violet would resist, had not been careful to check if the gun was actually loaded, and so on, then we may conclude that he is guided by deliberations more similar to someone who carjacks without an intention to cause harm. However, if he thought resistance was likely, made sure the gun was loaded, had already bought supplies to clean blood spills from the interior, and so on, then he seems guided by deliberations more similar to someone who carjacks with the intent to harm. In the first case, we might reasonably find him not guilty of the federal crime, but in the second, we might reasonably find him guilty.

This provides one way of understanding "purely" virtue testing entrapment cases. Realistically, many such officers, like Perry, are likely to be guided by deliberations similar to those of paradigmatic entrapping officers. For instance, by trying to get Dom to commit the crime, he knows that he is increasing the likelihood of it happening. If he is serious about his conditional intention to arrest him, he is likely also to have deliberated about how to arrest him,

how to prevent escape, made sure there is room for a suspect in the police car, and so on. In that case, treating *Virtue Testing* as equivalent to entrapment unconditionally aimed at punishment may be appropriate.

However, suppose Perry was mostly guided by his hope that Dom would not commit the crime. For instance, he has not made concrete plans for arresting him if the crime is committed. Imagine, indeed, that Perry does not want Dom to be punished, but that he knows he has to do his job of trying to arrest those who commit crimes. So, to increase the likelihood that Perry will go unpunished, he leaves his gun at home, knowingly making it easier for Dom to escape if Perry tries to arrest him. To his surprise, when Dom gives in to the virtue testing, another police officer steps in and arrests Dom. Here, although there is a conditional intention to have him arrested and punished, I am much less inclined to think it is an instance of entrapment which morally taints punishment.⁴⁹

In sum, *Virtue Testing* is not a counterexample to the Manipulation Account. There are ways for the account to treat some of these cases as equivalent to paradigmatic entrapment cases. Furthermore, those which cannot be treated as similar in this way do not seem to be relevantly similar regarding whether punishment is morally tainted either.

5 Defensive Entrapment

Returning to defensive entrapment, I believe the Manipulation Account can help us shed light on cases such as *Fight Wish* and *Infidelity*. When Bree goes into the park and tries to steer Alan towards deciding to attack her, with the intention of creating a situation in which she has to defensively harm him, she is guilty of intention-manipulating him into attacking her. Moreover, given that the manipulation *aims* at that harm, it also morally taints her act of self-defence. In

⁴⁹ The case is inspired by Tadros (2005, p. 318-9).

other words, the answer to Alexander's Challenge is that her defensive rights are compromised by the duty not to complete her intention-manipulation.

The account is also nuanced and compatible with different particular views on exactly what this means for her right to defend herself. First, it seems to me that the Draconian View is attractive in the sense that harming the aggressor will inevitably complete the manipulation wrong. Thus, any defensive action seems morally tainted in some fashion. Second, however, as also highlighted in Chapter 5, this does not entail that the use of any defensive harm will be all things considered impermissible. Plausibly, if the defensive harm is minor and the harm averted is serious enough, it will be disproportionate to demand that the victim not act in self-defence.

But are any defensive options that would be open to a perfect victim ruled out because of the duty not to complete the manipulative wrong? Even if one thinks that any use of defensive harm will wrong the aggressor insofar as it completes that wrong, one might disagree on whether any options are actually made impermissible. The reason is that, according to the Manipulation Account, the aggressor is liable to defensive harm *qua* culpable aggressor, but not liable to defensive harm *qua* victim of intention-manipulation.⁵⁰ Perhaps these dimensions should be treated separately. For instance, one might argue that the Draconian View is true insofar as any defensive action will complete the manipulation, but that the Lenient View is true regarding questions about what the victim may do to stop the wrongful threat. In other words, the entrapment does not change what the victim may permissibly do to the aggressor in self-defence, but it changes the overall moral quality of these permissible actions (that is, by making it so that they necessarily wrong the aggressor).⁵¹

⁵⁰ Borrowing from Quong (2020, p. 146), we might say that they are "partially" liable.

⁵¹ This might be a compromise between Alexander's scepticism that we saw earlier and the intuition that there is *something* problematic about a provocateur's use of defensive force.

However, I believe a version of the Moderate View will also be true here. Defensively harming the aggressor will fulfil the initial manipulative wrong, and the duty not to do so plausibly requires the victim to be willing to sacrifice *some* extra costs to avoid doing so. At the same time, the compromising effect that this will have on their defensive rights is likely weaker than the effect that provocation in the excuse-causing sense has. The reason is that the primary wrongs are importantly different. The wrongness of the provocation has to do with the wrongness of putting someone in danger of suffering substantial harm in a way that seriously limits their ability to avoid such harm. In provocateur cases, by contrast, the victim does not limit the aggressor's fair opportunities to avoid harm. Instead, the primary wrong in these cases is disrespectfully treating the person, or their agency, as a mere means. This difference also creates a difference in how stringent the respective *pro tanto* duties not to impose defensive harm are. For provokers, the reason not to harm the aggressor is the reason not to inflict harm on someone whom one has made vulnerable to suffering it by limiting their ability to avoid it. For provocateurs, however, the reason not to harm is the reason not to aggravate the disrespectful wrong. All else equal, the former seems morally worse than the latter.⁵² Thus, the compromising effect that the acts of provokers and provocateurs have on their defensive rights will also differ in severity.⁵³

Furthermore, cases like *Infidelity* and *Fight Wish* are similar to cases like *Dangerous Party* in that the duty to avoid trouble is violated in all of them. However, the defensive rights of provocateurs are more significantly compromised because of the presence of the intention-

⁵² Of course, these sources of wrongdoing often come together. In many cases, treating people as mere means is partly wrong because it involves harming or putting people at risk of harm in ways they cannot control or avoid. This is not the case in cases like *Infidelity* and *Fight Wish* because the aggressors choose to attack freely and culpably.

⁵³ Note, this is compatible with thinking that provocateurs are often more culpable or at least more nefarious people in virtue of intending and desiring to cause harm.

manipulation, which adds a further directed reason against imposing defensive harm.⁵⁴ Taken together, then, when it comes to defensive entrapment, we ought to conclude with a version of the Moderate View which is somewhere between the minimally compromising Moderate View that holds for imperfect victims who merely violate the duty to avoid trouble and the more radically compromising Moderate View that holds for provokers.

To end, I want to address two potential objections to applying the Manipulation Account to defensive entrapment. One is that Bree is not persuading or encouraging Alan to attack, which is different from entrapment cases like *Drugs*. Thus, she is not entrapping him. However, this objection fails to appreciate that encouragement and persuasion matter for intention-manipulation insofar as they involve trying to steer the target's decision-making by providing reasons to act that one thinks the target is likely to respond to. Bree fits this description even if her entrapping plan only involves *walking* near Alan (and so, no verbal interaction) because she knows that her presence will likely affect his reasoning in the direction of deciding to attack her.

Another objection concerns cases in which the aggressor has already decided to attack the victim. Consider, for instance:

No Retreat: Ben is alerted that Anwar is on his way to kill him. He could avoid a confrontation by going inside for a few hours or arming himself with a tranquilliser gun, but he wants to kill Anwar. So, instead, he arms himself with a gun, locks himself outside, and waits for Anwar to show up.

By deliberately removing the retreat option and least harmful defensive option because he wants to kill Anwar, Ben intuitively compromises the right to defensively harm him. In fact, some may think that the case is so relevantly similar to *Fight Wish* and *Infidelity* that the same

⁵⁴ That the duty to avoid trouble is also violated, and very culpably so, in these cases also matters. However, this alone cannot account for the sense that the defensive rights are more radically compromised, as I argued that the nature of this duty prevents it from significantly compromising defensive rights.

explanation for why the defensive rights are morally tainted should apply to all of them. However, there is no manipulation in this case. In fact, it does not count as entrapment since the case does not satisfy the defensive entrapment version of condition (2) of the definition of entrapment, which says that Ben has entrapped Anwar only if the latter would not have attacked him *but for* Ben's actions. Since Anwar has decided to attack him already, the condition is not satisfied. The Manipulation Account thus faces a problem. In general, it does not explain compromised defensive rights in cases where victims do not instigate attacks but nefariously take advantage of decisions to attack that are already in motion. At best, the account leaves a gap to be filled by another account; at worst, it fails to even explain the original cases insofar as the same explanation should apply to all of them.

An initial reply is to insist that the cases are relevantly different and that other arguments should explain cases like *No Retreat*. For instance, perhaps Ferzan's lack-of-standing-to-defend view should apply here. Alternatively, some of the compromised defensive rights may be explained just by the fact that Ben is particularly culpable for violating the duty to avoid trouble from Chapter 3. Still, I agree that Ben is intuitively also guilty of *wronging* Anwar in a way similar to how (say) Bo wrongs Adil in *Infidelity*.

However, the Manipulation Account might be compatible with this judgement in the end. It seems to me that the case has enough in common with paradigmatic intention-manipulation cases like *Infidelity* and *Fight Wish* for Ben's defensive rights to be morally tainted in the same way.

Recall, I argued that the wrong in intention-manipulation involves subjecting one will to that of another. By steering another person's agency towards a decision with hidden intentions, the manipulator treats the target's agency or decision-making as a mere means – a pawn or instrument – in his own plan. This wrong seems present in some form in *No Retreat* as well. Ben is not guiding Anwar's agency by providing reasons likely to make Anwar decide to attack.

But Ben's plan is to keep Anwar "on course" and not have him change his mind about whether to attack Ben. To achieve this, Ben must also keep his true intentions hidden from Anwar. If he was alerted to the fact that Ben had armed himself and was just waiting for a chance to shoot him, Anwar would likely change course, as he would presumably not willingly walk into a defensive trap. Ben does not steer Anwar's agency by *providing* reasons likely to get Anwar to act as he wants, but by *keeping* him in the dark regarding reasons likely to make him choose contrary to the plan. Thus, by keeping his true intentions hidden, Ben seems guilty of treating Anwar or Anwar's agency as a mere means in his own plan.

In that way, Ben also seems guilty of the wrong identified by the Manipulation Account. Moreover, since the wrong is *aimed* at defensively harming Anwar, the account also explains why Ben's defensive rights are morally tainted.

6 Conclusion

In this chapter, I have looked at provocateurs and their rights to self-defence. I argued that their rights are seriously compromised even though they do not necessarily affect their right not to be attacked or the culpability of the aggressor. Instead, I argued that we should understand the cases as similar to legal entrapment. I then outlined the novel Manipulation Account of entrapment, which holds that entrapment involves intention-manipulation aimed at punishment and therefore grounds a *pro tanto* duty not to punish. After defending this proposal against potential objections, I returned to defensive entrapment and argued that the account offers an intuitive and nuanced view of how provocateurs compromise their defensive rights.

Conclusion

1 Summing Up

This brings us to the end of the thesis. To quickly recap, we have looked at a variety of considerations that may be thought to impact the defensive rights of imperfect victims. In Chapter 2, I argued that imperfect victims do not generally owe it to prospective culpable aggressors to avoid their threats for their sake. In Chapter 3, I argued that they may nevertheless be under a duty to avoid trouble, but that this duty is limited in scope and its impact on defensive rights minimal. In Chapter 4, I rejected the suggestion that imperfect victims have their defensive rights compromised because of their voluntary risk-taking. In Chapter 5, I argued that imperfect victims who provoke aggressors endanger them and thus incur stringent protective duties to avoid imposing defensive harm. Finally, in Chapter 6, I argued that those who entrap culpable aggressors compromise their defensive rights for a different reason, namely because they manipulate the aggressors.

The result is a theory of imperfect victims and self-defence which captures the intuitive view that our rights to self-defence should be robust and provide us protection in a wide variety of cases, but that they can sometimes be, even radically, compromised by our immoral behaviour. The view is distinctive because it offers a unified and nuanced approach with firm theoretical underpinnings. It is based on the more fundamental idea, which I have pursued throughout the thesis, that to determine whether an imperfect victim has compromised their defensive rights, we must look at which, if any, duties concerning the aggressor's interest in avoiding defensive harm the victim disrespected in failing to avoid, or in contributing to, a threat. Such

duties can be grounded in importantly different ways, and investigating those grounds carefully is important because it matters for the extent to which the victim compromises their right to self-defence.

2 Looking Ahead

To end, let me briefly mention some of the future directions for this project. First, in the introductory chapter, I said that a full theory of imperfect victims and the right to self-defence needs to investigate other kinds of cases than the ones we have been looking at. We can now see how the arguments and views laid out in this thesis help us approach that further task. In dealing with cases involving aggressors with different degrees of responsibility and fault – for instance, non-responsible threats, people under duress, or mistaken aggressors – a key question to focus on is what a prospective victim's duties are with respect to the potential defensive harm these aggressors might suffer. Especially Chapter 5, on duties of non-endangerment and fair opportunities to avoid harm, can help us get a clearer picture of such cases.

Second, a related but more applied natural extension concerns police escalations. Sometimes, the presence and actions of police officers can escalate tense situations and make it necessary for the police to use substantial defensive force against a suspect. What should we say about a police officer's defensive rights in these cases? In line with the view presented here, we should approach this topic by investigating the duties of police officers vis-à-vis defensively harming suspects. Of course, the suspect's responsibility for partaking in the escalation will matter. However, there may also be special obligations that police officers owe citizens to avoid situations in which they have to defensively harm them. Pursuing this line of research can help bring the arguments here even further into the practical domain.

Lastly, the thesis may also be relevant to the issue of victim blaming in society more generally. People are often criticised for blaming the victim when they, especially in the

aftermath of sexual assault and rape, are less inclined to offer sympathy and compassion, to think they merit compensation, and/or to view the wrongdoer as less culpable. The criticism is well-placed, and I hope that the holistic line of thinking in Chapter 4 may help us make more sense of exactly what goes wrong here. In particular, focusing on the attitudinal aspect of victim blaming can open an interesting and quite different area of relevance for the claims in that chapter beyond its relevance to matters of defence, compensation, and punishment.

However, I leave these questions for another time.

Bibliography

- Alexander, L. (1999). "A Unified Excuse of Preemptive Self-Protection", *Notre Dame Law Review* 74(5): 1475–1506.
- Alexander, L. (1980). "The Doomsday Machine: Proportionality, Punishment and Prevention", *The Monist* 63(2): 199–227.
- Alexander, L. (1986). "Consent, Punishment, and Proportionality", *Philosophy and Public Affairs* 15(2): 178–182.
- Alexander, L. (1991). "Self-Defense, Punishment, and Proportionality", *Law and Philosophy* 10(3): 323–328.
- Alexander, L. (2013a). "Causing The Conditions of One's Defense: A Theoretical Non-Problem", *Criminal Law and Philosophy*, 7, 623–628.
- Alexander, L. (2013b). "Other People's Errors". *Ethical Theory and Moral Practice* 16: 1049–1059.
- Alexander, L. (2021). "Culpably Creating the Conditions of Justified Acts: Another Look", *Philosophia* 49: 107–112.
- Alexander, L. and Ferzan, K. K. (2009). *Crime and Culpability: A Theory of Criminal Law* (New York: Cambridge University Press).
- Alm, D. (2019). "Crime Victims and the Right to Punishment", *Criminal Law and Philosophy* 13: 63–81.
- Anderson, E. S. (1999). "What is the Point of Equality?", *Ethics* 109(2): 287–337.
- Amir, M. (1971). *Patterns in Forcible Rape*. Chicago: University of Chicago Press.
- Arneson, R. J. (1989). "Equality and Equal Opportunity for Welfare", *Philosophical Studies* 56: 77–93.
- Arneson, R. J. (2004). "Luck Egalitarianism Interpreted and Defended", *Philosophical Topics* 32(1/2): 1–20.
- Arneson, R. J. (2018). "Self-Defense and Culpability: Fault Forfeits First", *San Diego Law Review* 55: 230–264.
- Ashworth, A. J. (1976). "The Doctrine of Provocation", *The Cambridge Law Journal* 35: 292–320.
- Ashworth, A. J. (1999). "What is Wrong with Entrapment?", *Singapore Journal of Legal Studies* 40: 293–317.
- Bader, R. (2016). "Conditions, Modifiers, and Holism," in E. Lord and B. Maguire (eds.), *Weighing Reasons* (Oxford: Oxford University Press): 27–55.
- Barry, C. and Øverland, G. (2014). "The Implications of Failing to Assist," *Social Theory and Practice* 40(4): 570–590.
- Bazargan-Forward, S. (2018). "Dignity, Self-Respect, and Bloodless Invasions", in B. J. Strawser, R. Jenkins, and M. Robillard (eds.), *Who Should Die? The Ethics of Killing in War* (Oxford: Oxford University Press), 142–162.
- Benbaji, Y. (2005). "Culpable Bystanders, Innocent Threats and the Ethics of Self-Defense," *Canadian Journal of Philosophy*, 35(4), 585–622.
- Bergelson, V. (2005). "Victims and Perpetrators: An Argument for Comparative Liability in Criminal Law," *Buffalo Criminal Law Review* 8(2): 385–487.
- Bergelson, V. (2009). *Victims' Rights and Victims' Wrongs: Comparative Liability in Criminal Law* (Stanford: Stanford University Press).
- Berman, M. N. and Farrell, I. P. (2011). "Provocation Manslaughter as Partial Justification and Partial Excuse", *William and Mary Law Review* 52: 1027–1109.

- Billingham, P. and Parr, T. (2020a). "Online Public Shaming: Virtues and Vices," *Journal of Social Philosophy* 51(3): 343–500.
- Billingham, P. and Parr, T. (2020b). "Enforcing Social Norms: The Morality of Public Shaming," *European Journal of Philosophy* 28(4): 839–1109.
- Bloomfield, P. (2008). "The Harm of Immorality," *Ratio* 21(3): 241–256.
- Bloomfield, P. (2010). "Why It's Bad to be Bad," in P. Bloomfield (ed.) *Morality and Self-Interest* (Oxford: Oxford University Press): 251–271.
- Bloomfield, P. (2017). "Morality is Necessary for Happiness," *Philosophical Studies* 174(10): 2613–2628.
- Bowen, J. (2016). "Necessity and Liability: On an Honour-Based Justification for Defensive Harming," *Journal of Practical Ethics* 4(2): 79–93.
- Boxill, B. R. (1976). "Self-Respect and Protest," *Philosophy and Public Affairs* 6(1): 58–69.
- Bratman, M. (1987). *Intention, Plans and Practical Reason*. Cambridge: Harvard University Press.
- Brennan, S. (1995). "How Is the Strength of a Right Determined? Assessing The Harm View," *American Philosophical Quarterly* 32(4): 383–393.
- Brink, D. O. (2021). *Fair Opportunity and Responsibility*. Oxford: Oxford University Press.
- Brown, C. (2020). "Sex Crimes and Misdemeanours," *Philosophical Studies* 177: 1363–1379.
- Brownlee, K. (2019). "Acting Defensively for the Sake of Our Attacker," *Journal of Moral Philosophy* 16: 105–130.
- Bublick, E. M. (1999). "Citizen No-Duty Rules: Rape Victims and Comparative Fault," *Columbia Law Review* 99(6): 1413–1490.
- Buchak, L. (2014). "Risk and Tradeoffs," *Erkenntnis* 79(6): 1091–1111.
- Burri, S. (2020). "Morally Permissible Risk Imposition and Liability to Defensive Harm," *Law and Philosophy* 39: 381–408.
- Burri, S. (2022). "Defensive Liability: A Matter of Rights Enforcement, not Distributive Justice," *Criminal Law and Philosophy* 16: 539–553.
- Burri, S. and Christie, L. (2019). "On the Enforceability of Poverty-Related Responsibilities," *Ethics and Global Politics* 12: 68–75.
- Buss, S. (2005). "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints," *Ethics* 115(2): 195–235.
- Carlson, A. (2007). "Entrapment, Punishment, and the Sadistic State," *Virginia Law Review* 93: 1081–1134.
- Chandler, D. and Munday, R. (2016). *A Dictionary of Social Media*. Oxford: Oxford University Press.
- Chisholm, R. and Feehan, T. (1977). "The Intent to Deceive," *Journal of Philosophy* 74(3): 143–159.
- Christie, L. (2020). "Causation and Liability to Defensive Harm," *Journal of Applied Philosophy* 37(3): 378–392.
- Christie, N. (1986). "The Ideal Victim," in Ezzat A. Fattah (ed), *From Crime Policy to Victim Policy* (London: Palgrave Macmillan): 17–30.
- Clark, D. J. (2023). "Refusing Protection," *Philosophy and Public Affairs* 51(1): 33–59.
- Cohen, G. A. (1989). "On the Currency of Egalitarian Justice," *Ethics* 99: 906–44.
- Cohen, G. A. 2006. Casting the First Stone: Who Can, and Who Can't, Condemn the Terrorists? *Royal Institute of Philosophy Supplement* 58: 113–136.
- Cohen, G. A. (2008). *Rescuing Justice and Equality*. Cambridge: Harvard University Press.
- Cornell, N. (2016). "A Complainant-Oriented Approach to Unconscionability and Contract Law," *University of Pennsylvania Law Review* 164(5): 1131–1176.
- Cruft, R. (2013). "Why Is It Disrespectful to Violate Rights?" *Proceedings of the Aristotelian Society* 113(2): 201–224.

- Dancy, J. (2004). *Ethics Without Principles* (Oxford: Clarendon Press).
- Dein, J. and Collier, L. V. (2014). "Non-State Agent Entrapment – The X Factor," *Archbold Review* 9: 4–6.
- Dempsey, M. M. (2016). "Against Liability: Toward a Reasons-Based Account of Self-Defense", in C. Coons & M. Weber (eds.), *The Ethics of Self-Defense* (Oxford: Oxford University Press): 211–232.
- Doggett, T. (2017). "Killing Innocent People", *Noûs* 52(3): 645–666.
- Doggett, T. (2020). "Letting Others Do Wrong", *Noûs* 56(1): 40–56.
- Draper, K. (1993). "Fairness and Self-Defense", *Social Theory and Practice* 19(1), 73–92.
- Draper, K. (2009). "Defense", *Philosophical Studies* 145: 69–88.
- Draper, K. (2016a). *War and Individual Rights: The Foundations of Just War Theory*. Oxford: Oxford University Press.
- Draper, K. (2016b). "Necessity and Proportionality in Defense", in C. Coons and M. Weber (eds.), *The Ethics of Self-Defense* (New York: Oxford University Press): 171–184.
- Draper, K. (2018). "Defensive Liability: Four Common Mistakes", in B. J. Strawser, R. Jenkins, and M. Robillard (eds.), *Who Should Die? The Ethics of Killing in War* (Oxford: Oxford University Press): 80–101.
- Dressler, J. (1982). "Rethinking Heat of Passion: A Defense in Search of A Rationale", *Journal of Criminal Law and Criminology* 73: 421–470.
- Dressler, J. (1988). "Provocation: Partial Justification or Partial Excuse?", *The Modern Law Review* 51: 476–480.
- Driver, J. (1992). "The Suberogatory", *Australasian Journal of Philosophy* 70 (3): 286–295.
- Duff, R. A. (2001). *Punishment, Communication and Community*. Oxford: Oxford University Press.
- Duff, R. A. (2010). "Blame, Moral Standing and The Legitimacy of The Criminal Trial", *Ratio* 23(2): 123–140.
- Duff, A., L. Farmer, S. Marshall, and V. Tadros. (2007). *The Trial on Trial Volume 3*. Cornwall: Hart Publishing.
- Duus–Otterström, G. (2017). "Benefiting from Injustice and the Common–Source Problem", *Ethical Theory and Moral Practice* 20: 1067–1081.
- Duus–Otterström, G. and Page, E. A. (forthcoming). "Defeating Wrongdoing: Why Victims of Unjust Harm Should Take Priority over Victims of Bad Luck", *Social Theory and Practice*.
- Dworkin, G. (1985). "The Serpent Beguiled Me and I Did Eat: Entrapment and The Creation of Crime", *Law and Philosophy* 4(1): 17–39.
- Dworkin, R. (2000). *Sovereign Virtue*. Cambridge: Harvard University Press.
- Eigenberg, H. and T. Garland. (2003). "Victim Blaming", in L. J. Moriarty (ed.) *Controversies in Victimology* (Newark: LexisNexis): 21–36.
- Eldar, S. and Laist, E. (2014). "The Misguided Concept of Partial Justification", *Legal Theory* 20: 157–185).
- Enoch, D. (2018). "Against Utopianism: Noncompliance and Multiple Agents", *Philosophers' Imprint* 18(16): 1–20.
- Eskens, R. (unpublished manuscript). "Reciprocity, Inequality, and Unsuccessful Rescues", on file with author.
- Eyal, N. (2007). "Egalitarian Justice and Innocent Choice", *Journal of Ethics and Social Philosophy* 2(1): 1–18.
- Fabre, C. (2000). *Social Rights Under the Constitution*. Oxford: Oxford University Press.
- Fabre, C. (2012). *Cosmopolitan War*. Oxford: Oxford University Press.
- Fabre, C. (2022). "Harming, Rescuing and the Necessity Constraint on Defensive Force", *Criminal Law and Philosophy* 16: 525–538.

- Fabre, C. and Lazar, S. eds. (2014). *The Morality of Defensive War*. Oxford: Oxford University Press.
- Farrell, D. (1985). "The Justification of General Deterrence", *The Philosophical Review* 94(3): 367–394.
- Farrell, D. (1990). "The Justification of Deterrent Violence", *Ethics* 100(2): 301–317.
- Farrell, D. (1995). "Deterrence and The Just Distribution of Harm", *Social Philosophy and Policy* 12(2): 220–240.
- Farrell, D. (2013). "What Should We Say about Contrived "Self-Defense" Defenses?", *Criminal Law and Philosophy*, 7, 571–585.
- Feinberg, J. (1965). "The Expressive Function of Punishment", *The Monist* 49: 397–423.
- Ferzan, K. K. (2005). "Justifying Self-Defense", *Law and Philosophy* 24: 711–749.
- Ferzan, K. K. (2012). "Culpable Aggression: The Basis for Moral Liability to Defensive Killing", *Ohio State Journal of Criminal Law* 9: 669–697.
- Ferzan, K. K. (2013). "Provocateurs", *Criminal Law and Philosophy* 7: 597–622.
- Ferzan, K. K. (2016). "Forfeiture and Self-Defense", in C. Coons & M. Weber (eds.), *The Ethics of Self-Defense* (Oxford: Oxford University Press): 233–253.
- Ferzan, K. K. (2018). "Defending Honor and Beyond: Reconsidering the Relationship between Seemingly Futile Defense and Permissible Harming", *Journal of Moral Philosophy* 15: 683–705.
- Ferzan, K. K. (2019). "Stand Your Ground", in L. Alexander and K. K. Ferzan (eds), *The Palgrave Handbook of Applied Ethics and the Criminal Law* (Cham: Palgrave Macmillan): 731–750.
- Ferzan, K. K. (2020). "Losing the Right to Assert You've Been Wronged", in P. B. Miller and J. Oberdiek (eds.), *Civil Wrongs and Justice in Private Law* (Oxford: Oxford University Press): 111–129.
- Finlay, C. J. (2015). *Terrorism and The Right to Resist: A Theory of Just Revolutionary War*. Cambridge: Cambridge University Press.
- Firth, J. M. and Quong, J. (2012). "Necessity, Moral Liability, and Defensive Harm", *Law and Philosophy* 31: 673–701.
- Fishback, I. (2016). "Necessity and Institutions in Self-Defense and War", in C. Coons & M. Weber (eds.), *The Ethics of Self-Defense* (Oxford: Oxford University Press): 274–291.
- Flanigan, J. (2014). "A Defense of Compulsory Vaccination", *HEC Forum* 26: 5–25.
- Flanigan, J. (2019). "Duty and Enforcement", *Journal of Political Philosophy* 27(3): 341–362.
- Frick, J. (2016). "What We Owe to Hypocrites: Contractualism and The Speaker-Relativity of Justification", *Philosophy and Public Affairs* 44(4): 223–265.
- Fritz, K. (2019). "Hypocrisy, Inconsistency, and the Moral Standing of the State", *Criminal Law and Philosophy* 13: 309–327.
- Frowe, H. (2014). *Defensive Killing*. Oxford: Oxford University Press.
- Frowe, H. (2016). "The Role of Necessity in Liability to Defensive Harm," in C. Coons and M. Weber (eds.), *The Ethics of Self-Defense* (Oxford: Oxford University Press): 152–170.
- Frowe, H. (2021). "The Moral Irrelevance of Moral Coercion", *Philosophical Studies* 178: 3465–3482.
- Frowe, H. (2022). "Intervening Agency and Civilian Liability", *Criminal Law and Philosophy* 16: 181–191.
- Frowe, H. and Lang, G. eds. (2014). *How We Fight*. Oxford: Oxford University Press.
- Frowe, H. and Parry, J. (2019). "Wrongful Observation", *Philosophy and Public Affairs* 47(1): 104–137.
- Gardner, J. (2011). "What is Tort Law For? Part I. The Place of Corrective Justice", *Law and Philosophy* 30: 1–50.
- Garnett, M. (2018). "Coercion: The Wrong and The Bad", *Ethics* 128: 545–573.

- Gill, M. (2022). “#MeToo is over if we don’t listen to “imperfect victims” like Amber Heard,” *The Guardian* May 22, 2022: <https://www.theguardian.com/commentis-free/2022/may/22/metoo-is-over-if-we-dont-listen-to-imperfect-victims-like-amber-heard>
- Go, J. (2021). “Two-Level Luck Egalitarianism: Reconciling Rights, Respect, and Responsibility”, *Journal of Value Inquiry* 55: 543–566.
- Gordon–Solmon, K. (2018). “What Makes a Person Liable to Defensive Harm?”, *Philosophy and Phenomenological Research* 97(3): 543–567.
- Gorin, M. (2014). “Towards a Theory of Interpersonal Manipulation”, in C. Coons and M. Weber (eds.), *Manipulation: Theory and Practice* (Oxford: Oxford University Press): 73–97.
- Haeg, J. (2022). “Entrapment and Manipulation”, *Res Publica* 28(4): 557–583.
- Halstead, J. (2016). “The Numbers Always Count”, *Ethics* 126: 789–802.
- Haque, A. A. (2015). *Law and Morality at War*. Oxford: Oxford University Press.
- Hecht, L. (2019a). “Provocateurs and Their Rights to Self-Defence”, *Criminal Law and Philosophy* 13: 165–185.
- Hecht, L. (2019b). *Victims of War*. PhD Thesis, Stockholm University.
- Hecht, L. (2022). “Activating the Right to Be Rescued”, *Journal of Moral Philosophy* (Online First): 1–24.
- Hereth, B. (2017). “Against Self-Defense”, *Social Theory and Practice* 43(3): 613–635.
- Hereth, B. (2022). “Honorable Survivors: A Feminist Reply to Statman”, *Public Affairs Quarterly* 36(2): 121–135.
- Hidalgo, J. (2015). “Resistance to Unjust Immigration Restrictions”, *Journal of Political Philosophy* 23(4): 450–470.
- Hidalgo, J. (2019). *Unjust Borders: Individuals and The Ethics of Immigration*. New York: Routledge.
- Hill, D. J., McLeod, S. K., and Tanyi, A. (2018). “The Concept of Entrapment”, *Criminal Law and Philosophy* 12: 539–554.
- Ho, H. L. (2011). “State Entrapment”, *Legal Studies* 31: 71–95.
- Hohfeld, W. (1919). *Fundamental Legal Conceptions*, W. Cook (ed.). New Haven: Yale University Press.
- Holton, S. and Shute, S. (2007). “Self-Control in the Modern Provocation Defence”, *Oxford Journal of Legal Studies* 27(1): 49–73.
- Horder, J. (1992). *Provocation and Responsibility*. Oxford: Clarendon Press.
- Horton, J. (2018). “Always Aggregate”, *Philosophy and Public Affairs* 46(2): 160–174.
- Hosein, A. (2017). “Responsibility and Self-Defense: Can We Have It All?”, *Res Publica* 23: 367–385.
- Howard, J. (2016). “Moral Subversion and Structural Entrapment”, *Journal of Political Philosophy* 24 (1): 24–46.
- Howard, J. (2018). “Kidnapped: The Ethics of Paying Ransoms”, *Journal of Applied Philosophy* 35(4): 675–688.
- Howard, J. (2019a). “Dangerous Speech”, *Philosophy and Public Affairs* 47(2): 208–254.
- Howard, J. (2019b). “Free Speech and Hate Speech”, *Annual Review of Political Science* 22: 93–109.
- Howard, J. and Pasternak, A. (2023). “Criminal Wrongdoing, Restorative Justice, and the Moral Standing of Unjust States”, *Journal of Political Philosophy* 31(1): 42–59.
- Hughes, P. M. (2004). “What Is Wrong with Entrapment?”, *The Southern Journal of Philosophy* 42: 45–60.
- Hurd, H. M. (2001). “Is It Wrong to Do Right When Others Do Wrong?”, *Legal Theory* 7: 307–340.

- Hurd, H. M. (2005). "Blaming the Victim: A Response to the Proposal That Criminal Law Recognize a General Defense of Contributory Responsibility", *Buffalo Criminal Law Review* 8(2): 503–522.
- Hurd, H. M. (2016) "Stand Your Ground", in Christian Coons and Michael Weber (eds.), *The Ethics of Self-Defense* (Oxford: Oxford University Press), 254–273.
- Hurka, T. (2005). "Proportionality in the Morality of War", *Philosophy and Public Affairs* 33(1): 34–66.
- Hurka, T. (2019). "More Seriously Wrong, More Importantly Right", *Journal of the American Philosophical Association* 5(1): 41–58.
- Kamm, F. M. (1993). *Morality, Mortality Volume 1*. Oxford: Oxford University Press.
- Kamm, F. M. (2014). "Self-Defense, Resistance, and Suicide: The Taliban Women", in H. Frowe and G. Lang (eds.), *How We Fight* (Oxford: Oxford University Press): 75–86.
- Kamm, F. M. (2022). *Rights and Their Limits*. Oxford: Oxford University Press.
- Karhu, T. (2019). "Non-Compensable Harms", *Analysis* 79(2): 222–230.
- Karmen, A. (1991). "The Controversy over Shared Responsibility: Is Victim-Blaming Ever Justified?" in D. Sank and D. I. Caplan (eds.), *To Be a Victim: Encounters with Crime and Injustice* (New York: Springer): 395–408.
- Karnein, A. (2014). "Putting Fairness in Its Place: Why There is a Duty to Take Up The Slack", *Journal of Philosophy* 111(11): 593–607.
- Katz, L. (1990). "The Assumption of Risk Argument", *Social Philosophy and Policy* 7(2): 138–169.
- Katz, L. (2013). "Entrapment Through the Lens of Actio Libera in Causa", *Criminal Law and Philosophy* 7: 587–595
- Kaufman, W. (2008). "Torture and the "Distributive Justice" Theory of Self-Defense: An Assessment", *Ethics and International Affairs* 22(1): 93–115.
- Kerstein, S. J. (2013). *How To Treat Persons*. Oxford: Oxford University Press.
- Kernohan, A. (2022). "How To Modify The Strength of a Reason", *Philosophical Studies* 179: 1205–1220.
- Kim, H. (2019). "Entrapment, Culpability, and Legitimacy", *Law and Philosophy* 39: 67–91.
- Lang, G. (2014). "Why Not Forfeiture?", in H. Frowe and G. Lang (eds.), *How We Fight* (Oxford: Oxford University Press): 38–61.
- Lang, G. (2017). "What Follows from Defensive Non-Liability?", *Proceedings of the Aristotelian Society* 117(3): 231–252.
- Lang, G. (2021). *Strokes of Luck: A Study in Moral and Political Philosophy*. Oxford: Oxford University Press.
- Lang, G. (2022). "Defensive Escalations", *Journal of Ethics* 26: 273–294.
- Lazar, S. (2009). "The Nature and Disvalue of Injury", *Res Publica* 15: 289–304.
- Lazar, S. (2012). "Necessity in Self-Defense and War", *Philosophy and Public Affairs* 40(1): 1–77.
- Lazar, S. (2014). "National Defence, Self-Defence, and The Problem of Political Aggression", in C. Fabre and S. Lazar (eds.), *The Morality of Defensive War* (Oxford: Oxford University Press): 11–39.
- Lazar, S. (2015). *Sparing Civilians*. Oxford: Oxford University Press.
- Lazar, S. (2016). "Authorization and The Morality of War," *Australasian Journal of Philosophy* 94(2): 211–226.
- Lazar, S. (2017a). "Just War Theory: Revisionists Versus Traditionalists", *Annual Review of Political Science* 20: 37–54.
- Lazar, S. (2017b). "Response: Limiting Defensive Rights", *Journal of Applied Philosophy*, 34(1): 19–23.
- Leverick, F. (2006). *Killing in Self-Defence*. Oxford: Oxford University Press.

- Lippke, R. L. (2017). "A Limited Defense of What Some Will Regard as Entrapment", *Legal Theory* 23: 283–306.
- Löschke, J. (2021). "Agent-Relative Reasons and Normative Force", *Philosophia* 49: 259–372.
- Macklem, T. and Gardner, J. (2001). "Provocation and Pluralism", *The Modern Law Review* 64: 815–830.
- MacLagan, W. G. (1955). "How Important is Moral Goodness?", *Mind* 64(254): 213–225.
- Martin, A. (2021). "Personal Bonds: Directed Obligations without Rights", *Philosophy and Phenomenological Research* 102(1): 65–86.
- McAuley, F. (1987). "Anticipating the Past: The Defence of Provocation in Irish Law", *The Modern Law Review* 50: 133–157.
- McMahan, J. (1994). "Innocence, Self-Defence and Killing in War", *Journal of Political Philosophy* 2(3), 193–221.
- McMahan, J. (2004a). "War as Self-Defense", *Ethics and International Affairs* 18(1): 75–80.
- McMahan, J. (2004b). "The Ethics of Killing in War", *Ethics* 114: 693–733.
- McMahan, J. (2005a). "The Basis of Moral Liability to Defensive Killing", *Philosophical Issues* 15(1): 386–405.
- McMahan, J. (2005b). "Self-Defense and Culpability", *Law and Philosophy* 24: 751–774.
- McMahan, J. (2009). *Killing In War*. Oxford: Clarendon Press.
- McMahan, J. (2010). "Humanitarian Intervention, Consent, and Proportionality," in A. Davis, R. Keshen, and J. McMahan (eds.), *Ethics and Humanity: Themes from the Philosophy of Jonathan Glover* (Oxford: Oxford University Press): 44–72.
- McMahan, J. (2012). "Individual Liability in War: A Response to Fabre, Leveringhaus and Tadros", *Utilitas* 24(2): 278–299.
- McMahan, J. (2014). "What Rights May Be Defended by Means of War?" in C. Fabre and S. Lazar (Eds.), *The Morality of Defensive War* (Oxford: Oxford University Press): 115–156.
- McMahan, J. (2016). "The Limits of Self-Defense," in C. Coons and M. Weber (eds.), *The Ethics of Self-Defense* (Oxford: Oxford University Press): 185–211.
- McMahan, J. (2017). "Proportionate Defense," in J. D. Ohlin, L. May, and C. Finkelstein (eds.), *Weighing Lives in War* (Oxford: Oxford University Press): 132–152.
- Mendelsohn, B. (1976). "Victimology and Contemporary Social Trends", *Victimology* 1(1): 8–28.
- Miller, D. (2020). "The Nature and Limits of the Duty of Rescue", *Journal of Moral Philosophy* 17(3): 322–23.
- Miller, S. (1992). "Self-Defence and Forcing the Choice between Lives", *Journal of Applied Philosophy* 9(2): 239–243.
- Montague, P. (1981). "Self-Defense and Choosing Between Lives", *Philosophical Studies*, 40(2): 207–219.
- Montague, P. (1989). "The Morality of Self-Defense: A Reply to Wasserman", *Philosophy and Public Affairs* 18(1): 81–89.
- Montague, P. (1995). *Punishment as Societal Defense*. Maryland: Rowan & Littlefield Publishers, Inc.
- Montague, P. (2010). "Self-Defense, Culpability, and Distributive Justice", *Law and Philosophy* 29(1): 75–91.
- Mund, J. (2021). "Buying Humane Animal Products is Wrong Because it Wrongs the Farmer", *Public Affairs Quarterly* 35(4): 315–337.
- Muñoz, D. and T. Pummer. (2022). "Supererogation and Conditional Obligation", *Philosophical Studies* 179: 1429–1443.
- Narveson, J. (1965). "Pacifism: A Philosophical Analysis", *Ethics* 75: 259–71.

- Nathan, C. (2017). "Liability to Deception and Manipulation: The Ethics of Undercover Policing", *Journal of Applied Ethics* 34: 370–388
- Nino, C. S. (1986). "Does Consent Override Proportionality?", *Philosophy & Public Affairs* 15(2): 183–187.
- Nozick, R. (1974). *Anarchy, State and Utopia*. New York: Basic Books.
- Oberman, K. (2013). "Can Brain Drain Justify Immigration Restrictions?", *Ethics* 123(3): 427–455.
- Otsuka, M. (1994). "Killing the Innocent in Self-Defense", *Philosophy & Public Affairs* 1, 23(1): 74–94.
- Otsuka, M. (2016). "The Moral Responsibility Account of Liability to Defensive Killing", in C. Coons and M. Weber (eds.), *The Ethics of Self-Defense* (Oxford: Oxford University Press): 51–68.
- Owens, D. (2008). "Rationalism about Obligation", *European Journal of Philosophy* 16(3): 341–482.
- Owens, D. (2012). *Shaping the Normative Landscape*. Oxford: Oxford University Press.
- Øverland, G. (2010). "Conditional Threats", *Journal of Moral Philosophy* 7: 334–345.
- Øverland, G. (2011a). "Dividing Harm", *Journal of Moral Philosophy* 8(2): 547–566.
- Øverland, G. (2011b). "Moral Taint: On the Transfer of the Implications of Moral Culpability", *Journal of Applied Philosophy* 28(2): 122–136.
- Pasternak, A. (2019). "Political Rioting: A Moral Assessment", *Philosophy and Public Affairs* 46(4): 384–418.
- Parr, T. (2016). "The Moral Taintedness of Benefiting from Injustice", *Ethical Theory Moral Practice* 19: 985–997.
- Peterson, S. (1991). "Victimology and Blaming the Victim: The Case of Rape," in D. Sank and D. I. Caplan (eds.) *To Be a Victim: Encounters with Crime and Injustice* (New York: Springer): 171–178.
- Eigenberg, H. and Garland, T. (2003). "Victim Blaming", in L. J. Moriarty (ed.) *Controversies in Victimology* (Newark: LexisNexis): 21–36.
- Pummer, T. (2019). "All or Nothing, But If Not All, Next Best or Nothing", *Journal of Philosophy* 116(5): 278–291.
- Pummer, T. (2022). "Contrastive Consent and Secondary Permissibility", *Philosophy and Phenomenological Research* Online First: 1–15.
- Pummer, T. (2023). *The Rules of Rescue: Cost, Distance, and Effective Altruism*. Oxford: Oxford University Press.
- Quinn, W. (1985). "The Right to Threaten and the Right to Punish", *Philosophy and Public Affairs* 14(4): 327–373.
- Quinn, W. (1989). "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing", *The Philosophical Review* 98(3): 287–312.
- Quong, J. (2012). "Liability to Defensive Harm", *Philosophy and Public Affairs* 40(1): 45–77.
- Quong, J. (2015). "Proportionality, Liability, and Defensive Harm", *Philosophy and Public Affairs* 43(2): 145–173.
- Quong, J. (2020). *The Morality of Defensive Force*. Oxford: Oxford University Press.
- Quong, J. (2022). "The Morality of Defensive Force: Replies to Otsuka, Frowe, Fabre, and Burri", *Criminal Law and Philosophy* (Online First): 1–20.
- Ramakrishnan, K. (2016). "Treating People as Tools", *Philosophy and Public Affairs* 44(2): 134–165.
- Raymond, M. (2010). "Looking for Trouble: Framing and the Dignitary Interest in the Law of Self-Defence", *Ohio State Law Journal* 71: 287–339.
- Raz, J. (1984). "On the Nature of Rights", *Mind* 93(370): 194–214.
- Raz, J. (1986). *The Morality of Freedom*. Oxford: Oxford University Press.

- Raz, J. (1994). *Ethics in the Public Domain*. Oxford: Oxford University Press.
- Raz, J. (1999). *Practical Reason and Norms*. Oxford: Oxford University Press.
- Renzo, M. (2017). "Rights Forfeiture and Liability to Harm", *Journal of Political Philosophy* 25(3): 324–342.
- Renzo, M. (2019). "Political Authority and Unjust Wars", *Philosophy and Phenomenological Research* 99(2): 336–357.
- Renzo, M. (2021). "Manipulation and Liability to Defensive Harming", *Philosophical Studies* 178: 3483–3501.
- Ripstein, A. (2016). *Private Wrongs*. Harvard: Harvard University Press.
- Robillard, M. (2018). "Fighting For One's Self", in B. J. Strawser, R. Jenkins, and M. Robillard (eds.), *Who Should Die? The Ethics of Killing in War* (Oxford: Oxford University Press): 103–117.
- Robinson, P. H. (1985). "Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine", *Virginia Law Review* 71(1): 1–63.
- Rodin, D. (2002). *War and Self-Defense*. Oxford: Oxford University Press.
- Rodin, D. (2011). "Justifying Harm", *Ethics* 122(1): 74–110.
- Rodin, D. (2014a). "The Myth of National Defence", in C. Fabre and S. Lazar (eds.), *The Morality of Defensive War* (Oxford: Oxford University Press): 69–89.
- Rodin, D. (2017). "The Lesser Evil Obligation", in C. Coons and M. Weber (eds.), *The Ethics of War* (Oxford: Oxford University Press): 28–46.
- Roemer, J. (1995). "Equality and Responsibility", *Boston Review* 20: 3–7.
- Rozelle, S. D. (2005). "Controlling Passion: Adultery and The Provocation Defense", *Rutgers Law Journal* 37: 197–233.
- Satz, D. (2012). "Countering The Wrongs of The Past: The Role of Compensation", *Nomos* 51: 129–150.
- Schopp, R. F. (1998). *Justification Defenses and Just Convictions*. Cambridge: Cambridge University Press.
- Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Scanlon, T. M. (2014). *Being Realistic about Reasons*. Oxford: Oxford University Press.
- Schafer, S. (1968). *The Victim and His Criminal: A Study in Functional Responsibility*. New York: Random House.
- Schwartz, D. (2020). "Necessity Historically Considered", *Journal of Moral Philosophy* 17: 591–605.
- Segev, R. (2005). "Fairness, Responsibility and Self-Defense", *Santa Clara Law Review*, 45(2), 383–460.
- Singer, P. (1972). "Famine, Affluence, and Morality", *Philosophy and Public Affairs* 1 (3): 229–243.
- Slavny, A. (2014). "Negating and Counterbalancing: A Fundamental Distinction in The Concept of a Corrective Duty", *Law and Philosophy* 33: 143–173.
- Stark, F. (2018). "Non-State Entrapment", *Archbold Review* 10: 6–9.
- Statman, D. (2008). "On the Success Condition for Legitimate Self-Defense", *Ethics* 118(4): 659–686.
- Statman, D. and Benbaji, Y. (2021). "Against Moral Taint", *Ethical Theory and Moral Practice* 24: 5–18.
- Steel, S. (2020). "Compensation and Continuity", *Legal Theory* 26: 250–279.
- Stefánsson, H. O. and R. Bradley. (2019). "What Is Risk Aversion?", *British Journal for the Philosophy of Science* 70(1): 77–102.
- Steinhoff, U. (2006). "Torture – The Case for Dirty Harry and against Alan Dershowitz", *Journal of Applied Philosophy* 23(3): 337–353.

- Steinhoff, U. (2013). “Rodin on Self-Defense and the “Myth” of National Self-Defense: A Refutation”, *Philosophia* 41: 1017–1036.
- Steinhoff, U. (2017). “Proportionality in Self-Defense”, *The Journal of Ethics* 21(3): 263–289.
- Stemplowska, Z. (2009). “Making Justice Sensitive to Responsibility”, *Political Studies* 57: 237–259.
- Stemplowska, Z. (2013). “Rescuing Luck Egalitarianism”, *Journal of Social Philosophy* 44(4): 402–419.
- Stemplowska, Z. (2016). “Doing More Than One’s Fair Share”, *Critical Review of International Social and Political Philosophy* 19(5): 591–608.
- Stitt, G. and James, G. (1974). “Entrapment and The Entrapment Defense: Dilemmas for a Democratic Society”, *Law and Philosophy* 3: 111–131.
- Strawser, B. J., Jenkins, R., and Robillard, M. (eds.). (2017). *Who Should Die? The Ethics of Killing in War*. Oxford: Oxford University Press.
- Sussman, D. (2005). “What’s Wrong with Torture?”, *Philosophy and Public Affairs* 33(1): 1–33.
- Tadros, V. (2009). “Poverty and Criminal Responsibility”, *Journal of Value Inquiry* 43: 391–413.
- Tadros, V. (2011). *The Ends of Harm: The Moral Foundations of Criminal Law*. Oxford: Oxford University Press.
- Tadros, V. (2012). “Duty and Liability”, *Utilitas* 24(2): 259–277.
- Tadros, V. (2014). “Resource Wars”, *Law and Philosophy* 33: 361–389.
- Tadros, V. (2016a). *Wrongs and Crimes*. Oxford: Oxford University Press.
- Tadros, V. (2016b). “Permissibility in a World of Wrongdoing”, *Philosophy and Public Affairs* 44(2): 101–132.
- Tadros, V. (2018). “The Moral Distinction between Combatants and Noncombatants: Vulnerable and Defenceless”, *Law and Philosophy* 37: 289–312.
- Tadros, V. (2020a). “Distributing Responsibility”, *Philosophy and Public Affairs* 48(3): 223–261.
- Tadros, V. (2020b). “Secondary Duties,” in P. B. Miller and J. Oberdiek (eds.), *Civil Wrongs and Justice in Private Law* (Oxford: Oxford University Press): 186–207.
- Tadros, V. (2020c). *To Do, To Die, To Reason Why: Individual Ethics in War*. Oxford: Oxford University Press.
- Taurek, J. (1977). “Should the Numbers Count?”, *Philosophy and Public Affairs* 6: 293–316.
- Timmerman, T. (2015). “Sometimes There is Nothing Wrong With Letting a Child Drown”, *Analysis* 75(2): 204–212.
- Tomlin, P. (2017a). “On Limited Aggregation”, *Philosophy and Public Affairs* 45(3): 234–260.
- Tomlin, P. (2017b). “Innocence Lost: A Problem for Punishment as Duty”, *Law and Philosophy* 36: 225–254.
- Thomson, J. A. (2022). “Relief from Rescue”, *Philosophical Studies* 179: 1221–1239.
- Thomson, J. J. (1977). “Some Ruminations on Rights”, *Arizona Law Review* 19(1): 45–60.
- Thomson, J. J. (1990). *The Realm of Rights*. London: Harvard University Press.
- Thomson, J. J. (1991). “Self-Defense”, *Philosophy and Public Affairs* 20(4), 283–310.
- Tunick, M. (2011). “Entrapment and Retributive Theory”, in M. White (ed), *Retributivism: Essays on Theory and Policy* (Oxford: Oxford University Press): 171–191.
- Uniacke, S. (1994). *Permissible Killing: The Self-Defence Justification of Homicide*. Cambridge: Cambridge University Press
- Uniacke, S. (2011). “Proportionality and Self-Defense”, *Law and Philosophy* 30: 253–272.
- Venezia, L. (2022). “Locke on Conditional Threats”, *Southern Journal of Philosophy* 60: 696–713.
- Voorhoeve, A. (2014). “How Should We Aggregate Competing Claims?”, *Ethics* 125: 64–87.

- Wallace, R. J. (2010). "Hypocrisy, Moral Address, and the Equal Standing of Persons", *Philosophy and Public Affairs* 38(4): 307–341.
- Wasserman, D. (1987). "Justifying Self-Defense", *Philosophy and Public Affairs* 16(4): 356–378.
- Wenar, L. (2013). "The Nature of Claim-Rights", *Ethics* 123(2): 202–229.
- White, S. (2015). "The Problem of Self-Torture: What's Being Done?", *Philosophy and Phenomenological Research* 94 (3): 584–605.
- White, S. (2017a). "On the Moral Objection to Coercion", *Philosophy and Public Affairs* 45 (3): 199–231.
- White, S. (2017b). "Responsibility and the Demands of Morality", *Journal of Moral Philosophy* 14: 315–338.
- White, S. (2021). "Self-Prediction in Practical Reasoning: Its Role and Limits", *Noûs* 55: 825–841.
- Wolfendale, J. (2016). "Provocative Dress and Sexual Responsibility", *The Georgetown Journal of Gender and The Law* 17: 599–524.
- Woollard, F. (2015). *Doing and Allowing Harm*. Oxford: Oxford University Press.
- Yaffe, G. (2004). "Conditional Intent and Mens Rea", *Legal Theory* 10: 273–310.
- Yaffe, G. (2005). "'The Government Beguiled Me': The Entrapment Defense and The Problem of Private Entrapment", *Journal of Ethics and Social Philosophy* 1: 1–50.
- Yaffe, G. (2010). *Attempts: In the Philosophy of Action and the Criminal Law*. Oxford: Oxford University Press.