

A RIGHTS-BASED PERSPECTIVE ON PERMISSIBLE HARM

Susanne Burri

A thesis submitted to the Department of Philosophy, Logic and Scientific Method
of the London School of Economics and Political Science
for the degree of Doctor of Philosophy, 14 August 2014.
Revised and final draft submitted 14 February 2015.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work. The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without the prior written consent of the author. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 70,201 words.

Susanne Burri

Abstract

This thesis takes up a rights-based perspective to discuss a number of issues related to the problem of permissible harm. It appeals to a person's capacity to shape her life in accordance with her own ideas of the good to explain why (i) her death can be bad for her, and why (ii) each of us should have primary say over what may be done to her.

The thesis begins with an investigation of the badness of death for the person who dies. If death is bad for us, this helps explain the wrongness of killing. The thesis defends the deprivation account—i.e. the idea that death is bad for us when and because it deprives us of good life—against two Epicurean challenges. It adds that death is also bad when and because it thwarts our agency.

Next, the thesis deals with the logic of our moral rights to non-interference. It proposes a conception of rights according to which the stringency of our rights derives from and is justified by the rational aspect of our human nature. It argues that this conception of moral rights solves the paradox of deontology.

While our rights to non-interference are stringent, they are not absolute. The thesis considers two possible exceptions to the general rule that it is impermissible to harm an innocent person against her consent. First, using an actual case from WWII, it investigates the circumstances under which a government may expose some parts of its population to an increased risk of harm in order to decrease the risk to others. Second, it considers the permissibility of self-defence against an innocent threat. It argues that the potential victim of an innocent threat has a justice-based reason to treat her own interests as on a par with those of the threat.

Contents

	Page
List of Tables	7
Acknowledgements	8
1 Introduction	10
1.1 On the Ethical Significance of Our Rationality	12
1.2 Two Provisos on Permissible Harm	17
1.3 Understanding What is at Stake in a Decision Situation	22
2 How Death is Bad for the Person Who Dies	24
2.1 Introduction	24
2.2 Two Epicurean Arguments Against the Badness of Death	26
2.3 The Deprivation Account of the Badness of Death	28
2.3.1 Comparativism	29
2.3.2 The Deprivation Account and the First Epicurean Argument	32
2.3.3 The Deprivation Account and the Timing Puzzle	35
2.3.4 A Dialectical Stalemate	37
2.4 Death from a Deliberative Perspective	41
2.4.1 Choosing the Best Expected Life	41
2.4.2 Epicurean Maximax	43
2.4.3 Focusing on Intrinsic Value	46
2.5 An Agency-Based Account of the Badness of Death	48
2.5.1 A Problem For the Deprivation Account	48

2.5.2	Death as a Frustrator of Categorical Desires	49
2.5.3	How Death Is Bad For Us as Agents	52
2.5.4	Agential Reasons to Put Death Off Into the Future	56
2.5.5	Solving the Problem for the Deprivation Account	58
2.6	Conclusion	60
3	Moral Status and the Paradox of Deontology	63
3.1	Introduction	63
3.2	The Paradox of Deontology	66
3.3	The Agent-Focused Approach to Solving the Paradox of Deontology	67
3.4	The Inviolability Approach	70
3.5	Inviolability and Moral Importance	73
3.6	The Saveability Criticism of the Inviolability Approach	74
3.7	Moral Status, Moral Importance, and the Saveability Criticism	76
3.7.1	Moral Status and Moral Importance	76
3.7.2	Do Groups Have a Moral Status?	78
3.7.3	The Weights-Based View of Moral Reasons	80
3.7.4	From Weights to Trumps	85
3.8	A Moral Status That Fits Us	94
3.8.1	Inviolability and Our Capacity for Rational Agency	95
3.9	The Enforceability Criticism	97
3.10	Why the Enforceability Criticism Fails	101
3.11	Effective Rights to Non-Interference	106
3.12	Conclusion	110
4	The Redirection of Bombs During World War II	113
4.1	Introduction	113
4.2	The Historical Case	115
4.3	The Feasibility of Confusing the Enemy	120
4.4	The Historical Case in the Philosophical Literature	122
4.4.1	The Difference Between Doing Harm and Allowing it to Occur	123
4.4.2	The Doctrine of Double Effect	127

4.5	Showing Equal Concern	131
4.5.1	Did Redirection Show Equal Concern?	134
4.6	The Right to Life	138
4.6.1	Negative and Positive Aspects of the Right to Life	138
4.6.2	The Significance of the ‘No Involvement’ Baseline	142
4.7	The Defensibility of the Redirection Strategy	146
4.8	Conclusion	153
5	A Profiting Innocent Threat and You: Why It’s a Toss-Up	155
5.1	Introduction	155
5.2	Arguments in Favour of Permissibility	156
5.3	Arguments Against Permissibility	158
5.4	The Toss-Up Between a Profiting Innocent Threat and His Victim . .	160
5.4.1	The Misconception	160
5.4.2	The Differences Between a Threat and His Victim	161
5.4.3	Weighing Claims	168
5.4.4	The Distributive Perspective	170
5.5	Comparing Innocent Threat Cases	172
5.6	A Prerogative to Save One’s Own Life?	174
	Bibliography	180

List of Tables

	Page
1 The Enforceability Dimension According to Otsuka	99
2 Personal Sovereignty's Enforceability Dimension	104

Acknowledgements

As a supervisor, Alex Voorhoeve has proved incredibly fast at being extremely helpful. I have often received his constructive and careful feedback to a piece of written work not just within days of sending it to him, but within hours, allowing me to work on a particular problem in the most effective way possible. Our discussions greatly advanced my understanding of the material, and they consistently left me inspired. I appreciate how Alex never failed to treat me respectfully and fairly, and how he managed to be genuinely encouraging. I thank Alex for his invaluable support during four years of outstanding supervision.

I am also very grateful to Luc Bovens for acting as a second supervisor. He has provided insightful and interesting comments and was available for extremely useful discussions with respect to chapters 2 and 3 of this thesis. No matter how busy he was, Luc never turned me away when I knocked on his door with a favour to ask or an issue to discuss. I am convinced that the LSE Philosophy Department is such a great place largely because it is headed by such a compassionate and cheerful character.

I thank Ralph Oppenheimer for his generous financial support through the *LSE Department of Philosophy PhD Partnership Scholarship*. I am honoured by the keen interest that Ralph has taken in my work and in my progress, and in the confidence that he has put in me. Without his support, I would not have been able to pursue my passion for philosophy.

I thank my partner Thomas Seiler for his unwavering patience and understanding especially during the last few months of my PhD. Thomas kept me sane by convincing me to go swimming or running on a daily basis. He was always willing to engage in philosophical discussion and help me clarify my thoughts. His loving encouragement proved vital to the completion of this thesis.

I am profoundly grateful to my parents and my two sisters. Their unconditional and uncomplicated love and support sustained me throughout my PhD, and I have drawn many philosophical inspirations from the extended conversations we tend to carry on while hiking or over a glass of wine.

I owe a large debt of gratitude to Dieter Thomä. With his extraordinarily well-taught course “Der Liberalismus und seine Kritiker”, he inspired me to complement my undergraduate degree in economics with postgraduate studies in both economics and philosophy, which eventually led me to pursue a PhD in philosophy. Dieter has been an invaluable source of encouragement and support throughout my academic career.

Karin Ohashi, Francesca Pero, Esha Senchaudhuri, Chlump Chatkupt, Foad Dizadji-Bahmani, Constantin Bürgi, Ben Ferguson, Roberto Fumagalli, Sascha Gysel, Dominik Hangartner, Peter Helles, Tom Hinrichsen, Andreas Pähler, Mantas Radžvilas, Alan Robinson, Hamid Seyedsayamdost, and Francesco Spöring are some of the friends who have shared in the ups and downs of my life as a PhD student, thereby making that life meaningful, fun, and memorable. Special thanks go to my very good friend Katya Pirozhkova. Without Katya, I would not have stumbled upon Vladimir Nabokov’s translation of Eugene Onegin, which figures importantly in the introduction to this thesis.

1 | Introduction

“The trouble is, as so often in philosophy, it is hard to improve intelligibility while retaining the excitement.”

– Donald Davidson (1974, p. 5)

“In art as in science, there is no delight without the detail, and it is on details that I have tried to fix the reader’s attention. Let me repeat that unless these are thoroughly understood and remembered, all ‘general ideas’ (so easily acquired, so profitably resold) must necessarily remain but worn passports allowing their bearers short cuts from one area of ignorance to another.”

– Vladimir Nabokov (1975, p. 8)

In February of 2013, my friend Katya and I went to see a wonderful production of Tchaikovsky’s *Eugene Onegin* at the Royal Opera House. After the performance, we walked around Covent Garden, feeling inspired and deeply moved. Katya recounted how she had read Pushkin’s novel in the original Russian many years ago, and she told me that she still remembered it as one of the most beautiful books she had ever read. Back at home, I ordered Nabokov’s English translation of Pushkin’s novel. When it arrived in the mail a few days later, I unwrapped it with great anticipation, and immediately started reading it. Nabokov’s introduction was intriguing. He emphasized that he had taken no liberties in his translation; that his aim had been to express in English precisely what Pushkin had written in Russian. Nabokov admitted that some of the beauty of the original was lost in his literal translation, but he explained that preserving the beauty of Pushkin’s verses would have come at the

cost of distorting their meaning. I turned the page, and started with the main text. After twenty minutes or so, I gave up. Nabokov's translation was unbearably dull and awkward; it was more painful to work through than most philosophy texts I have encountered. I never bothered giving it a second try. At least in art, I believe that there is no understanding without emotional engagement. It is funny that Nabokov—literary genius and admirer of Pushkin—assumed there was a way of staying true to the original text without trying to preserve its poetic beauty.

When it comes to philosophy, I am willing to slowly work my way through a problem or a difficult text. Here, what Nabokov (1975, p. 8) says seems exactly right: unless we pay careful attention to the details, we end up taking “short cuts from one area of ignorance to another.” In philosophy, precision is thus more important than beauty, and emotional involvement, rather than promoting our understanding, can sometimes keep us from thinking clearly. Still, when it comes down to it, what makes thinking about philosophical problems worthwhile is the fact that they have an emotional hold on us. As philosophers, we grapple with problems we find exciting and important, and we grapple with them *because* we find them exciting and important. When we lose sight of this—as sometimes happens when we struggle to think clearly about some small aspect of a much larger issue—doing philosophy can start to feel as cumbersome and pointless as reading an ugly translation of *Eugene Onegin*.

How do we retain our excitement for a philosophical problem when we are working on clarifying some small aspect of it? Personally, I have noticed that I remain excited about a problem if I periodically take a step back, and contemplate what I am working on from a certain distance. Hence this introduction, which I use to take a step back from the chapters of my thesis, so as to contemplate them from a distance. In the following pages, I explain what intrigues me about the different questions that my thesis tackles, and I speculate about some of the wider implications of my main conclusions.

The four chapters of my thesis focus on four different questions. Nonetheless, as I shall now explain, they are unified by a number of shared methodological assumptions and recurring themes.

1.1 On the Ethical Significance of Our Rationality

Especially the first two chapters of this thesis are based on the methodological assumption that a human person is a certain kind of creature, and that we can use this idea to justify some of the philosophical claims that we make. I particularly heavily rely on the idea that a human person is a rational being who lives her life in a purposeful way.

In chapter 2, I investigate how death can be bad for the person who dies. The ancient Greek philosopher Epicurus (1926, p. 85) famously argued that there is no way in which a person's death can be bad for her, and that we should thus all "[become] accustomed to the belief that death is nothing to us." According to the currently most favoured position in the philosophical literature, Epicurus' claim rests on an important oversight, and a person's death is bad for her just in case and to the extent that she would have lived an overall better life had her death not occurred when it did in fact occur. The basic idea behind this *deprivation account of the badness of death* is simple and compelling. But it is not clear how well the deprivation account can deal with a number of Epicurean challenges. In chapter 2, I defend the insights of the deprivation account against two Epicurean challenges. I show how from a deliberative perspective, the deprivation account lays out an intelligible and practically relevant way in which death can be bad for the person who dies. The Epicurean resistance to calling death 'bad' unhelpfully obscures this fact.

When I first encountered the deprivation account of the badness of death, I felt that there was something amiss. I am much more confident in my judgement that it would be bad for me if I died tonight than I am in my judgement that I will eventually have lived a better life if I do not die tonight. Yet according to the deprivation account, these two seemingly distinct judgements come to one and the same judgement, so that one should feel similarly confident in both.

My hunch that there was something amiss with the deprivation account led me to come up with and defend an *agency-based account of the badness of death*. The agency-based account is not meant to replace the deprivation account, but to *complement its insights*. As I see it, the deprivation account accurately captures the conditions under

which a person has an all-things-considered reason to avoid her own death. But it fails to make clear that a person's death can be non-comparatively bad for her. I argue that this failure stems from the fact that defenders of the deprivation account see a human person primarily as a *bearer of experiences*, or as someone to whom good and bad things happen. But a human person is also an *agent*, or someone who lives her life guided by goals and principles that she has set for herself. Once we bear in mind that a human person is an agent, it becomes apparent that the deprivation account is importantly incomplete as a description of how a person's death can be bad for that person.

To the extent that we live our lives as agents, we live a life that fits our nature as rational beings. Robert Nozick (1974, pp. 49-50) suggests that the distinguishing feature of a rational being is its capacity to govern itself in the light of reasons. To Nozick, a rational being is "able to formulate long-term plans for its life, able to consider and decide on the basis of abstract considerations or principles it formulates to itself and hence not merely the plaything of immediate stimuli, a being that limits its own behavior in accordance with some principles or picture it has of what an appropriate life is for itself and others" (Nozick, 1974, p. 49). Insofar as we are rational beings, our lives are meaningful and valuable if we are able to live them in a self-determined way, i.e. by pursuing plans and projects in accordance with our own ideas of the good.

In chapter 3, I argue that an appeal to the rational aspect of our human nature helps justify and explain why as human persons, we have stringent moral rights to non-interference. Moral rules built around stringent rights to non-interference neither make it morally permissible to minimize the harm that comes to innocent persons, nor do they make it morally permissible to distribute harm equally among similarly undeserving recipients. But with stringent rights to non-interference, each of us enjoys an almost exclusive authority over the 'territory' of his or her own body and mind (while being in turn required to steer clear of the sovereign territories of other people). As rational beings with a strong social bent, this sovereignty is something we rightly value, and that therefore fits the kind of creature we are. Being sovereign over our bodies and minds enables us to pursue our plans and projects, and to shape our lives

as we please, while living together in peace.

How should our moral rules deal with the fact that the sovereignty of the individual is not always respected? Suppose a vicious murderer is threatening to kill five innocent people, thereby arrogating to himself a decision over their fate. If he manages to kill the five, he thwarts their plans and projects, and he destroys their capacity for rational agency. Now suppose that there is a way to stop the murderer, but it involves killing one innocent person as a means to secure the safety of the five. Is it morally permissible to kill the one as a means to secure the safety of the five? Intuitively, the answer is 'no'. But this intuitive judgement is difficult to justify. To see why, suppose that it is permissible to kill the one, and you go ahead and kill him. Then only one person has his agency destroyed and his plans and projects thwarted. From the point of view of what our rights to non-interference protect, this would seem to be a better outcome than the outcome in which five people have their agency destroyed and their plans and projects thwarted. This means that if our intuitive judgement is correct, then our rights to non-interference sometimes make it impermissible to bring about what is a best possible outcome *from their own point of view*. How can this be? In its general form, this problem is standardly referred to as the *paradox of deontology* (Nozick, 1974, pp. 28-30).

In chapter 3, I follow Frances Kamm (1996) in defending a status-based approach to answering the paradox of deontology. Kamm's basic idea is that the paradox of deontology rests on a mistaken understanding of the purpose of our rights to non-interference. To Kamm, we do not have rights to non-interference only because, or even primarily because, they help protect us against unwanted interference. Rather, the primary purpose of our rights to non-interference is to mark us out as beings who are *owed* respectful treatment, and who *rightfully* have the final say over what should happen to them. To put the same idea differently, the main point of our moral rights to non-interference is to protect and give expression to our *inviolability*. The more stringent our rights, the more inviolable we are. The more inviolable we are, the fewer the circumstance under which morality endorses our being harmed against our consent, and the higher our moral status. Kamm (1996, p. 272) argues that having a high moral status is good for us: it means that we are morally important beings who

are “worthy of respect, owed respect.” As a solution to the paradox of deontology, this suggests that it may be impermissible to kill one person as a means to save five others because this implies that we are morally highly important beings who are owed respectful treatment.

I argue that Kamm’s status-based approach to solving the paradox of deontology can be fleshed out into a viable solution to the paradox. I consider two criticisms to the status-based approach: a *saveability criticism* and an *enforceability criticism*. The saveability criticism (Kagan, 1991, pp. 919-20) accuses the status-based approach of arbitrarily favouring the importance captured by our inviolability over the importance captured by another aspect of our moral status—our ‘saveability’. A person’s saveability increases as more may permissibly be done to others to save her from harm. I show that the saveability criticism is well-founded, and that we should therefore not appeal to considerations of moral importance to defend the idea that being highly inviolable is good for us. I suggest that we should instead appeal to considerations of *fittingness*. On the fittingness defence of the status-based approach, we have reason to believe that each of us is highly inviolable just in case this fits the kind of moral creature we are. I propose that as rational beings, we are capable of shaping our life in accordance with our own ideas of the good, and we aspire to lead our lives in a self-determined way. This makes it fitting that each of us should have primary say over what may be done to him or her (Quinn, 1989b, p. 309). To the extent that our moral rules make us sovereign over our bodies and minds, they therefore recognize and give expression to our capacity for rational agency—they endow us with a moral status that fits the *rational* aspect of our human nature.

The enforceability criticism (Otsuka, 2011, pp. 51-4) grants that our moral status should fit the kind of creature we are. It also grants that as rational beings who aspire to lead self-determined lives, it is fitting that we should be highly inviolable, as being highly inviolable makes us personally sovereign by putting us in charge of our bodies and minds. But it contends that the fittingness defence of the status-based approach depends on too narrow an understanding of personal sovereignty. More specifically, it holds that an appropriate conception of personal sovereignty cannot be limited to rights to non-interference, but has to include enforcement rights also.

If this is correct, then our personal sovereignty might increase not only with the stringency of our rights to non-interference, but also with our moral authority to defend ourselves against illegitimate interference. Michael Otsuka (2011) claims that our authority to defend ourselves consists of and increases with enforcement rights we have against culpable attackers as well as against innocent third parties (whom it may sometimes be helpful to harm in order to fend off an attack). He then shows that on his conception of personal sovereignty, the fittingness defence of the status-based approach is undermined.

I maintain that the enforceability criticism fails. I argue that while Otsuka is right to insist that there is an enforceability dimension to personal sovereignty, he is wrong to believe that it extends to enforcement rights we have against innocent third parties. This restores the fittingness defence of the status-based approach to solving the paradox of deontology.

However we spell out its details, it is an intuitively plausible and promising idea that the rational aspect of our human nature helps justify and explain the *stringency* of our moral rights to non-interference. But what explains their *limits*? Why do we not have absolute rights to non-interference? In chapter 3, I argue that we care about promoting best possible outcomes for a number of reasons, and that these reasons need to be balanced against status-based considerations. For one thing, a capacity for rational agency is not the only morally relevant aspect of human nature. Another relevant aspect is sentience, or the fact that a human person is a “complex, feeling creature who cares about whether it lives or dies” (Kamm, 1996, p. 277). It is reasonable to assume that our sentience grounds a moral concern with what actually happens to us, thus underpinning consequentialist ideas according to which we should try and do what best promotes the general welfare. Insofar as these ideas are on the right track, we can understand conflicts between respecting people’s rights to non-interference and promoting the general welfare as clashes between the moral implications of our rationality and our sentience—the former giving rise to our moral status as sovereign individuals who must not be interfered with against their consent, the latter grounding a consequentialist concern with the general welfare.

This rough sketch helps clarify why I am intrigued by the idea that there is a

connection between our moral rules and certain basic features of our human nature. In addition to being intuitively plausible, it allows us to frame morally difficult decision situations as clashes between the moral implications of different aspects of our human nature. In this way, it enables us to understand why reasonable people sometimes disagree about the right thing to do. Unlike many other moral philosophers, I do not think that there is always one right ‘solution’ to a morally difficult decision situation, a solution that we are able to identify if only we give the matter due consideration. It seems to me that where the implications of different but fundamental values pull apart, there will sometimes be no unambiguously right thing to do—only differently unsatisfactory courses of action we can take. This does not mean that thinking about morally difficult decisions strikes me as a waste of time. Quite to the contrary, I believe that much is gained if we are able to improve our understanding of the nature of moral conflicts, and are able to identify what is at stake in any given difficult decision situation.

1.2 Two Provisos on Permissible Harm

Our moral rights to non-interference are stringent, but they are not absolute. Under certain circumstances, it becomes permissible to infringe them. In chapters 4 and 5, I look at two exceptions to the general rule that it is morally impermissible to harm an innocent human person against her consent.

Chapter 4 is a philosophical investigation of an actual historical case. During World War II, the German Wehrmacht started terror bombing London. The Germans were assumed to be targeting the area around Charing Cross in Central London, but the actual ‘mean point of impact’ (MPI) of their bombs lay in Dulwich, some four miles south-east of Charing Cross. With the help of double agents that the British authorities had recruited among German spies, it would have been possible to redirect the bombs even further away from Central London. Implementing this *redirection strategy* would likely have saved a great number of civilian lives in Central London, while resulting in an increase of civilian deaths in certain areas in South London. The expected net effect was a significant reduction in the overall number of civilian

deaths. While the British deception authorities were convinced that redirecting the bombs was the right thing to do, the War Cabinet's Ministers were doubtful. They felt that it was "not for the Government to authorise action which would result in certain areas in South London [...] sustaining heavier casualties than they do at the present time", and they therefore issued an order not to redirect the bombs, but to "confuse the enemy" (PREM 3/111A, 3 August 1944).

After taking a close look at the historical details of the case, I discuss whether the Ministers were right to feel hesitant about the redirection strategy. Were there any considerations that spoke against it? If yes, did these considerations make it the case that confusing the enemy would have been morally preferable to redirection?

Implementing the redirection strategy would have significantly reduced the expected overall number of civilian deaths. While redirecting the bombs would have caused some harm, it would thus also and expectably have prevented much greater harm. This means that the British authorities had a *lesser evil justification* for adopting the redirection strategy. Moreover, it is plausible to assume that redirecting the bombs need not have involved intending harm. I use some of the ideas behind Kamm's *Doctrine of Triple Effect* to argue that even though the redirection strategy *depended* on the occurrence of bomb incidents in certain parts of London, none of these incidents had to be *intended as a means* to implement the redirection strategy (Kamm, 2007, p. 95-120).

Non-intentional harm that comes with a lesser evil justification is the textbook example of harm that may permissibly be inflicted on innocent persons. But then why did the Ministers feel uneasy about redirection? I argue that they had good reason for doing so. A government owes its duty of protection equally to all of its citizens. This means that the protective measures which it implements should show a sufficiently equal regard for the safety of all of its citizens. But there is an important sense in which the redirection strategy did *not* do this. To adopt it meant to try and keep the people living in Central London safe by making those living in South London bear the brunt of the German attack. The fact that redirecting the bombs meant singling out a part of the population to bear the brunt of the German attack provided a *pro tanto* reason against adopting it. I argue that under the circumstances,

it would nevertheless have been right to redirect the German bombs, as avoiding to show unequal concern would have come at too high a price in human lives. But I also argue that it would *only* have been right to adopt the redirection strategy subject to the qualification that this would have been accompanied by a serious effort to put a swift end to the bombings. Why this qualification? I propose that it is an application of the following more general proviso:

Proviso on Redirection. If you redirect a threat away from more and towards fewer people, then you ought to subsequently expend the same effort and expenses to protect the smaller number as you ought to have expended to protect the larger number in the absence of redirection.

According to the proviso on redirection, what makes redirecting a threat justifiable is partly a commitment to do just as much to stop the threat as you would have done—and should have done—in the absence of redirection. If you do less to stop the threat once you have contained its potential impact through redirection, this makes it the case that you redirected the threat partly to save expenses and effort, and not just to save lives. But the consideration that it would save expenses is not a valid reason in favour of exposing innocent people to grave danger. If you put innocent people's lives at risk through redirection, it should be only because this serves to significantly reduce a comparable risk to more numerous others.

The redirection of bombs during WWII would have been morally permissible partly because it came with a lesser evil justification, and because redirecting the bombs need not have involved intending harm. By contrast, chapter 5 argues that there are circumstances under which it is permissible to harm an innocent person when both of these factors are absent. To get a feel for these kind of circumstances, imagine that you find yourself at the bottom of a deep well (Nozick, 1974, p. 34). Thugs have picked up an innocent person—call him Bob—and have thrown him down the well. Bob is now falling towards you. If you do nothing, your body will cushion Bob's otherwise lethal fall. This will guarantee his survival, but it will kill you. If you shoot your ray gun, you vaporize and kill Bob, thereby saving your life. Are you

morally permitted to shoot your ray gun?

If you are, then it is permissible to *intentionally* kill an innocent person in the *absence* of a lesser evil justification. What could possibly ground such a permissibility? Some moral philosophers appeal to the idea of an agent-relative prerogative to argue that you may permissibly kill Bob (see e.g. Quong, 2009). Their idea is that because Bob will kill you unless you kill him—and because you haven’t forfeited any of your rights to non-interference—you may permissibly save your own life by vaporizing Bob. To their mind, our morality would be overly demanding if you lacked a right to defend yourself against being crushed by Bob.

I argue against the idea of an agent-relative prerogative to kill an innocent threat such as Bob. Intentionally killing an innocent person merely because one puts one’s own interests first is a gravely disrespectful thing to do. Of course, it can be extremely difficult to keep a cool head when one’s own life is at stake, so that it may sometimes be appropriate to *excuse* someone who has killed an innocent other in order to secure their own survival. But this in no way suggests that it would be *morally justifiable* to shoot Bob. In light of the fact that our moral rules are based on a strong presumption against harming innocent others, it is difficult to see how you could be permitted to kill Bob merely because it would be good for you if you were.

Based on these considerations, I reject a prerogative-based justification for harming Bob. But I still think that you may permissibly harm Bob. More precisely, I argue that you have a *justice-based reason* to inflict on Bob the harm that he would be duty-bound to suffer for your sake. To see what I mean, suppose that Bob could redirect himself away from you at the cost of a broken leg (Kamm, 1996, p. 96). It then seems that Bob would be under a duty to do so, even though an innocent bystander would not have to break one of her legs to keep Bob from harming you (ibid.). Why would Bob have to incur substantial harm to avoid harming you if he could? In essence, because our morality is built on a strong presumption against harming innocent others. Metaphorically speaking, if Bob were physically able to avoid trespassing on your sovereign territory, then he would be morally required to take great pains to do so.

Yet as things are, Bob is unable to avoid harming you. You, on the other hand,

are able to avoid harming Bob. Given our morality's strong presumption against harming innocent others, isn't it simply your tough luck that you have to let yourself be crushed by Bob? I argue that it is not. In brief, the protective logic of our moral rights to non-interference is built around—and for its functioning depends on—the idea of mutual recognition between moral equals. This means that you owe it to Bob not to just vaporize him because he is the kind of creature who can understand that if he could avoid crushing you, then he would have to go to great lengths to do so. But if you must constrain yourself towards Bob partly *because* Bob would have to suffer substantial harm to avoid crushing you if he could, it would be unjust if you could not impose that harm on Bob on his behalf.

How much harm would Bob have to impose on himself to avoid harming you if he could? I argue that if Bob were falling towards you with a ray gun at his disposal, he would have to treat your interests as on a par with his own. This means that he would have to flip a coin to choose between vaporizing himself and saving his life by crushing and killing you. It follows that in the original set-up, you may burden Bob with a 50% chance of death, as this is the amount of harm that he would have to inflict on himself for your sake.

I contend that it is impermissible to burden Bob with more harm than a 50% chance of death. Given the circumstances, there is a certain amount of harm that may rightly be assigned to Bob, i.e. the amount of harm that he would have to suffer to avoid harming you if he could. Once you impose on Bob harm in excess of this amount, you violate his rights to non-interference. These considerations suggest a second proviso on permissible harm. It can be stated as follows:

Constraint on Agent-Relative Prerogatives. If you harm an innocent person only so as to avoid harm to yourself, you must not impose on the innocent person harm in excess of the amount that he would have a duty to inflict on himself if he were able to.

As a constraint on agent-relative prerogatives, this proviso states that even if it would greatly benefit you, you must not do to Bob what he would not have to do to himself if he were able to.

1.3 On the Importance of Understanding What is at Stake in a Decision Situation

Normative ethical theorizing should provide us with practically relevant insights: it should have the potential to help us determine what we ought to do in a given decision situation (see e.g. Rachels, 2000, p. 89, Feldman, 2006, p. 49). But when we stop to think about it, it may appear doubtful whether I have managed to clarify *in a practically relevant way* what was at stake in the British decision to redirect the German bombs.

As mentioned in section 1.2, I argue that if the British authorities had adopted the redirection strategy, they would not have shown equal concern for the safety of all of their citizens. I suggest that this consideration provided a *pro tanto* reason against adopting the redirection strategy. But then I go on to conclude that *all things considered*, the German bombs *should* have been redirected, as avoiding to show unequal concern would have come at too high a price in human lives. From a practical point of view, this discussion would seem useful only if it specified how the trade-off between showing equal concern and minimizing losses ought to be decided. But I make no suggestions to this effect. Practically speaking, what is the point of arguing for the *importance* of a consideration without specifying the conditions under which it is *decisive*?

I agree that the arguments presented in chapter 4 would gain in practical relevance if they were more specific about the proper trade-off between showing equal concern and other considerations. But this does not make them useless as they stand. If you know what considerations are important in a given decision situation, this allows you to try and find an alternative that accommodates them all. To see what I mean, suppose that you are one of the War Cabinet's Ministers. You have to decide whether to support or oppose the redirection strategy. You know that redirection is a protective measure that does *not* show equal concern for the safety of all, and you understand that this gives you a reason to avoid it. Even if you do not know *precisely how important* it is to show equal concern, the awareness *that* it is important suffices

to guide deliberation: it tells you to look for alternatives to redirection that better express an equal concern for the safety of all citizens, and to give these alternatives serious consideration. If you raise the issue with your fellow Ministers, and if you and your fellow Ministers are a dedicated and resourceful bunch, you may be able to come up with an ingenious alternative to redirection that avoids showing unequal concern, but that nevertheless promises to be highly effective as a protective measure. You might also find that given the circumstances, *all* of the available alternatives for action—including doing nothing—would amount to showing *similarly* unequal concern. If so, you would be justified in implementing the most effective protective measure. It is of course possible that no such clearly best alternative can be found. When this is the case, knowing what is of importance may help guide you in the right direction, but it will not lead you all the way to a satisfactory practical conclusion. As mentioned further above, I believe that there can also be circumstances under which no entirely satisfactory practical conclusion exists.

When Nabokov decided to translate *Eugene Onegin*, he failed to properly take into account considerations of style and beauty. Instead of being painfully aware of the difficult trade-off between preserving the beauty of Pushkin's verses and staying close to their literal meaning, Nabokov resolutely set considerations of style aside, and concerned himself only with literal meaning. The result proved disastrous. Had Nabokov felt the full force of both considerations, he probably would have struggled to juggle them, and there is no guarantee that he would have managed to settle on a suitable compromise between them. But in light of his literary genius, he just *might* have come up with a wonderful translation well worth reading.

2 | How Death is Bad for the Person Who Dies^{*}

“When you are dead, you don’t know that you are dead. It is difficult only for other people.

It is the same when you are stupid.”

– Unknown

2.1 Introduction

It is uncontroversial that a person’s death can be a terrible misfortune for other people. To lose someone you love is often an immensely painful and disruptive experience. But can death also be bad for the person who dies?¹ Most people intuitively think that it can. The Greek philosopher Epicurus (1926, p. 85), however, has argued against this intuitive judgement, and has put forward the idea that we should all become “accustomed to the belief that [our own] death is nothing to us.” Today, a number of *Epicurean arguments* based on Epicurus’ insights are still used to challenge the wide-spread belief that death can be bad for the person who dies.

In the philosophical literature, the currently most favoured explanation of the badness of death is offered by the *deprivation account*. According to the deprivation

^{*} An earlier version of this chapter was presented at the Stockholm Philosophy Graduate Conference. I thank Jens Johansson for a constructive response and several helpful discussions. I am grateful to Katie Steele for providing useful feedback on the decision-theoretic aspects of this chapter. I also thank Ben Bradley and Steven Luper for helpful comments on earlier versions of this chapter.

¹ In what follows, I put aside the bad effects someone’s death can have on other people, and I use terms such as ‘the badness of death’ as shorthand for ‘the potential badness of death for the person whose death it is’.

account, a person's death is bad for her just in case—and to the extent that—she would have led a better life had she not died when she did in fact die. The deprivation account is more congenial to our intuitions than the Epicurean position. But it is not clear whether defenders of the deprivation account are able to provide satisfactory answers to the challenges raised by the Epicurean arguments against the badness of death.

In this chapter, I defend the deprivation account against two Epicurean challenges. I argue that the deprivation account lays out an intelligible way in which our own death can be bad for us, and that the Epicurean resistance to calling death 'bad' unhelpfully obscures this fact. I also put forward an *agency-based account of the badness of death* which complements the insights of the deprivation account.

The structure of this chapter is as follows. In section 2.2, I present two Epicurean arguments against the badness of death. According to a first argument, death cannot be bad for the person who dies because it lacks the kind of features that could make it either good or bad for her. According to a second argument, death cannot be bad for the person who dies because there is no time at which she could possibly suffer its supposed misfortune. In section 2.3, I introduce the deprivation account, and I discuss how we ought to resolve the stand-off between an Epicurean and a defender of the deprivation account. I suggest that Epicureans and defenders of the deprivation account talk past each other because they do not mean the same thing when they talk about harm. I also suggest that we do not have to resolve their fundamental disagreement before we can make philosophical progress. As long as there are practical implications which make it clear *how* our death can be bad for us, we need not establish whether our death can *harm* us.

In section 2.4, I argue that from a deliberative perspective, the deprivation account shows that we can have *reason to avoid our own death*, i.e. reason to put its occurrence off into the future at some cost to ourselves. In addition to this, I show that an Epicurean ought not to resist the idea that we can have reason to avoid our own death. But if the Epicurean should accept that there are valid reasons to avoid death, then he should also accept that there is at least one clear sense in which death can be bad for the person who dies. In section 2.5, I argue that a person's death can be

bad for her not only when, and not only because, it deprives her of a better life. I put forward an agency-based account of the badness of death according to which our death is bad for us *as agents* when and because it frustrates some of our plans and projects. I show that despite appearances to the contrary, the deprivation account is unable to adequately capture this ‘agential’ aspect of the badness of death. Section 2.6 concludes.

2.2 Two Epicurean Arguments Against the Badness of Death

In his *Letter to Menoecus*, Epicurus advises his friend that death is “nothing” (Epicurus, 1926, p. 85) or at least “no concern” (Epicurus, 2006, sec. 125) to the person who dies:

“Become accustomed to the belief that death is nothing to us. For all good and evil consists in sensation, but death is deprivation of sensation. [...] So death, the most terrifying of ills, is nothing to us, since so long as we exist, death is not with us; but when death comes, then we do not exist. It does not then concern either the living or the dead, since for the former it is not, and the latter are no more” (Epicurus, 1926, p. 85).

This passage provides material for two Epicurean arguments² in favour of the view that death cannot be bad for the person who dies. According to a first argument, a person’s death cannot be bad for her because it lacks the kind of features that could make it either good or bad for her. This argument rests on two premises: the termination thesis³ and hedonism. The *termination thesis* asserts that a person ceases to exist when she dies. *Hedonism* is a theory of value according to which pleasure is the only intrinsic good, and pain the only intrinsic bad. If something is *intrinsically*

² I refer to the two arguments as ‘Epicurean’ because they are based on the passage just quoted, but are not intended as a reconstruction of Epicurus’ views. Rather, they are meant to capture two challenges to the view that a person’s death can be bad for her that are frequently discussed in the philosophical literature on the topic.

³ This is Feldman’s (1992, p. 89) terminology.

valuable—so that it is either intrinsically good or intrinsically bad—it is good or bad because of the way it is in itself (Olson, 2013, p. 67); its goodness or badness “is not derivative from or due to the [value] of anything beyond itself” (Luper, 2013b, p. 318). Whenever something is good or bad without being intrinsically valuable, it is thereby *extrinsically valuable* (Luper, 2013a, p. 100). For a hedonist, ‘catching a cold’ is extrinsically bad because it causes us to feel pain (if only a few days later); ‘taking an aspirin’ when we are in pain is extrinsically good because it reduces the amount of suffering that we have to go through.

If the termination thesis is correct, a person’s death marks the end of her conscious experience, so that neither the event of her death nor her being dead can cause her to feel pain or pleasure.⁴ By hedonism, this would seem to imply that neither the event of her death nor the state of her being dead can be good or bad for her. But if this is right, then a person’s death cannot be bad for her. In sum, the first Epicurean argument challenges those who believe that death can be bad for the person who dies to explain *how* a person’s death can affect her adversely.

According to the second Epicurean argument, death cannot be bad for the person who dies because there is no point in time during which she could possibly suffer its supposed misfortune. As Epicurus puts it, “[...] when we exist, death is not present, and when death is present, we do not exist” (2006, sec. 125). Like the first Epicurean argument, this second argument depends on the termination thesis. Only if we cease to exist when we die is it true that “when death is present, we do not exist.” In addition to the termination thesis, the second Epicurean argument depends on a *temporal goods assumption*, according to which an event *E* is bad (good) for a person only if there is a time during which *E* affects her adversely (beneficially) (Luper, 2009b, p. 124; see also Broome, 2004, p. 237). While the second Epicurean argument is compatible with hedonism, it does not require it, as it puts no constraints on what it means for a person to ‘be affected adversely (beneficially)’, other than that

⁴ This is not to deny that the process leading up to a person’s death can be painful (or pleasurable) for her. In the context of this chapter, I do not mean to include a person’s dying process when I speak of her death. As Kai Draper (2013a, p. 298) puts it, “‘death’ is to be understood not as the dying process which [...] can involve prolonged and intense suffering [...]. Rather ‘death’ is to be understood either as the event of annihilation (i.e., ceasing to exist) that follows the dying process, or as ‘being dead’, the perpetual posthumous nonexistence that begins with annihilation.”

‘being affected adversely (beneficially)’ is something we are, or that happens to us, at particular points in time. If—contra the Epicurean—we believe that our own death can be bad for us, the second Epicurean argument challenges us to explain *when* a person’s death can affect her adversely. In light of this, I will refer to the second Epicurean argument as the *timing puzzle*.⁵

2.3 The Deprivation Account

In his seminal article *Death* (1974), Thomas Nagel argues that death is usually bad for the person who dies. At the core of Nagel’s argument is the idea that if we did not die when we do in fact die, we would (at least for a while) continue to live, and that any given death thus deprives the person who dies of life that she would otherwise have led. Nagel then argues that a given death is bad for the person who dies whenever—and to the extent that—the additional life that she would have enjoyed by dying a later death would have been good on balance. Because he assumes that life is usually good, he concludes that death is usually bad. By arguing in this way, Nagel was the first philosopher to present a *deprivation account of the badness of death*.

The deprivation account is built around the idea that death can be *extrinsically bad* for the person who dies. Similar to the way in which ‘taking an aspirin’ is extrinsically good for us to the extent that it reduces the amount of pain that we have to suffer, our death is held to be extrinsically bad for us to the extent that it reduces the amount of ‘good life’ that we are able to live through. If a version of the deprivation account is correct, then the Epicurean arguments presented in section 2.2 are not sound. But does the deprivation account stand up to scrutiny?

Given the way in which Nagel presents his ideas, the deprivation account seems to involve a potentially incoherent comparison. If death is bad just in case it would be better to go on living, then it seems that in order to assess the badness of death, we need to compare how good it is to be dead with how good it would be to remain alive. But the thought that there is some ‘goodness level’ that can be assigned to the state of being dead—for example a goodness level of zero, so that being dead is ‘intrinsically

⁵ This is Luper’s (2009a, sec. 4) terminology. Silverstein (1980, p. 405) and Bradley (2004, p. 6) refer to the timing puzzle as the “*no subject difficulty*” and the “*no subject problem*”, respectively.

neutral’—begs the question against the Epicurean position, which plausibly asserts that the state of being dead lacks intrinsic value (Hanser, 2008, p. 440). On the assumption that ‘being dead’ lacks intrinsic value, the intrinsic value of ‘being dead’ simply *does not exist*, and we cannot thus compare it to something else (Silverstein, 1980, pp. 404-5).⁶

2.3.1 Comparativism

Fred Feldman (1992), John Broome (1993, 2004), and Ben Bradley (2004, 2009) have built on Nagel’s ideas to develop more refined versions of the deprivation account. They make it clear that the deprivation account does not have to rely on a potentially incoherent comparison between the state of being dead and that of continuing to live, but can instead be made to rest on a conceptually straightforward *comparison between two lives*. According to their versions of the deprivation account, a specific death is *extrinsically bad* for the person who dies if and only if the life that is terminated by that death is *intrinsically worse* than the life the person would have led had she not died when she did in fact die. Analogously, a given death is *extrinsically good* for the person who dies if and only if the life that is terminated by that death is *intrinsically better* than the life the person would otherwise have led.⁷

Feldman, Broome, and Bradley argue not only that our death can be extrinsically

⁶ Ben Bradley (2004, 2009) argues that we should assign a well-being level of zero to the state of being dead. I briefly discuss Bradley’s ideas in section 2.3.3.

⁷ There are two interesting and related questions here, one metaphysical, the other epistemic. The metaphysical question asks about the right counterfactual: how should we settle the meaning of the expression ‘the life that someone would have led had they not died when they did in fact die’? The epistemic question asks whether—and if so, how—we can know the intrinsic goodness of the life that someone would have led had they not died when they did in fact die. Broome, Feldman, and Nagel largely ignore these questions, presumably because they see them as technicalities that can be left for others to solve. Jeff McMahan is a notable exception: when he develops his *time-relative interest account* (which is a complex deprivation account) he grapples with both questions at considerable length (see McMahan, 1988, pp. 41-54, and McMahan, 2002, pp. 103-35). Bradley (2009, pp. 47-60) carefully discusses some of the issues that McMahan raises. Both McMahan (1988, p. 45; 2002, pp. 113-4) and Bradley (2009, pp. 48-9) suggest that an appeal to *possible worlds* can help us get a better grip on the questions that the deprivation account raises. Based on a framework developed by David Lewis and Robert Stalnaker, the *life that a person would have led if she had not died when she did in fact die* can be understood as the life that this person leads in the possible world in which the suitably described particular death that ends her life in the actual world does not occur, but which is otherwise as ‘close as possible’, or most similar, to our actual world (Bradley, 2007, p. 112).

bad for us; they also claim that whenever a person's death is extrinsically bad for her, it is at the same time and to the same extent *overall bad* for her, or bad for her *all things considered* (Luper, 2013a, p. 100). Feldman (1992, p. 138 & p. 151) defines the overall badness (goodness) of an event E for a person P as follows:

An event E is [overall]⁸ bad for a person P whenever P 's life would have contained more intrinsic goodness had E not happened. E is [overall] good for P whenever P 's life would have contained less intrinsic goodness had E not happened.

If something is overall bad for us, then it is intrinsically or extrinsically bad for us (or both). But not everything that is intrinsically or extrinsically bad for us is overall bad for us. On the assumption of hedonism, it is intrinsically bad to undergo a painful treatment at the dentist, but if undergoing the treatment minimizes your tooth-related pain in the long run, it is overall good to undergo it. Catching a cold is extrinsically bad, but if it causes you to stay at home, enter a prize draw out of boredom, win £100,000, and embark on your dream vacation, its extrinsic badness is outweighed by its extrinsic goodness, so that catching the cold was only *partially bad*⁹ for you—overall, it turned out to be very good indeed (Luper, 2009b, p. 83).¹⁰

Comparativism is a popular account of what it means for something to benefit or harm us. Feldman, Broome, and Bradley, as well as most other defenders of the deprivation account, subscribe to it. According to comparativism, an event is against our interests and its occurrence harms us if and only if its occurrence is overall bad

⁸ Feldman uses 'comparatively' instead of 'overall'. On my understanding of the terms, something is *overall* bad for you if and only if it is bad for you all things considered, so that your life would have gone better had it not occurred. Something is *comparatively* bad for you if and only if it is worse for you than some relevant alternative. On the deprivation account, when your death is bad for you, it is both bad for you all things considered, and bad for you compared to the alternative of continued life.

⁹ Bradley (2013) uses *pro tanto* good or bad in place of Luper's *partially* good or bad.

¹⁰ In this chapter, I assume that pains and pleasures are all measurable on a common cardinal scale. One implication of this is that pleasures can offset pains, and vice versa. I also assume that pleasure and pain are aggregative goods, so that other things equal, more pleasure is better than less pleasure, and more pain is worse than less pain. It follows from these two assumptions that a hedonist judges some life A to be better (or to contain more intrinsic goodness) than some other life B just in case A contains a higher amount of net pleasure than B . The textual evidence suggests that Epicurus defended a very idiosyncratic form of hedonism that is incompatible with these assumptions (see e.g. Warren, 2004, pp. 127-35; Hetherington, 2013, esp. pp. 84-5). See also footnotes 2 and 27.

for us, so that our lives would have gone better had it not occurred (Luper, 2013b, p. 319). An event is in our interest and its occurrence benefits us if and only if its occurrence is overall good for us, so that our lives would have gone worse had it not occurred (ibid.). On the deprivation account, whenever a person's death is extrinsically bad for her, it is at the same time and to the same extent overall bad for her. Combined with a comparativist understanding of harm, this implies that whenever a person's death is bad for her, it harms her and is against her interests.

The deprivation account and the comparativist understanding of harm that underlies it are compatible with a wide variety of theories of value, including hedonism. As long as a theory of value permits us to compare whole lives in terms of their intrinsic goodness, the deprivation account will at least in principle allow us to assess whether a given death was overall good or bad for the person who died under that theory of value.

A theory of value which allows us to compare whole lives in terms of their intrinsic value may merely allow us to rank different lives in order, i.e. from best to worst. If so, it measures the values of different lives on an ordinal scale. Alternatively, it may in addition to this allow us to evaluate *how much* better or worse a given life was compared to another. If so, it measures the values of different lives on a cardinal scale, and we can meaningfully interpret the differences in the 'goodness numbers' that we assign to different lives. In this chapter, I assume that our theories of value allow us to measure goodness cardinally. At least in principle, it is then possible to determine just *how* good or bad a given death was. More precisely, the extent of the goodness or badness of a given death will be determined by the size of the difference between the intrinsic goodness of the counterfactual life that a person would have led had she not died when she did in fact die and the intrinsic goodness of her actual life. A large difference corresponds to a bad death. As the difference becomes smaller and turns negative, death becomes less bad and turns from a harm into a benefit.

Many of us find it plausible to assume that continuing to live is usually a good thing, and that longer lives tend to be more intrinsically valuable than shorter lives (see e.g. Nagel, 1974, p. 80; Broome, 2004, p. 237; McMahan, 2002, pp. 103-4). If this is right, then death will usually be bad for the person who dies, and death at an

earlier age will tend to be worse than death at a later age.¹¹

2.3.2 The Deprivation Account and the First Epicurean Argument

The first Epicurean argument challenges us to explain *how* death can be bad for the person who dies. If pleasure is the only intrinsic good, and pain the only intrinsic bad, and if death can be characterized as the absence of both, how can death possibly be good or bad?

The deprivation account gives a clear answer. A person's death is overall and extrinsically bad for her just in case (and to the extent that) she would have lived an intrinsically more valuable life had it not occurred. For a hedonist, a life's intrinsic value is determined by its *balance of pleasure over pain*. The higher that balance, the more intrinsically valuable the life (see footnote 10). On the assumption of hedonism, a person's death is thus bad for her just in case she would have lived a life with a higher balance of pleasure over pain had it not occurred.

This seems to refute the first Epicurean argument. But an Epicurean might beg to differ. Consider the following counterargument: "When we ask whether a person's death can be bad for her, we are trying to establish whether death can *genuinely harm* the person who dies, or be a *true misfortune* for her. This is what we are really interested in. Contrary to what comparativism asserts, being overall bad is *not* sufficient for some X to genuinely harm us, as not everything that is overall bad for us harms us. This is not to deny that 'being overall bad' may be a necessary condition for some X to constitute a harm. But what is *definitely* a necessary condition for genuine harm is that either X or its consequences (or both) be intrinsically bad. Death is

¹¹ This holds true only under the assumption that someone would live longer if they did not die when they do in fact die. But this need not necessarily be the case (see my comments in footnote 7 about the metaphysical question that the deprivation account raises). It is possible that the closest possible world in which an event *E* does not occur would *not* be the same as our actual world *A* up until *E* occurs in *A*. The idea that our world might have been different at points in time before *E* occurred had *E* not occurred is captured by the concept of a "backtracking counterfactual" (Bradley, 2004, p. 11). Advocates of the deprivation account often regard the assumption that someone would have lived longer had they not died when they did in fact die as self-evidently true, thus implicitly denying or at the very least disregarding the possibility of backtracking counterfactuals (cf. Nagel, 1974; Feldman, 1992, p. 139; Broome, 2004, p. 238). In what follows, I put aside the possibility of backtracking counterfactuals.

not intrinsically bad, and it has no intrinsically bad consequences. Hence our death cannot be bad for us *in the relevant sense*—it cannot genuinely harm us.”

Kai Draper (2013b, p. 72) refers to something that is intrinsically bad, or that has intrinsically bad consequences (or both) as “bad in the absolute (i.e. non-comparative) sense.” To Draper (2004, 2013a, 2013b), it is likely—and it makes for the most charitable interpretation of Epicurus’ views—that Epicurus took *absolute badness* to be the relevant notion for his purposes.

How appealing is the Epicurean counterargument just presented? Suppose we grant for the time being that we are interested in establishing whether death can genuinely harm the person who dies. Why shouldn’t the fact that X is overall bad for a person be sufficient to establish that X harms that person? Draper (2013a, p. 299) argues as follows:

“It is [overall] bad for me that I am not omnipotent, that I am not universally loved, and that I do not receive a relaxing massage whenever I would like one. I do not, however, count any of those deprivations among my misfortunes, nor do I regard them as harms, nor do I find anything troubling in their nature. Or [...] suppose that I am receiving a massage, a wonderful massage from Bjorn at Bjorn and Sven’s House of Swedish Massage. Suppose further that were I not enjoying that massage, I would be enjoying an even better massage from Bjorn’s even more talented partner, Sven. Then Bjorn’s giving me a wonderful massage is [overall] bad [...]. Nevertheless, I should scarcely be entitled to count my massage among my misfortunes, or to accuse poor Bjorn of having harmed me.”

Levenbook (2013, pp. 188-9, note 2) agrees with Draper that comparativism is inadequate as it stands, or that “unless some other necessary condition is added [to the comparativist definition of harm], there will be no way to distinguish between failure to benefit and harm.”

What should we make of this criticism of comparativism? I agree that it is highly

counterintuitive that Bjorn’s wonderful massage should constitute a harm.¹² Also, if our death can be bad for us *merely* in the way in which a wonderful massage can be bad, this would seem to take most of the sting out of death’s badness. But it does not follow from these two considerations that we should add ‘being absolutely bad’ as a necessary condition for ‘being a harm’. There might be events that we clearly consider harmful but that are not absolutely bad.¹³ Also, once we start to doubt comparativism, we may start to doubt that ‘being overall bad’ is a necessary condition for ‘being a harm’. Luckily, it is not the purpose of this chapter to find an adequate definition of harm. For our purposes, the important points to note are the following:

1. If ‘being absolutely bad’ is a necessary condition for ‘being a harm’, then the deprivation account does not establish that death can harm the person who dies.
2. It is not unreasonable to think that ‘being absolutely bad’ is a necessary condition for ‘being a harm’.

The idea that only absolutely bad things can harm us—the *absolute badness thesis*—is similar to the timing puzzle, but the two should be kept apart. Both can be understood as resting on the assumption that a person’s death is *relevantly* bad for her only if it *harms* her, and both can be understood to give a necessary condition for harm. Moreover, read as challenges to the deprivation account, both can be understood to give a necessary condition for harm which the mortal ‘harm’ identified by the deprivation account is then claimed not to meet. But at least on the face of it, the timing puzzle and the absolute badness thesis give different necessary conditions for harm. The timing puzzle asserts that X harms you only if there are *identifiable*

¹² Levenbook (2013, p. 189) points out that “[o]ne of Webster’s definitions for ‘harm’ is physical or mental damage.”

¹³ I do not list any contenders because I cannot think of any. It seems to me that *intuitively clear-cut examples of harmful events* are in fact examples of *absolutely* bad events (cf. also footnote 12). On the assumption of hedonism, ‘falling into a painless coma instead of going on a dream vacation’ at first sight seems to come close. But if the incident *never* causes you any emotional distress—e.g. because when you wake up from the coma, you no longer remember your vacation plans, and your friends and family take great care not to remind you—it is not intuitively obvious to me that ‘falling into a coma’ harmed you.

points in time at which X affects you adversely. By contrast, the absolute badness thesis makes no mention of time. Of course, if it is assumed that X affects you adversely *just in case* X is intrinsically bad for you or has intrinsically bad effects on you, then the timing puzzle becomes a more demanding version of the absolute badness thesis. It then asserts that X can harm you only if X is absolutely bad for you *and* this badness is locatable in time. On a theory of value where intrinsic badness for a person is always and straightforwardly locatable in time—such as hedonism—the timing puzzle and the absolute badness thesis will then coincide. But if it is possible for some X to affect you adversely *without* being absolutely bad for you, the timing puzzle and the absolute badness thesis remain distinct.

2.3.3 The Deprivation Account and the Timing Puzzle

Can defenders of the deprivation account provide a satisfactory solution to the timing puzzle? Suppose we grant that my actual death will harm me because it will deprive me of a better life that I would have led had I not died when I will in fact die. If so, when does my death affect me adversely? Nagel (1970, p. 78) thinks that I suffer the misfortune of my death *atemporally*:

“[When] is a subject’s actual life, taken as a whole, worse than the life he would have lived had he not died his actual death? The answer would seem to be that it is timelessly worse. On [Nagel’s] analysis, the harm of death cannot be located temporally at all.”
(Hanser, 2008, p. 448; cf. also Bradley, 2004, p. 2)

This position begs the question against the *temporal goods assumption*, according to which an event X is bad for a person only if there is a time at which it affects her adversely. Hanser (2008) finds this unacceptable. He argues that we should reject the deprivation account because it fails to provide an adequate solution to the timing puzzle.

Defenders of the deprivation account disagree. They think that the timing puzzle can be dealt with in one way or another. Some back Nagel’s atemporal view (e.g. Broome, 2004); others argue that we suffer the misfortune of our death eternally (e.g.

Feldman, 1991); or while we are alive (e.g. Pitcher, 1984; Luper, 2004); or when our death occurs (e.g. Lamont, 1998); or for a time after our death (e.g. Bradley, 2004; 2009). In this section, I consider as examples Bradley's and Broome's proposed answers to the timing puzzle.¹⁴ While each of them holds considerable intuitive appeal, I argue that they both fall short of providing a compelling solution to the timing puzzle.

Bradley (2004, 2009, ch. 3) suggests a version of the deprivation account according to which we suffer the misfortune of our death during a specific period of time *after our death*, roughly for as long as we would have continued to live a good life had we not died when we did in fact die. Granting hedonism, he argues that we should assign a well-being level of zero to the state of being dead, and that a person is affected adversely by her death at all those times after her death at which her welfare level would have been positive had she not died when she did in fact die. Analogously, a person is affected beneficially by her death at all those times after her death at which her welfare level would have been negative had she not died when she did in fact die. This neatly yields the comparativist conclusion that the overall harm which our death causes us is the total net welfare that it deprives us of.

Bradley's solution to the timing puzzle is elegant, but it argues against intuitively highly plausible Epicurean assumptions, so that an Epicurean is unlikely to be moved its elegance. By assigning a welfare level of zero to the state of being dead, and by claiming that we suffer the harm of our death once we no longer exist, Bradley refuses to let himself be constrained by the Epicurean presupposition that 'being dead' lacks intrinsic value.

Broome defends Nagel's idea that if a person's death is bad for her, it affects her adversely in an *atemporal* way. Broome (2004, p. 237) concedes to the Epicurean that many things that are good or bad for us are 'temporal goods', in that they are good

¹⁴ Bradley's and Broome's answers are both consistent with hedonism. Pitcher's and Luper's proposals are not considered here because they are inconsistent with hedonism, and would thus be unacceptable to an Epicurean. Feldman's proposal is not considered because it seems to me to conflate the time at which it is *true* that X caused me harm with the time at which X actually affects me adversely (Luper, 2009a, sec. 4.2; cf. Bradley, 2004, pp. 7-8). Lamont's proposal is not considered because it seems to me to conflate the time when a harm is *caused* with the time at which it is suffered (see Broome, 2013, p. 220).

or bad for us only because there is a time during which we enjoy their benefits or are in some way worse off because of them. What Broome (ibid.) denies is that all things that are good or bad for us are temporal in this way. To Broome, there is at least one “obvious way for an event to harm you without harming you at any time: it only has to shorten your life” (ibid.). Based on this, Broome (2004, p. 238) recommends we reject the Epicurean temporal goods assumption, and with it the challenge of the timing puzzle.

An Epicurean would not feel the force of Broome’s argument. To Broome, it is “obvious” that an event which shortens your life can harm you without affecting you adversely at any time because he accepts a comparativist account of harm. But from an Epicurean point of view, comparativist accounts of harm appear deficient. It follows that what seems “obvious” from Broome’s point of view requires careful elaboration from an Epicurean point of view.

2.3.4 A Dialectical Stalemate

Epicureans claim that our own death cannot harm us. They reach this conclusion based on hedonism and the termination thesis. Defenders of the deprivation account are willing to grant both hedonism and the termination thesis. They nevertheless arrive at a different conclusion. Neither side is convinced by the arguments of the other. In one way or another, the two sides are talking past each other. But where does their fundamental disagreement lie?

As I see their stand-off, the two sides do not mean the same thing when they talk about harm. Both sides would be willing to grant that death is bad for you only if it harms you. But the defender of the deprivation account relies on a comparativist notion of harm that the Epicurean rejects. To an Epicurean, X can harm a person only if X is absolutely bad. If X is absolutely bad, then it is either intrinsically bad, or it has intrinsically bad consequences. The Epicurean’s idea is thus the following: if X never affects you adversely by being in one way or another intrinsically bad for you, then X does not harm you. This fits well with how we commonsensically think about harm, and with how we use the term in everyday language (see footnote 12). As I see it, the timing puzzle is more aptly understood as a rhetorical device than an

actual argument. On the assumption of hedonism, the timing puzzle adds nothing of substance to the absolute badness thesis. I believe that it is instead meant to bring home the point that death cannot be absolutely bad for the person who dies, and therefore cannot harm her.

If this is correct, then a defender of the deprivation account is right to reject the timing puzzle, or to answer it in a somewhat question-begging way. The deprivation account does not contest the Epicurean idea that death is not absolutely bad. Rather, it is based on the idea that your own death can harm you *even if it does not have any intrinsically bad effects on you*. In light of this, the timing puzzle’s challenge to make tangible the harm of death by locating it in time appears misguided even on the assumption of hedonism. A harmful death *robs the person who dies of a better life*—surely this makes clear how our death can harm us! What could possibly be gained by locating the harm of death in time?

I believe that the situation between Epicureans and defenders of the deprivation account represents what John Martin Fischer (2006, p. 366) calls a “dialectical stalemate”. Here is Fischer’s characterization of the term:

“Frequently in philosophy we are engaged in considering a certain argument [...] for some claim C .¹⁵ The argument employs a principle P .¹⁶ Allegedly, P supports C . Now the proponent of the argument may be called upon to support the principle, and may do so by invoking a set of examples (or other considerations).¹⁷ Based on these examples (or other considerations), he argues that the principle and thus also the philosophical claim are to be accepted.

But the opponent of the argument may respond as follows. The examples are not sufficient to establish the principle P . One could embrace all the examples and not adduce P to explain them; rather,

¹⁵ In our case, C is the claim that a person’s death cannot be bad for her, or that it cannot harm her.

¹⁶ In our case, P is the principle that ‘being absolutely bad’ is a necessary condition for ‘being a harm’.

¹⁷ In our case, an Epicurean would invoke examples such as Bjorn’s wonderful massage to support the claim that if X is overall bad for me but not absolutely bad for me, then it does not harm me. See section 2.3.2.

it is alleged that a weaker principle, P^* ,¹⁸ is all that is decisively established by the examples (or other considerations). Further, P^* , in contrast to P , does not support C . Finally, it is very hard to see how one could decisively establish P . One reason it is so difficult is that it at least appears that one cannot invoke a particular example which would *decisively* establish P without begging the question in a straightforward fashion against either the opponent of P or the opponent of C . Further, it *also* seems that one cannot invoke a particular example which would *decisively* refute P without begging the question against the proponent of P or the proponent of C . These conditions mark out a distinctive—and particularly precarious—spot in dialectical space” (Fischer, 2006, p. 367, emphases in original).

How can we make philosophical progress in a dialectical stalemate? Fischer (2006, p. 372) suggests that “philosophical creativity can issue in a restructuring of the problem”, and that this may allow us to sidestep some of the deadlocks we run into. I suggest that a restructuring of the problem is in order in our case at hand. To see why, suppose we do *not* restructure the problem. Further suppose we find an excellent reason why only absolutely bad things should qualify as harms. We could then confidently claim that when a person’s death deprives that person of a better life, it does not thereby *harm* her. But what exactly would we gain from this? It would not seem to render the deprivation account’s insights irrelevant. Our own death would still be able to deprive us of a better life, and this would still appear significant. Analogously, suppose we manage to come up with an excellent reason why the comparativist definition of harm should be accepted. We could then confidently claim that when a person’s death deprives her of a better life, it thereby harms her. But the main Epicurean insight would remain untouched by this—it would still be true that our own death is not *absolutely* bad for us, and this would still appear significant. In short, if we do not restructure the problem, it is not clear what we can

¹⁸ In our case, P^* could be the principle that most things that harm us are absolutely bad, but that not all of them are.

gain from further progress.

I suggest we restructure the problem by thinking about the badness of death in a distinctly ethical or practical way. Why are we trying to settle whether our own death can be bad for us? What are we hoping to clarify? At least in part, we are trying to find answers to a number of practical questions, such as:

1. Do we have reason to avoid our own death? If some course of action involves an increased risk of premature death, does this fact speak against adopting it? If so, why?
2. What emotional attitude should we have towards our own death? Is it rational to fear your own death, or to be disappointed if you are told that you will die soon? What sort of “self-interested distress” (if any) is warranted in the face of our own death (Draper, 2013a, p. 298)?

If we can have reason to avoid our own death, then there is one practical sense in which our death *can* be bad for us. If it is never reasonable to fear death, then there is a different practical sense in which our death *cannot* be bad for us. Clarifying the ways in which our death can and cannot be practically bad for us seems to me a much more worthwhile project than trying to overcome the stand-off between Epicureans and defenders of the deprivation account.

It could of course be the case that we *need* to overcome the stand-off between Epicureans and defenders of the deprivation account *before* we can tackle any practical questions. But this is not true. Take the second set of questions. The textual evidence suggests that they were Epicurus’ main concern, or that he was ultimately trying to establish that we have *no reason to fear our own death* (Epicurus, 1926, sec. 125). If this is right, then Epicurus was interested in the badness of death because the badness of some X can sometimes provide us with a reason to fear that X. He would then have been right to focus on absolute badness, as ‘being absolutely bad’ is plausibly a necessary condition for ‘meriting fear’.¹⁹ At the very least, the comparativist insight that our own death can be overall bad for us does not suggest that our own death

¹⁹ Draper (2013a, 2013b) argues that Epicurus was right in his assertion that we have no reason to fear our own death. But Draper (ibid.) also suggests that other forms of self-interested emotional distress, for example disappointment, may sometimes be warranted in the face of our own death.

merits fear. On a comparativist understanding of harm, many harms ought not to be feared. Consider Draper's example of the wonderful massage that you are about to receive from Bjorn. In light of the fact that this massage is *wonderful*, it would be crazy to dread it, even if receiving it harms you by depriving you of an even better massage. Despite their differences, Epicureans and defenders of the deprivation account need thus not necessarily be divided over the question whether we have reason to fear our own death.²⁰

In the remainder of this chapter, I will be concerned with the first set of questions, or with the reasons we have to avoid our own death. While each of us has to die eventually, many choices we make can affect *when* we will die. Should we take this into consideration when deliberating about the prudentially right thing to do? From a self-interested point of view, do we have reason to try and put our own death off into the future?

2.4 Death from a Deliberative Perspective

2.4.1 Choosing the Best Expected Life

The deprivation account implies that you have reason to put your own death off into the future whenever this maximizes the value of the life that you can expect to lead.

To give an example, suppose you are on holiday in New Zealand, and you are thinking about doing a bungee jump. You are quite tempted to jump: the rush of adrenaline in the time leading up to the jump seems desirable to you, and you would love to live through the intense experience of a free fall. But suppose you also know that if you jump, you will die. Instead of oscillating to a halt on the bungee cord, you will in fact hit the ground and die a quick and painless death. Should you jump? The deprivation account can help you think about this question in a structured manner. It tells you that your deliberation should take the form of a comparison between the goodness of two possible lives. More specifically, it advises you to compare the goodness of the shorter life that you will lead if you decide to jump to the expected

²⁰ Considering whether we have reason to fear our own death need not only be of practical interest. If our own death does not merit fear, this might speak against thinking of it as harmful.

goodness of the longer life that you think you will lead if you do not jump. If your best estimate is that the shorter life will all things considered be more intrinsically valuable than the longer life,²¹ you should jump; else you should not.

Now as a deliberating agent, you do not usually know with certainty what different options have in store for you. When on a specific occasion you think about whether or not to bungee jump, your epistemic situation will most likely be such that it is reasonable to assume that when compared to the alternative of not jumping, the option of jumping brings with it an *increased* risk of dying today, a risk that need not however be anywhere near certainty. This type of epistemic limitation poses no fundamental difficulty for the deprivation account, as we can still apply its basic framework once we pair it with some standard decision-theoretic axioms. To give an example, suppose that after gathering some information, you estimate there is a 5% chance that your bungee jump will be lethal. You furthermore regard the risk of dying today if you do not jump as close to zero. Should you jump? The deprivation account tells you to evaluate the value of the two options—not jumping vs. jumping—as follows. First, the value of not jumping is the goodness of the life you expect to lead if you do not jump. Suppose your best guess is that if you do not jump, you will live a long and fairly fulfilled life whose value is somewhat diminished by the fact that it lacks a bungee jump experience. Let's say you assign an overall value of 400 to this life. Next suppose that if you jump and die, you estimate that your short life will have an overall value of 200. It will be a brave and impulsive but tragically short life. If you jump and do not die, you expect to lead a life with an intrinsic value of 410. It will be a life similar to the one you expect to lead if you do not jump, but you think it will be slightly better because of the bungee jump experience it will contain. The expected value of the lives you might lead if you jump will then be 399.5,²² which is less than the value of 400 that you have assigned to not jumping. Assuming risk-neutrality,²³ it follows that you should not jump. The deliberative

²¹ On whatever theory of value you endorse.

²² This is the sum of the probability-weighted values of the two possible outcomes if you jump, i.e. $200 \times 0.05 + 410 \times 0.95 = 399.5$. Note that it is assumed here that you measure 'goodness of a life' on a cardinal scale (see footnote 10).

²³ We can easily accommodate risk-averse and risk-loving attitudes by applying suitable concave or convex functions to the intrinsic values of different possible lives.

framework implied by the deprivation account can easily be generalized. Whenever you are faced with a choice where it is reasonable to assume that the different options come with different distributions of the risk of death over time, you should choose an option which you calculate to have the best expected life in store for you. If longer lives tend to be more intrinsically valuable than shorter lives, deliberation based on the deprivation account will sometimes lead you to renounce otherwise worthwhile activities or valuable experiences simply because they involve an increased risk of premature death.²⁴ Think about decisions to eat healthier, to drink less wine, or not to travel to an unsafe destination that you would be interested in visiting. To the extent that you take such decisions because they reduce your chances of a premature death,²⁵ you may sometimes wonder whether you are missing out on the good things in life because of an irrational aversion to death. But as long as you renounce valuable experiences based on an application of the insights of the deprivation account, your awareness of the badness of death does not keep you from appreciating the good things in life. Quite to the contrary, it makes you act in such a way that you can hope to maximize your enjoyment of them in the long run.

2.4.2 Epicurean Maximax

As I have presented their position, Epicureans do not deny that our own death can deprive us of a better life. They simply don't think that this matters. More specifically, they argue that a person's death cannot be absolutely bad for her, and that 'absolute badness' is the relevant notion when it comes to determining whether death is bad for the person who dies. I have suggested that when our focus is on the question whether a person should fear her own death, the Epicurean stance is reasonable (see section 2.3.4). But can we make sense of it in our current context also? Can we construct an Epicurean deliberative framework which yields that we do

²⁴ By an 'otherwise worthwhile activity or valuable experience' I mean an activity or experience that you would expect to be *overall good* for you if it did not increase your risk of dying a premature death.

²⁵ Of course we also take such decisions for reasons other than their effect on the risk of dying prematurely. I believe that the bungee jump example is a relatively 'pure' case, in that it seems reasonable to suppose that at least for some people, the only consideration that keeps them from jumping is the thought that they might die.

not have reason to avoid our own death?

I think we can. Consider again the bungee jump example. Suppose you are trying to decide whether you should jump, and you focus your attention on the fact that your death would not be absolutely bad for you. Your reasoning might then go something like this: “If I do not jump, I will live. But I will wonder what it would have felt like to jump; I will think that I have missed out on something. Also, I will think of myself as a coward. If I jump, there are two possible outcomes. Either I die, or I live. Suppose I live. It will then be a good thing for me to have jumped. I will feel exhilarated and proud. The experience of the jump will make my life richer. Now suppose that jumping kills me. I will not then be around to regret my decision to jump. If I die, it will simply not matter to me that I have decided to jump. It follows that I should jump: I will only ever be glad to have done so.” To put the same point more generally and more technically, we can interpret the Epicurean stance as advocating a Maximax strategy²⁶ for a specific type of decision situation, namely the type of situation where we have to decide whether to opt for an otherwise valuable experience or take up an otherwise worthwhile activity when these come with an increased risk of premature death. In this type of situation, the Epicurean position can be taken to imply that we should *never* renounce an otherwise valuable experience *merely* because the experience involves an increased risk of premature death.²⁷

²⁶ In decision theory, an agent adopts a *Maximax rule* if she always chooses that option for action which gives her a chance at the best possible outcome, even if that chance is slim. If Sally can either go to the beach or stay home, and she expects to have (i) a reasonably good time if she stays home, (ii) a great time if she goes to the beach and it is sunny, and (iii) a horrible time if she goes to the beach and it starts to rain, then Maximax tells her to go to the beach as long as she assigns a probability greater than zero to the weather being sunny.

²⁷ I do not mean to suggest that Epicurean Maximax would be endorsed by a modern day Epicurean, or would have been endorsed by a historical Epicurean. For the modern day Epicurean’s likely views, see section 2.4.3. Regarding the historical Epicurean, Epicurean Maximax does not fit well with what we know of Epicurus’ ideas about the good life (Epicurus, 2006; see also Warren, 2004). Epicurus rejected an aggregative idea of pleasure—he did not think that having *more* pleasurable experiences would make one’s life go *better*. He did not advocate thrill-seeking, either; instead he told his followers to opt for a life of moderation. However, when it comes to the badness of death, Epicurus’ disciple Lucretius argues that we wrongly believe death to be bad because we erroneously picture how once we are dead, we regretfully look back at all the good things we have lost (see Nussbaum, 1999, p. 811). It was one of the aims of historical Epicureanism to rid us of this erroneous picture. Once we combine this aim with the now standard view that pleasure is an aggregative good (see footnote 10), we arrive at what I have called *Epicurean Maximax*.

Epicurean Maximax and deprivation account-based deliberation are inconsistent. Where the former tells you to take the plunge, the latter may advise restraint. Can one be defended over the other?

I believe so. Succinctly put, if you apply the deliberative framework implied by the deprivation account, you pick that option which gives you the *best expected life*. Contrary to what Epicurean Maximax implies, a guaranteed absence of regret does not give you reason to pick a different option. Consider the following analogy. Suppose you can either enter lottery *A* in which you will win £900,000 with probability 1, or lottery *B* in which you will win £1,000,000 with probability 0.1, and nothing otherwise. Further suppose that if you enter lottery *B* and do not win, you will instantly forget that you have had a choice between *A* and *B*, and that you chose *B*. Which lottery should you choose? (Lottery *A* and its prize represent the option of not jumping and the expected value of a long life without a bungee jump experience. Lottery *B* represents the option of jumping. Its £1,000,000 prize represents the expected value of a long life that contains a bungee jump experience; winning nothing and forgetting about the whole incident corresponds to dying in the bungee jump example). Epicurean-style deliberation would have you pick *B*, as you will only ever be glad to have done so. By construction, there is no way you will ever regret choosing *B*. Whereas if you pick *A*, you will have to live with the nagging feeling of having missed out on the £1,000,000 that *B* might have offered.

I believe this analogy clarifies what is wrong with Epicurean Maximax. First, if you choose *A*, you do not—or at least should not—afterwards regret having chosen it (just as you should not regret your choice not to jump in the bungee jump example). Lottery *A* gave you the highest expected reward, and it was thus rational to choose it. You do not have to feel that you have ‘missed out’ on the £1,000,000 that *B* ‘offered’. What *B* offered was a 10% *chance* at £1,000,000, which simply was not good enough given the alternative of *A*. What about the fact that you would never regret choosing *B* if you chose it? Regret decreases the intrinsic value we derive from payoffs that are not already specified in intrinsic value terms (see Broome, 1991b, ch. 5). In intrinsic value terms, the payoff of ‘not winning anything in lottery *B* when you *know* you could have gone home with £900,000’ is not zero, but negative. The expected

value of lottery *B* is therefore not 10% of the intrinsic value that you would derive from winning £1,000,000, but 10% of the intrinsic value of winning £1,000,000 *less 90% of the intrinsic disvalue you would derive from being left with nothing when you know you could have gone home with £900,000*. With regard to gambles that involve regret-tinted payoffs, we will therefore be more risk-averse than we would otherwise be. If we are guaranteed *not* to feel regret, we should accordingly be *less* risk-averse than we would otherwise be. If lottery *B* offered you a 95% chance at a £1,000,000, it could be rational for you to choose *B* even if you would be too risk-averse to choose it if you weren't guaranteed to forget about the whole incident if you lost out. To return to life-and-death cases, *if* in cases where some options involve an increased risk of premature death we have risk-averse attitudes *partly because of considerations of regret*, then our risk-aversion is *to that degree* unfounded, and we should shift our attitudes towards risk-neutrality. But there is no reason why in the absence of regret we should disregard probabilities entirely, and always choose that option with the best possible outcome no matter how unlikely that outcome's occurrence. Yet this is exactly what Epicurean Maximax would have us do.

2.4.3 Focusing on Intrinsic Value

Epicurean Maximax is not a tenable approach. But an Epicurean need not be committed to it (Draper, 2013b, p. 74). It is open to an Epicurean to argue that we can have reason to avoid some X without X thereby being bad for us. To return to Draper's example, suppose you are booking an appointment at Bjorn and Sven's House of Swedish Massage (Draper, 2013b, p. 78; Draper, 2013a, p. 299). The receptionist informs you that there are a number of open slots, and asks you what slot would be convenient for you. If you know that you will enjoy a massage from Sven *more* than you will enjoy a massage from Bjorn, you have reason to inquire about an appointment with Sven, and to avoid booking a slot where you end up with a massage from Bjorn. This holds true even if a massage from Bjorn is a genuinely wonderful experience.

Here is how an Epicurean might argue: "When picking an option or deliberating about the right thing to do, you should let yourself be guided by the *intrinsic value*

that you expect to derive from the different options. More specifically, you should always pick that option which promises to be the most intrinsically valuable. But this does not imply that all the options which you should not pick are thereby bad. A wonderful massage does not turn into something bad or harmful just because there is an even better alternative available.”²⁸

If we focus on the intrinsic value that we expect to derive from different alternatives for action, and if we always pick that option which we expect to have the highest value, we deliberate in exactly the same way that we would deliberate in if we applied deprivation account-based decision rules. That is, we consistently avoid otherwise worthwhile activities and experiences that expose us to an increased risk of premature death if avoiding them maximizes the value of the life that we can expect to lead.

The position just sketched seems to me a tenable one, but I am not sure how reasonable it is to want to defend it. Someone who defends it agrees that we have reason to avoid alternatives that a comparativist would refer to as ‘overall bad’. Moreover, he agrees that we have reason to avoid these alternatives *because* they satisfy the criterion that makes them overall bad, or because some other alternative would be *better*. But he refuses to call an overall bad alternative bad unless it is also absolutely bad. Why not just grant that an *overall bad* alternative is *by definition one that we have reason to avoid*? Maybe not everything that is overall bad for us should be thought of as a harm (see section 2.3.2). But the fact that we have reason to avoid some X makes for an intelligible sense in which that X is bad, and it is not clear to me what is gained by refusing to call it (overall) bad (cf. Olson, 2013). Suppose an Epicurean tells you: “Become accustomed to the belief that death is nothing to you! As long as you are here, death is not; once death is here, you are no more. It follows that your death cannot be bad for you.” You are impressed by the Epicurean, and in an attempt to follow his advice, you start crossing the street without looking left or right. He grabs you by the shoulder and yells: “Are you out of your mind? I never said that you should not avoid death!” The point I am trying to illustrate is the following: an Epicurean cannot insist that a person’s death cannot be bad for her without at least suggesting that we have no self-interested reason to avoid our own

²⁸ That this would be the modern day Epicurean’s view was suggested to me by Jens Johansson.

death. But if he believes that we have such a reason, he should be willing to make this clear. Just like a defender of the deprivation account should agree that death is probably *not* bad in the sense that it merits fear, an Epicurean should therefore agree that death *is* bad in the sense that we can have reason to avoid it.

2.5 An Agency-Based Account of the Badness of Death

2.5.1 A Problem For the Deprivation Account

Suppose that you are confronted with a choice between two options. Let us refer to them as *Ten* and *Twenty*. If you choose option *Ten*, you will live for ten more years; if you choose option *Twenty*, you will live for twenty more years. Suppose it is guaranteed that no matter which option you choose, the intrinsic value of the life you will have led when you die will be the same. Let's say that for each life, its eventual value will be 500. If you choose *Ten*, you will thus live a shorter life with a value of 500; if you choose *Twenty*, you will live a longer life with the exact same value. Which option should you choose? According to the deliberative framework implied by the deprivation account, it is a matter of indifference which option you choose.²⁹ Personally, I do not however feel indifferent about this choice: I would clearly prefer to live the longer life. Based on the deprivation account, this preference cannot be explained. It can only be explained away, e.g. by arguing that I am under the misapprehension that *Twenty* implies an all things considered better life than *Ten*. A defender of the deprivation account could argue that my misapprehension is understandable, as living a longer life will usually be better than living a shorter life, though in the case at hand, *Twenty* is by stipulation no better than *Ten*. But even on reflection, it does not seem to me that my preference for *Twenty* rests on a mistake.

In this section, I propose a new way of looking at the badness of death that is

²⁹ It is also a matter of indifference according to the Epicurean framework which tells you to focus on the intrinsic value that you can expect to derive from different options (see sec. 2.4.3). Since the two frameworks effectively coincide, and since the Epicurean refuses to call 'bad' what we merely have reason to avoid, I present the problem that I introduce in this section as a *problem for the deprivation account*.

meant to complement the deprivation account, and which renders my preference for the longer life reasonable. More precisely, I argue that we have a reason to avoid death which is *also* a reason to prefer *Twenty* over *Ten*. Kamm (1993) agrees that such a reason is called for. She argues that as human persons, “we could prefer to postpone things being all over even if this did not increase the total amounts of good we had in our life”, and concludes that “we must [thus] be trying to avoid something about death other than that it diminishes the amount of goods of life we have” (Kamm, 1993, p. 19). In essence, I will argue that to appreciate what “we are trying to avoid about death” when we choose *Twenty* over *Ten*, we have to move away from one basic understanding of human persons—one where we focus on the fact that we are bearers of experiences—towards another: one where we see ourselves first and foremost as agents.

2.5.2 Death as a Frustrator of Categorical Desires

For a first stab at this new way of looking at the badness of death, return to the bungee jump example. In section 2.4.1, I have argued that the deliberative framework implied by the deprivation account can help you structure your thinking about whether you should jump. But is deprivation account-based deliberation realistic? That is, does it resemble the way in which we would actually reach a decision? Try to imagine what considerations would cross your mind if you considered jumping. Assuming that I were tempted to jump, it seems to me that my main consideration would be whether the risk of dying was negligible.³⁰ If I thought it was, I would jump; otherwise, I would not. If I decided not to jump and someone asked me why I decided not to take the plunge, I would tell them that it was too risky—that I did not want to die, and that jumping just did not seem worth it. If the person then asked me why I did not want to die, I would tell them that there were still so many things that I wanted to do and see. That I wanted to run a marathon at some point in my life. That I wanted to continue my studies in philosophy. That I wanted to learn Spanish, read Borges in the original, and travel across South America. That I wanted to see my niece grow up.

³⁰ At first, I might also be worried about the risk of serious injury, but I would then reason myself to the conclusion that any serious accident was extremely likely to be fatal.

That I might want to have my own children some day.³¹ If asked about why I did not want to die, I would thus list all of my possibly merely potential plans and projects for the future, implying that my death would be bad for me because it would rule out their realisation. Bernard Williams (1973b, p. 86) has coined the term “categorical desires” to describe such plans and projects. He calls them “categorical” because they make us want to stay alive in order to be able to realize them—they “[propel us] forward into longer life” (Williams, 1973b, p. 91). This distinguishes them from “conditional desires” (Williams, 1973b, p. 85) whose fulfillment we care about only on condition that we stay alive, such as the desire not to go hungry. While categorical desires give us a reason to stay alive, conditional desires do not.

Two things are notable here. First, when I tell my interlocutor that I do not want to die because there is a long list of things that I still want to do, it may seem that I am merely restating the deprivation account’s core insight in an imprecise way. Second, this seeming resemblance between the deprivation account and what I say does not withstand scrutiny, as what I imply about the badness of death does not in fact involve a comparison between different lives, and is thus *structurally different* from the explanation put forward by the deprivation account.³² Let us take a closer look at each of these observations in turn.

As for the seeming resemblance between the deprivation account and my statement that I do not want to die today because I have plans and projects for the future, we can interpret this statement to mean that I expect to live a better life if I do not die today, but continue to live for a good while longer. If I continue living, I can expect to realize at least some of my plans, and it seems reasonable to assume that a longer life with more plans come to fruition is a more valuable life than a shorter one with many plans thwarted by an early death. While on reflection I may well confirm that my death today would seem to be overall bad in this way, this comparative reasoning is *not* what underlies the assertion that dying today would be bad for me because it

³¹ And so on. I hope my interlocutor would make it clear after a few examples that I had made my point.

³² Williams (1973b) does not seem to notice this. He locates the main difference between his ideas about the badness of death and Nagel’s deprivation account in the fact that his ideas do not require a theory of value according to which something can have value for us independently of our desires (Williams, 1973b, pp. 88-9).

would mean the frustration of my categorical desires.³³ The idea that death is bad because it frustrates categorical desires is not a comparative idea—it is simply the insight that there is something bad about death insofar as we have such desires: it thwarts them irrevocably. To say that my death today would be bad in this way, I do not need to know or even take a guess what life I would lead if I did not die today. Suppose the life I would lead would be terrible. Suppose it would be one in which I struggled just to make ends barely meet, and in which none of my plans or projects ever came to fruition. Even if this were the case, it would still be true that dying today would be bad for me in the sense that it would still mean the frustration of my categorical desires. The crux is that staying alive would in such a case be *even worse* than dying, as it would mean the frustration of my categorical desires as well as hardship. When staying alive is worse for me than dying, dying is *overall good* for me, and is thus good for me on the deprivation account. But insofar as my death is bad for me because it frustrates my categorical desires, it is non-comparatively or *absolutely bad for me*—it remains bad for me no matter how awful the alternative of continued life becomes.

Very often, the option of dying today will appear to us both overall bad and bad because it would mean the certain frustration of our categorical desires. Moreover, ‘dying today’ will usually appear overall bad partly because we expect that fewer of our categorical desires would be frustrated if we did not die today. Whenever this is the case, the two structurally very different explanations for the badness of death are easily conflated.

³³ It is not strictly speaking correct that my death today would thwart *all* of my categorical desires. Certain other-regarding categorical desires, such as the desire that my sister’s marriage be a happy one, would not be thwarted by my death. Another type of categorical desires that would not be thwarted by my death are self-regarding desires whose fulfilment does not require my existence, e.g. the desire that no one speak ill of me behind my back. To make it clear that I wish to exclude categorical desires whose fulfilment does not depend on my existence, I prefer to speak of ‘plans and projects’. Throughout this chapter, I will however use Williams’ ‘categorical desires’ interchangeably with my ‘plans and projects’.

2.5.3 How Death Is Bad For Us as Agents

Epicureans and defenders of the deprivation account are in agreement that a person's death is at most comparatively bad for her (and Epicureans for this reason refuse to call death bad). But if what I have just suggested is correct, a person's death is absolutely bad for her when she has plans and projects that it thwarts.³⁴ In this section, I defend the claim that a person's death can be absolutely bad for her by laying out how this claim is to be understood.

To understand how death can be absolutely bad, a shift in perspective is called for. Epicureans and defenders of the deprivation account both assess the value of a human life in a similar way. To them, a human life is an accumulation of events and experiences that a particular person was exposed to and lived through. They then assume that the value of a human life derives—in possibly complex ways—from the value of the events and experience that it consists of. On this view, what is ultimately good or bad for a person is the extent to which the accumulation of events and experiences that makes up her life is an intrinsically valuable one.³⁵ On the termination thesis, this in turn seems to imply that the significance of a person's death can only lie in the fact that it *fixes the intrinsic value of her life at a particular level*.³⁶ But if our death merely fixes the value of our life, then it can never be good or bad for us except in comparison to something else. To see this more clearly, consider the following example. Sally lives a miserable life, and then dies. Her death fixes the value of her life at -100 . Does this negative value mean that her death was in any way bad for her? Not necessarily. It only means that her life was miserable. She

³⁴ Not every death is absolutely bad for the person who dies. As with the comparative badness that the deprivation account captures, the existence and the extent of any given death's non-comparative badness will depend on the details of that death. When I talk about the non-comparative badness of death, I thus use this as shorthand for 'death's non-comparative badness when death is non-comparatively bad'.

³⁵ Thinking of a human life in this way is compatible not only with hedonism, but with a wide range of theories of value. To give an example, Nagel (1974, p. 76) suggests that it is bad for me if my friends betray me even if their betrayal neither causes me any pain nor diminishes the amount of pleasure in my life (e.g. because I never find out). What Nagel suggests is incompatible with hedonism, but entirely compatible with the view that the value of a human life is a function of the value of the events and experiences that it consists of.

³⁶ If there are posthumous harms and benefits, a person's death fixes the value of her life only tentatively.

might have lived an even worse life had she had not died when she did in fact die. If so, it is (comparatively and overall) good for her that her death fixed the value of her life at -100 . Suppose I exclaim: “Sally’s death was a terrible misfortune for her! It frustrated so many of her plans and projects! I feel sorry for the poor woman.” On the assumption that a death merely fixes the value of a life, you should correct me roughly as follows: “Well, her death fixed the value of her life at -100 . This means that she led a horrible life. It is appropriate to feel sorry for her because of the life that she led. But whether it was a misfortune for her that her death frustrated her projects depends on what we take to be the relevant alternative.”

I don’t think that it is wrong to assess the value of a human life the way I have suggested Epicureans and defenders of the deprivation account do. It may even be the only way in which this can sensibly be done. The problem is that it is potentially misleading (Korsgaard, 1996). More precisely, it suggests that as human persons, what we value for its own sake—or what we properly ought to assign intrinsic value to—are the events and experiences that make our life good. But this is a flawed idea. It works well enough on the assumption of hedonism. But I believe that hedonism is false, and that the way in which Epicureans and defenders of the deprivation account assess the value of a human life can make this difficult to see.

As human persons, we are more than sentient beings capable of feeling pleasure and pain: we are also rational beings who live our lives as *agents*. Once we focus on what we value insofar as we are agents, the non-comparative badness of death becomes perspicuous.

What does it mean to be an agent? In essence, it means to be someone who lives her life guided by goals and principles she has set for herself (Korsgaard, 1996, pp. 387-8). To the extent that we live our lives as agents, we live a life that fits our nature as rational beings. Following Kant, Nozick (1974, pp. 49-50) suggests that the distinguishing feature of a rational being is its capacity to govern itself in the light of reasons. To Nozick, a rational being is “able to formulate long-term plans for its life, able to consider and decide on the basis of abstract considerations or principles it formulates to itself and hence not merely the plaything of immediate stimuli, a being that limits its own behavior in accordance with some principles or picture it

has of what an appropriate life is for itself and others” (Nozick, 1974, p. 49). As rational beings, our lives are meaningful and valuable only if we are able to live them as agents—i.e. purposefully, and in accordance with our own ideas of the good (see Rawls, 1971b, pt. 1, sec. 22).

Insofar as we are agents, what we intrinsically value is *setting our own goals and pursuing them in an effective way*. What promotes our doing so is good for us; what detracts from it is bad for us. If our plans and projects come to fruition, this is good for us because it makes it the case that we were pursuing our goals in an effective way. If our plans and projects are thwarted, this is bad for us because it means that our agency was ineffective. Crucially, we do *not* value pursuing our goals effectively because ‘achieving a goal’ is a good event, and we attempt to ‘collect’ as many good events as we possibly can. This way of looking at things puts the cart before the horse.

To see this more clearly, consider again Sally, whose death frustrates her plans and projects, and fixes the value of her miserable life at -100 . On the view that I have ascribed to Epicureans and defenders of the deprivation account, it would have been good for Sally to complete her projects *because ‘completing a project’ is a good event that would have added to the value of her life*. In a very similar way, it would have been good for Sally to receive a massage from Sven. ‘Completing a project’ and ‘receiving a massage’ are both good events that add to the value of a person’s life. As a matter of fact, there are infinitely many good events that *would* have made Sally’s life better had they occurred during her lifetime. Do we want to say that all of these possible events made her death bad for her? More precisely, do we want to say that a person P ’s death is bad for P as long as there is at least one possible event X so that P ’s life would have gone better had it contained X ? Probably not.³⁷ This view would imply that death is bad for a person who will remain in severe chronic pain for the rest of her life, and who therefore longs for death. But if not every possible event that would have made Sally’s life better makes her death bad for her, it depends on a *relevant alternative scenario* whether it was (comparatively and overall) bad for Sally

³⁷ Nagel might disagree. He suggests that “[if] there is no limit to the amount of life that it would be good to have, then it may be that a bad end is in store for us all” (Nagel, 1974, p. 80).

that her death kept her from completing her projects.

By contrast, on the agency-based view just introduced, ‘having one’s projects thwarted’ and ‘not receiving a massage’ are structurally different. While it is *absolutely bad* for Sally to have her projects thwarted, it is *not* absolutely bad for her to miss out on pleasurable experiences (such as a massage from Sven). By thwarting her projects, her death makes it the case that Sally was not pursuing her goals in an effective way; it makes it the case that the hardships which she endured and the effort that she put in to bring her ideas to fruition were largely in vain. Insofar as we are agents, these things are bad for us in exactly the same way as getting punched in the nose is bad for us insofar as we are beings who are capable of feeling pain.

What about the Epicurean challenges? On my interpretation of the first Epicurean argument (see section 2.3.2), its core claim is that a person’s death cannot be absolutely bad for her. This argument depends on the assumption of hedonism. I agree that death can neither be painful nor have painful consequences for the the person who dies. But a person’s death can clearly make it the case that her projects remain unfinished. I have argued that when death frustrates our projects, it is absolutely bad for us as agents. To make it the case that a thwarted project is bad for us agents, it is not necessary that a thwarted project causes us to feel pain. Rather, sadness and disappointment are appropriate when we learn that one of our projects has failed because we learn that something bad has happened to us (cf. Nagel, 1974, p. 76).³⁸

When do we suffer the misfortune that our death causes us when it frustrates our plans and projects? I am not sure that this question needs an answer. Once we give up hedonism, it is not clear to me why everything that is bad for us should affect us adversely *at a particular time*. Having said that, *priorist* defenders of the deprivation account answer the timing puzzle by arguing that when our death is comparatively bad for us, it affected us adversely *while we were alive*. On their preferentialist view of welfare, a person’s death affected her adversely whenever she held desires that her death ensured would never be fulfilled (but that would have been fulfilled had she not died when she did in fact die). Suppose my death is comparatively bad for me because it thwarts one of my projects that I would have completed had I not died when I did

³⁸ This is not to deny that sadness and disappointment can *add* to the badness of a failed project.

in fact die. To a priorist, I might then have suffered the misfortune of my death at all those times at which I was actively pursuing my project while alive (Luper, 2004, and Luper, 2009b, 134-36; see also Scanlon, 1998, pp. 119-23).³⁹ Similarly, if my death is absolutely bad for me by thwarting one of my projects, it could be argued that I suffered its misfortune at all those times at which I was actively pursuing my project. On this priorist way of looking at things, a person's death affects her adversely *insofar as she is an agent* whenever she makes an effort that her death subsequently renders futile (cf. Scanlon, 1998, p. 121).⁴⁰

2.5.4 Agential Reasons to Put Death Off Into the Future

On the deprivation account, and on at least one interpretation of the Epicurean position, you have an all-things-considered reason to try and put your own death off into the future whenever it is plausible to assume that this maximizes the value of the life that you can expect to lead. In this way, you can sometimes have reason to avoid an intrinsically valuable activity or experience merely because engaging in it or enjoying it would involve an increased risk of premature death.

³⁹ According to the preferentialism that Steven Luper (2009b, pp. 93-4) argues should form part of a correct theory of welfare, it is intrinsically good for a subject *S* if at time *t*, some proposition *P* (e.g. 'I will be married some day') holds and *S* desires *P*; it is intrinsically bad for *S* if at *t*, *not P* holds and *S* desires *P*. On this view of what is intrinsically good and bad for us, if my death thwarts one of my projects, this was intrinsically bad for me at all those times at which I *desired* that I would complete my project. By contrast, Thomas Scanlon (1998, p. 121) defends an *achievement preferentialism* where the fulfillment of an aim "does little or nothing to make my life better" if I *merely desired* to reach that aim. To Scanlon, the "aims whose fulfillment makes a significant contribution to a person's well-being are ones that the person has actually acted upon, and, typically, given a role in shaping his or her other activities and plans. The fulfillment of that aim then makes a difference to the person's life by making these plans and activities successful ones" (ibid.). While Scanlon's achievement preferentialism fits my agency-based ideas about the badness of death, it is Luper who (based on Pitcher, 1984) argues that on a preferentialist view of welfare, we can provide a priorist answer to the timing puzzle.

⁴⁰ On a preferentialist view of welfare (see footnote 39), it follows rather straightforwardly that death can be absolutely bad for the person who dies. Given that Luper (2009b, pp. 92-101; 134-6) endorses a priorist answer to the timing puzzle based on a preferentialist view of welfare, it is puzzling that he does not draw our attention to this fact. Luper (2009b, p. 101) even makes it clear that if absolute badness were a necessary condition for harm, then our death would not be able to harm us. What explains this oversight? I am not sure. As a comparativist, Luper (2009b, p. 100) defends the view that death harms the person who dies if and only if it is overall bad for her. On the further assumption that death is relevantly bad for the person who dies only if it harms her (see sec. 2.3.4), it follows that death is relevantly bad only if it is overall bad. As a comparativist, Luper may thus have focused on overall badness, and he may have treated absolute badness as somewhat of a side issue.

Insofar as you are an agent—i.e. insofar as you are someone who values setting your own goals and pursuing them in an effective way—you have a number of further reasons to put your own death off into the future. Unlike the reason that the deprivation account provides you with, these *agential reasons* are not necessarily all-things-considered reasons. That is, even when they speak in favour of putting death off into the future, they need not do so decisively, as opposing reasons may outweigh them.

First—and most obviously—you have an agential reason to avoid your own death whenever you have ongoing plans and projects that it would thwart. In the bungee jump example, if you are in the process of writing a book, this gives you a reason not to jump even if you consider a bungee jump a valuable experience. Of course, if the risk of death involved in a bungee jump is negligible, or if the value of a bungee jump experience is extremely high, or if you will most likely not finish your book project anyway, jumping may still be the all-things-considered right thing to do even if your death today would frustrate your book project.

Second—and less obviously—when it comes to decisions to eat healthier or to smoke less, you have an agential reason to try and put your own death off into the future because this gives you time. As agents and at any given point in time, we value the temporal space ahead of us as a precondition for living a purposeful life. Perceiving the future as open enables us to think of our ongoing plans and projects as meaningful, and it allows us to change our plans and to start new projects without having to fear that we are involved in a futile activity. Imagine that you are told by a psychic whose opinion you have reason to trust that you will die within the next two years. The psychic's announcement would probably render several of your current projects meaningless, and it would lead you not to take up new projects that you would otherwise have taken up. With your options as an agent seriously curtailed, you might shift your focus towards pleasurable experiences instead. To give an analogy, as agents we value having time at our disposal just like a painter values having canvas at her disposal. If a painter knows or has to fear that she only has a few sheets of canvas left, this seriously constrains her. Knowing that she has a large supply of canvas is both liberating and empowering.

The painter analogy allows us to relate our agential reasons to put death off into the future to our worries about growing old. One thing that we worry about with respect to growing old is the fact that it brings with it a decline in mental and physical health. As agents, our mental and physical health is important for us because it co-determines the range of plans and projects that we have a realistic option of pursuing. As our mental and physical abilities decline, there are fewer goals that we may realistically pursue. Think of the level of an agent's mental and physical health as the number and quality of colours available to a painter. The more high-quality colours she has available, the more numerous the ways in which she can express her artistic ideas. To the painter, canvas and colours are complementary goods: once she starts running out of colours, she rationally puts less value on having canvas available. Of course, if a painter is extraordinarily creative, she may still produce great art with only two or three high-quality colours at her disposal. But once there is not a single colour of good quality left, the painter has no more use for her canvas. Insofar as growing old means a decline in mental and physical health, growing old thus reduces the strength of our agential reasons to put death off into the future. Moreover, to an old person who has no plans and projects left, there is nothing absolutely bad about death. This holds true even if her death is still comparatively bad for her, e.g. because she is not in pain, and because she still enjoys her morning tea, the view from her window, and the warmth of her cat resting on her lap.

2.5.5 Solving the Problem for the Deprivation Account

In section 2.5.1, I have raised a problem for the deprivation account. I have claimed that if confronted with a choice between living for ten more years or for twenty more years, it seems reasonable to prefer to live the longer life, even if both lives were known to be equally good. The deprivation account cannot make sense of this preference.

On the agency-based account, my preference for *Twenty* over *Ten* is reasonable. *Twenty* provides me with more time than *Ten*. Because of this, it leaves me with more options worth pursuing. As an agent, this is something I value even if I know that it will have no effect on the value of the life that I will eventually have led. If I know that I have twenty years left to live, it makes sense to consider starting a family,

or to change my career and become a veterinarian. It makes sense to consider these projects because I can hope to realize the main goals that are involved in them before I die. By contrast, if I know that I only have ten years left to live, I may not be able to start a family, or to radically change my career, without thinking of myself as selfish or foolish. This holds true even if having a family and finding a job I love form part of my idea of the good life.

If the life implied by *Ten* were better than the life implied by *Twenty*, I would face a trade-off. If I were to choose *Ten*, I would live an all-things-considered better life, but I would face more constraints on my agency. If I chose *Twenty*, I would live an all-things-considered worse life, but I would be able to pursue more plans and projects. Depending on the details of that trade-off, it seems to me that my decision could understandably go either way. All things considered, it would of course be better to choose *Ten*, simply because *Ten* is by assumption the overall better option than *Twenty*. But suppose I was informed that *Twenty* was worse for me than *Ten* not primarily because of failed projects and thwarted plans, but because I would be too busy to stop and smell the roses if I picked *Twenty*. If my agential side got the better of me, I might then pick *Twenty* despite knowing that *Ten* would all things considered be the better deal.

2.6 Conclusion

Epicureans claim that death cannot be bad for the person who dies. Based on the assumptions of hedonism and the termination thesis, they argue that a person's death cannot be absolutely bad for her. According to my understanding of the Epicurean position, they furthermore assume that death could harm us—and thus be *relevantly* bad for us—only if it were absolutely bad.

Defenders of the deprivation account beg to differ. On their comparativist understanding of harm, X harms us just in case X is overall bad for us, or just in case our lives would have gone better in the absence of X. On the comparativist understanding of harm, there is no doubt that a person's death can harm her.

In this chapter, I have argued that we do not need to settle the true meaning of harm to decide whether death can be bad for the person who dies. I have suggested that we should understand the badness of death—or its different 'badnesses'—in a practical sense. If we sometimes have reason to avoid death, then there is *one practical sense* in which our death can be bad for us; if it is never appropriate to fear death, then there is *a different practical sense* in which our death cannot be bad for us.

On the deliberative framework implied by the deprivation account, we have reason to avoid our own death just in case this maximizes the value of the life that we can expect to lead. When we have reason to avoid our own death, it is in our self-interest to forego an enjoyable activity or experience simply because it would come with an increased risk of premature death. On one interpretation of the Epicurean position, we have no reason to avoid our own death, as we will not feel regret once we are dead. I have shown that this interpretation of the Epicurean position does not yield a defensible decision rule. But the Epicurean need not be committed to the indefensible *Epicurean Maximax*; it is open to him to argue that when deliberating about the right thing to do, we should always focus on the intrinsic value that we can expect to derive from different options. If he argues in this way, he ends up with the same deliberative framework that a defender of the deprivation account would endorse. But if the

Epicurean agrees that we can have reason to avoid our own death, he should *also* agree that there is a practical sense in which our death can be bad for us. By denying that our death can be bad for us, the Epicurean at least suggests that we have no reason to avoid it.

As agents, we have reason to avoid our own death when we have plans and projects that it would thwart. I have argued that insofar as we are agents, we value setting our own goals and pursuing them in an effective way. If our death thwarts our projects, it renders some of our past efforts futile. As agents, this is bad for us whether or not we are around to lament it. Despite appearances to the contrary, the deprivation account is unable to capture the distinctive way in which our death is bad for us as agents. Most importantly, our death can be bad for us as agents even if it is good for us according to the deprivation account. If Sally's death thwarts her projects, her death is absolutely bad for her even if she would not have been able to complete her projects, and would have suffered terrible hardship had she stayed alive. A person's death can also be bad for her on the deprivation account without being bad for her as an agent. Death is not absolutely bad for a very old person who pursues no further plans or projects. But if this person is not in pain, and if she enjoys a strong cup of tea and the company of her cat, her death may well be overall bad for her.

When a person's death would be overall bad for her without being absolutely bad, this person lives her life primarily as a contented *bearer of experiences*. For such a person, the goodness of life and the badness of death resemble the goodness of life and the badness of death of a contented sentient being who lacks agency. These ideas seem to me beautifully captured in Edgar Lee Masters' poem *Willie Metcalf*, with which I end this chapter.

Willie Metcalf

I was Willie Metcalf.
They used to call me "Doctor Meyers"
Because, they said, I looked like him.
And he was my father, according to Jack McGuire.
I lived in the livery stable
Sleeping on the floor
Side by side with Roger Baughmans bulldog,
Or sometimes in a stall.
I could crawl between the legs of the wildest horses
Without getting kicked—we knew each other.
On spring days I tramped through the country
To get the feeling, which I sometimes lost,
That I was not a separate thing from the earth.
I used to lose myself, as if in sleep,
By lying with eyes half-open in the woods.
Sometimes I talked with animals—even toads and snakes—
Anything that had an eye to look into.
Once I saw a stone in the sunshine
Trying to turn into jelly.
In April days in this cemetery
The dead people gathered all about me,
And grew still, like a congregation in silent prayer.
I never knew whether I was a part of the earth
With flowers growing in me, or whether I walked—
Now I know.

Edgar Lee Masters

Spoon River Anthology (1915)

3 | Moral Status and the Paradox of Deontology^{*}

“Do no harm, but take no shit.”

– Stephi McKeague, via Facebook, 26 July 2014

3.1 Introduction

Is it morally permissible to kill one innocent person as a means to save five other people from being killed by a vicious murderer? If you care about moral rights to non-interference, your answer will most likely be ‘no’. But why shouldn’t your answer be an unambiguous ‘yes’? If you care about rights, shouldn’t you think that rights violations are bad and ought to be kept to a minimum, so that it should always be permissible to infringe one person’s right in order to keep five other people from having their comparable rights violated? Nozick (1974, pp. 28-30) was the first to draw attention to this problem. Today, it is standardly referred to as the *paradox of deontology*.

In this chapter, I argue that a version of Kamm’s inviolability approach to solving the paradox of deontology can be fleshed out into a viable solution to the paradox. The inviolability approach is a victim-focused approach to answering the paradox: it appeals to facts about the potential victim to explain why it should not always

* I thank Gabriel Wollner and Hamid Seyedsayamdost for helpful comments and discussions on earlier versions of this chapter. Earlier versions of this chapter were presented at the Meetings on Ethics and Political Philosophy in Braga, at the Association for Legal and Social Philosophy Annual Conference, at the Society for Applied Philosophy Annual Conference, at the Pavia Graduate Conference in Political Philosophy, and at the London-Berkeley Graduate Conference.

be permissible to infringe someone's rights to minimize the overall number of rights violations that occur. The basic idea behind the inviolability approach is that our moral rights to non-interference protect and give expression to our *inviolability*, which is an aspect of our moral status. The more stringent our rights to non-interference, the fewer the circumstance under which morality endorses our being harmed against our consent, and the higher our moral status. If we have a high moral status, we are morally important beings. To the extent that our rights are not permissibly infringeable for minimization's sake, we are thus morally important beings who are "worthy of respect, owed respect" (Kamm, 1996, p. 272). Crucially, our inviolability is not reduced when someone violates our rights, or harms us impermissibly. As an aspect of our moral status, our inviolability determines how we *should* be treated, and is for that reason not affected by how people *actually* treat us.

I shall consider two criticisms to the inviolability approach: a *saveability criticism* and an *enforceability criticism*. The saveability criticism accuses the inviolability approach of arbitrarily favouring the importance captured by our inviolability over the importance captured by another aspect of our moral status—our 'saveability'. According to the saveability criticism, inviolability and saveability are two aspects of our moral status that are directly opposed in minimization cases, and there is no reason to assume that inviolability is a more significant indicator of our moral importance than saveability (Kagan, 1991).

I show that the saveability criticism is well-founded. On at least one plausible interpretation of moral status and moral importance, it seems that our moral importance is not at stake in minimization cases: no matter how we decide these cases, our overall moral importance would seem to remain the same. We should therefore not appeal to considerations of moral importance to defend the idea that the trade-off between saveability and inviolability in minimization cases should at least sometimes be decided in favour of inviolability. I suggest that we should instead appeal to considerations of fittingness. On the *fittingness defence* of the inviolability approach, we have reason to believe that each of us is highly inviolable just in case this fits the kind of moral creature we are.

The enforceability criticism grants that our moral status should fit the kind of creature we are. It also grants that as rational beings who aspire to lead self-determined lives, it is apt that we should be highly inviolable, as being highly inviolable makes us *personally sovereign* by putting us in charge of our bodies and minds. But it contends that the fittingness defence of the inviolability approach depends on too narrow an understanding of personal sovereignty. More specifically, it holds that an appropriate conception of personal sovereignty cannot be limited to rights to non-interference, but has to include enforcement rights also. If this is correct, then our personal sovereignty increases not only with the stringency of our rights to non-interference, but also with our moral authority to defend ourselves against illegitimate interference. Otsuka (2011) claims that our authority to defend ourselves consists of and increases with enforcement rights we have against culpable attackers and against innocent third parties (the latter of whom it may sometimes be helpful to harm in order to fend off an attack). He then shows that on his conception of personal sovereignty, the fittingness defence of the inviolability approach is undermined.

I maintain that the enforceability criticism fails. I argue that while Otsuka is right to insist that there is an enforceability dimension to personal sovereignty, he is wrong to believe that it extends to enforcement rights we have against innocent third parties. This restores the fittingness defence of the inviolability approach.

The structure of this chapter is as follows. I first present the paradox of deontology (section 3.2). I then introduce an agent-focused approach to answering the paradox, and explain what makes this type of approach unsatisfactory (section 3.3). In sections 3.4 and 3.5, I discuss the victim-focused inviolability approach to answering the paradox. The saveability criticism is outlined in section 3.6. In section 3.7, I argue that on at least one plausible interpretation of moral status and moral importance, the saveability criticism is sound. In section 3.8, I argue that this should encourage us to shift our attention away from moral importance and towards fittingness. The enforceability criticism is introduced in section 3.9. Section 3.10 argues that the enforceability criticism fails. In section 3.11, I investigate what follows from the fact that as rational beings, we rightly care about having *effective* rights to non-interference. Section 3.12 concludes.

3.2 The Paradox of Deontology

Consider the following problem case:¹

The Murder Trolley Case (MTC). A vicious killer has tied five people to a trolley track and directed a runaway trolley towards them. You are standing on a bridge across the trolley track with a massive man standing next to you. You have the following options: you can either do nothing, in which case the trolley will hit and kill the five, or you can push the massive man from the bridge and onto the tracks, in which case the trolley will hit him before it reaches the five. This will crush and kill the massive man, but it will also bring the trolley to a halt before it reaches the five. Is it morally permissible to push the massive man?

Most subjects faced with this question do not think so.² They think that the massive man has a right not to be killed, a right that is not permissibly infringeable³ given the circumstances of the case. Can this view be justified? One might think that the massive man has a right not to be killed because it would be bad if he were killed. More generally, one might think that people have rights against being killed because killings are bad, and, for that reason, ought not to occur. Yet if this is correct, it seems that in the specific setup of the MTC, you should in fact be permitted to push the massive man. That way, you would be able to ensure that only one instead of five killings occur, thus keeping the badness of what is happening to a minimum.

At least in principle, the problem just raised applies to all of our moral rights to non-interference. Rights to non-interference protect us by making it impermissible to

¹ This case is a fleshed out version of Kamm's *Constraint Case* (Kamm, 1996, p. 207), following Thomson's 'footbridge version' of the *Trolley Case* (Thomson, 1976, pp. 207-8).

² See e.g. Hauser et al., 2007, p. 6. When the Trolley Case was introduced to them, the subjects in this study were not given any information as to how the five people on the trolley track had ended up there. They were simply told that "there are 5 people on the track" (ibid.). I agree with Thomson and Kamm that it does not make an intuitive difference to the permissibility of pushing the massive man whether the five were put on the track by a murderer, or innocently got stuck there by themselves. In both cases, it seems similarly impermissible to push the massive man. See Thomson, 1985, p. 1399, and Kamm, 1996, p. 207.

³ In this chapter, I follow Thomson in distinguishing between permissible and impermissible rights infringements, and in referring to the latter as rights violations. See Thomson, 1990, p. 122.

harm⁴ us against our consent. In doing this, rights⁵ act as “side constraints” (Nozick, 1974, p. 29)⁶ on action: from any agent’s set of feasible alternatives, they rule out as impermissible those actions which would violate others’ rights. Presumably, rights rule out actions as impermissible on grounds that these actions are bad and ought to be prevented. But if this presumption were correct, rights would have to follow a *minimization logic*: they would *always* and *necessarily* have to be permissibly infringeable for the sake of preventing more numerous violations of the same right. If they weren’t infringeable in this way, rights would risk defeating their own purpose. However, if our intuitions about cases such as the MTC are correct, rights do *not* in fact follow a minimization logic. Instead, it is at least sometimes wrong to infringe a right to keep a higher number of comparable rights from being violated.

How can we resolve this inconsistency between the presumed purpose of our rights to non-interference and our intuitive judgements about them? Should we dismiss our intuitions about cases such as the MTC, and accept that rights ought to follow a minimization logic? Or is there a way of understanding the purpose of rights that is compatible with our judgements about cases such as the MTC? In line with standard terminology, I will refer to this problem as the *paradox of deontology*.

3.3 The Agent-Focused Approach to Solving the Paradox of Deontology

Suppose you are contemplating whether or not to push the massive man in the MTC. If you push him, you minimize the killings that occur. This doubtlessly speaks in favour of pushing him. But from your point of view, this is not the only consideration

⁴ In this chapter, I use ‘rights to non-interference’ interchangeably with ‘rights against being harmed’. I understand that not all interferences that we have moral rights against may be characterizable as harmings, but for my purposes, this does not matter. In my examples, I focus on *serious harms* as a type of interference that we have stringent rights against, but the points I am making generalize to other types of unwanted interference.

⁵ In this chapter and unless indicated otherwise, I use the term ‘rights’ to refer to moral rights to non-interference. I understand such rights as Hohfeldian claim-rights that a right-holder has against other persons who have correlative duties towards him. For a discussion of the Hohfeldian analytical system, see e.g. Wenar, 2011, sec. 2.1.

⁶ Nozick was the first moral philosopher to draw attention to this problem. In this paragraph, I follow his way of presenting it. See Nozick, 1974, pp. 28-30.

that matters. What also matters is that if you push the massive man, *you thereby kill him*. From your perspective, the fact that you have to kill someone to minimize killings provides you with a strong reason against doing so. Why is that? As agents, we are prone to distinguish between killings that are attributable to ourselves, and those that we attribute to other agents. From our individual points of view, our own killings appear to us much more worrisome than the killings of others. But does this distinction capture no more than “self-indulgent squeamishness” from our part (Williams, 1973a, p. 102)? Or do we have reason to believe that each of us has a special responsibility for her own actions? Advocates of the agent-focused approach to solving the paradox of deontology think the latter. They grant that from an *impartial* point of view, everyone has reason to minimize the number of rights infringements that occur. But they argue that from an *individual* agent’s point of view, it matters *morally* whether the agent has to infringe a right to do so. More fundamentally, the agent-focused approach to solving the paradox of deontology is based on the idea that morality is a source not only of *agent-neutral reasons* that apply with equal strength to all moral agents, but also of *agent-relative reasons* that have force only from the points of view of particular agents (Nagel, 1986, p. 153).

If agent-relative reasons are to explain the non-minimization logic of moral rights, we need to understand what their moral significance consists in. Nagel (1979, p. 130) suggests that an agent who intentionally infringes her victim’s rights and harms her victim thereby “incorporates [...] evil into what [she does]: it is [her] creation and the reasons stemming from it are magnified from [her] point of view so that they tower over reasons stemming from greater evils that are more ‘distant’ because they do not fall within the range of intention.” Nagel’s idea is that the evil that an agent intends is thereby made worse or “magnified”⁷ from the agent’s point of view, so that the agent has a valid moral reason not to intend evil, even if that leads to a worse outcome, impartially considered (Nagel, 1986, pp. 181-5; Nagel, 1979, pp. 130-1).

Suppose this is right. It then seems plausible that agent-relative considerations will sometimes ground a *prerogative* to refrain from doing what would be best from an

⁷ Note that according to Nagel, the evil doesn’t just *appear* magnified from the agent’s point of view; “normative magnification” (Nagel, 1986, p. 184) *really does take place* from the agent’s point of view.

impartial point of view. For example, if in the MTC, the evil of pushing the massive man to his certain death is greatly magnified from your point of view, it may tower over the evil of the five people on the trolley track being killed by a vicious murderer. You may then not be morally *required* to kill the massive man. If killing him would be morally worse from your point of view, you may be released from a requirement to do so. But from what Nagel suggests, it does not become clear how agent-relative considerations could ground deontological *constraints*, or prohibitions to bring about the impartially best possible consequences (Scheffler, 1982, p. 22). In the MTC, as soon as we grant—as Nagel does—that from an impartial point of view, it would be *better* for you to push the large man, it becomes difficult to see why you should not be allowed to transcend your subjective perspective and act in accordance with an impartial perspective. Yet if the massive man’s right not to be killed constrains what you may permissibly do, you are in fact prohibited from doing what would be best from an impartial point of view. Given that Nagel does not provide an argument in favour of the view that agent-relative reasons should necessarily (or even just sometimes) take priority over agent-neutral ones, this is quite puzzling. In fact, Nagel (1986, p. 183) himself explicitly doubts that a compelling argument in favour of this view can be given—and no one seems to have risen to the challenge so far. This leaves the agent-focused approach to solving the paradox of deontology importantly incomplete.

Kamm (1996, pp. 242-3) has raised an additional objection to the agent-focused approach. She introduces a case that is structurally similar to an Amended Murder Trolley Case (AMTC) in which *you* are the one who in an evil mood strapped five people to a trolley track, and sent a runaway trolley towards them. Now you are standing on a bridge across the track, and you feel deep regret for your earlier actions. Alas, you cannot harmlessly undo what you have done. Your only way of stopping the trolley is to push a massive stranger (who happens to be standing next to you) onto the tracks, thus killing him. May you permissibly push the massive man? To Kamm (1996, p. 242), it is obvious that you may not. This seems right to me as well, and Judith Thomson (1985, p. 1399), who discusses a similar case, agrees. Kamm (1996, pp. 242-3) then goes on to show that on an agent-focused approach to solving the paradox, we cannot account for these judgements. To see Kamm’s point, suppose

that Nagel is right, and that the evil that an agent intends to bring about is magnified from the perspective of that agent. In the AMTC, this fact cannot provide you with a reason against pushing the massive man. For while it is true that *you* will have intentionally killed one person if you push the massive man, it is also true that if you do not push him, *you* will have intentionally killed five people. Put differently, even if we look at things from your point of view, the evil of killing the one cannot dwarf the evil of killing the five. More generally, as long as all the rights infringements in question are an agent's own, agent-relative considerations apply equally to all infringements, so that they cannot provide the agent with a reason against infringing a right to minimize overall infringements. It follows that in cases such as the AMTC, advocates of the agent-focused approach are forced to conclude that our intuitive judgements are unfounded. But this is unsatisfactory, as it makes the paradox of deontology reappear on a smaller scale. For now we can simply ask: do we *really* have to give up our intuitions about cases such as the AMTC? Or is there a way of understanding the purpose of rights that is compatible with our judgements about cases such as the AMTC?

3.4 The Inviolability Approach

Nagel's agent-focused approach to solving the paradox of deontology takes as its starting point the fact that insofar as we are *agents*, we can come up against other people's rights to non-interference, in which case these rights constrain what we may permissibly do. A victim-focused approach proceeds differently. Its primary interest lies with establishing what a right to non-interference does for the *person whose right it is*.

The victim-focused *inviolability approach*⁸ to solving the paradox of deontology takes as its starting point the observation that insofar as we are potential victims of a rights violation, our rights to non-interference make it impermissible for others to interfere with us in ways we do not approve of. Its core claim is then the following:

⁸ The idea that the concept of inviolability might be helpful to overcoming the paradox of deontology goes back to at least Nozick, but it is Kamm who has developed Nozick's suggestions into the fleshed out 'inviolability approach' presented here. See Nozick, 1974, p. 32, and Kamm, 1996, pp. 259-83.

by making it impermissible to interfere with us against our consent, rights to non-interference protect and give expression to our⁹ *inviolability* (Kamm, 1996, p. 271). In essence, our inviolability and our moral rights to non-interference are two sides of the same coin. The smaller the set of circumstances under which it is permissible to harm us against our consent, the more stringent our rights to non-interference, and the more inviolable we are. The stringency of our rights to non-interference thus determines the extent of our inviolability, and vice versa (Kamm, 1996, p. 272).

Inviolability forms part of the moral standing we enjoy as individuals, i.e. our *moral status*. An individual's moral status fixes how that individual should be treated—it delimits the kind of treatment the individual is owed. To respect an individual's moral status is to live up to the moral requirements that the individual's presence imposes on those around her. Inviolability as part of our moral status is a measure of our moral importance. The more inviolable we are, the more extensive the moral constraints that others have to observe when dealing with us, and, other things equal, the higher our moral importance. Kamm (1996, p. 273) puts the point as follows: "Individuals whose rights stand as a barrier to action are more potent individuals than they would be otherwise. There being rights and constraints with high thresholds is a mark that the person who has them is a stronger, more valuable type of thing [...]."

Importantly, it does not lower our inviolability when someone violates our rights, or harms us impermissibly. Nor does it promote our inviolability when someone prevents our rights from being violated. It is the details of our moral status that

⁹ When I talk about *our* inviolability in the context of this chapter, I refer to the inviolability that a *human person* enjoys. Similarly, by *our* moral status I mean the moral status a human person has. By a 'human person' I mean a human being who has the capacity for rational and moral agency. Further below, I will argue that an individual's morally relevant capacities determine the moral status that the individual enjoys. I will thus defend a version of what James Rachels (1990, p. 173) has termed *moral individualism*. Moral individualism is opposed to *speciesism*, i.e. the view that an individual's moral status is determined by the biological species of which the individual is a member. In the context of this chapter, I leave open whether we have good reason to treat all human beings the same, irrespective of whether they are also persons or not. I also leave open whether non-human persons should have the same moral status as human persons. For a defence of moral individualism, see Singer, 1980, and McMahan, 2005a. For a defence of the view that we have reason to treat all human beings the same even if their morally relevant capacities differ, see Scanlon, 1998, pp. 185-6. Liao (2010) suggests that all human beings may share the same morally relevant features.

determine how inviolable we are, and these details remain the same whether or not we are treated in accordance with them. As agents who come up against the rights of others, we are therefore simply asked to *conform* our actions to the inviolability of those around us by not violating their rights (Kamm, *ibid.*).

To illustrate these ideas with the MTC, consider the five people tied to the trolley track. If the vicious murderer kills the five, each of them is thereby *mistreated*—because the five have a certain moral status, it is wrong to kill them for one’s sadistic pleasure. If the murderer kills his victims, he therefore disregards their moral status. But to disregard someone’s moral status is not to lower it: we do not relax the moral rules by breaking them. In short, whether or not the five are killed, each of them is a morally significant being who ought not to be killed for fun.

Similarly to what goes for the murderer and the five, whether you push the massive man has no effect on his moral status. What does have an effect, however, is whether it is *permissible* to push him. Once we have settled on an answer to this question, we have settled a fact about his moral status, and, by extension, about the moral status of each of us (see footnote 9). If it is impermissible to push him—as our intuition tells us it is—the set of circumstances under which morality condemns our being harmed against our consent is larger than it would otherwise be, and we are more inviolable. Admittedly, if morality condemns pushing the massive man, then respecting his inviolability will come at the cost of the five people on the trolley track having their rights not to be killed violated. But this does not lower our inviolability, as it still holds true that the five ought not to have been killed: the murderer should not have tied them to the trolley track, and he should not have sent a runaway trolley towards them. If *you* must not *now* push the massive man, morality condemns the only available course of action that would save the five from being killed, but this does not mean that it sanctions their being killed. By analogy, the fact that you must not cheat on a test does not imply that it was acceptable not to study for it, or that there is nothing wrong with flunking it.

To recapitulate, the inviolability approach asserts that our moral rights to non-interference protect and give expression to our inviolability. Our inviolability increases with the stringency of our rights to non-interference, so that we are *more* inviolable

if our rights are *not* permissibly infringeable for minimization's sake. It follows that if rights protect our inviolability, there is nothing self-defeating about a right that is not permissibly infringeable for the sake of preventing more numerous violations of the same right.

As sketched, the inviolability approach provides a *possible* solution to the paradox of deontology, but it raises a number of questions as well. First, what gives rise to our moral status as highly inviolable individuals? Second, what is the point of having a moral status that condemns our being harmed, even when as a consequence of this, more of us may actually be harmed? Third, our inviolability seems high, but it does not seem maximal. If this is right, then what determines its limits? Only if there are satisfactory answers to these questions may the idea that rights protect and give expression to our inviolability lie at the core of a *compelling* solution to the paradox.

3.5 Inviolability and Moral Importance

The most elaborate defence of the inviolability approach has been put forward by Kamm. Even so, it is not entirely clear how Kamm means to—and whether she is able to—answer the questions that this approach raises. She provides a number of suggestive but sketchy comments that are not easily assembled into a coherent picture.

Regarding the source of our inviolability, Kamm (1996) suggests that it might derive from “certain properties [...] we have as individuals” (p. 273), most likely from our “having a rational will” (p. 276). Her hunch is that having a rational will endows each of us with moral worth, and that our inviolability captures this moral worth. If we are highly inviolable, there are hardly any circumstances under which morality endorses our being harmed. Whenever it condemns our being harmed, it expresses the fact that we have moral worth.

Kamm argues that being highly inviolable is good for us because it means that we are *morally important beings*. This is “in itself a benefit *to us*”, because it implies that we are “worthy of respect, owed respect” (Kamm, 1996, p. 272, original emphasis). Kamm adds that there are other benefits to being morally important as well, such as the fact that as important beings, “we are creatures whose interests as recipients of

such ordinary benefits as welfare are more worth serving” (Kamm, 1992b, p. 383). Regarding the coherence of a moral status that on the one hand condemns certain types of harming, yet on the other hand rules out minimizing instances of them, Kamm thus seems to propose that the point of such a status is to capture our moral worth and reveal our moral importance.

3.6 The Saveability Criticism of the Inviolability Approach

Can Kamm’s suggestions be fleshed out into satisfactory answers to the questions that the inviolability approach raises? Some comments by Shelly Kagan may make us doubtful.¹⁰ Kagan (1991, p. 919) agrees with Kamm that there is a sense in which each of us is morally more important if our inviolability is higher. But he also thinks that there is a sense in which each of us is morally more important if more may be done to save us from being harmed (*ibid.*). He then notes that “given the restriction that all people are to have the same [moral status]”, it is not clear which set of moral rules would reveal “us to be the most important sorts of creatures, the ones whose interests are most worth serving” (Kagan, 1991, p. 920).

Based on Kagan’s comments, we can construct a ‘saveability criticism’ of Kamm’s ideas. For the sake of argument, we first grant Kamm the following two claims:

1. The higher our moral status, the higher our moral importance;
2. It is appropriate that our moral status should reveal each of us to be a highly important being.

¹⁰ Kagan (1991, pp. 919-20) was the first to provide an outline of the saveability criticism. The relevant passage is however very brief, and is not in fact intended as a criticism of Kamm’s answer to the paradox of deontology. Rather, Kagan’s aim in that passage is to rebut Kamm’s contention that a consequentialist morality is unable to capture the high moral importance we have as individuals. Kagan’s cursory but suggestive comments are then taken up by Kamm (1996, pp. 275-7), who fleshes them out in order to try and counter them. In this chapter, I develop Kagan’s ideas even further, so that he might not want to subscribe to the saveability criticism as I present it here. The saveability criticism as I present it here should thus be understood as a criticism that I develop based on a few comments that can be found in Kagan, 1991, pp. 919-20.

We then argue, *contra* Kamm, that (1) and (2) do *not* support the claim that our rights to non-interference should not be permissibly infringeable for minimization's sake, as we are *not* morally more important beings if our rights are not infringeable that way.

The idea behind the saveability criticism can be explained with the help of the MTC. In the MTC, there are two moral alternatives: either it is permissible to push the massive man to save the five, or it is impermissible. If it is impermissible, then there are *fewer purposes for which any of us may be harmed against our consent*, and we are more inviolable. Other things equal, if we are more inviolable, then this means that we are morally more important. But if it is permissible to push the massive man, then *more may be done to save us from being harmed*, and we are more *saveable*.¹¹ Other things equal, being more saveable *also* reveals us to be morally more important beings. It follows that inviolability and saveability are two different aspects of our moral status that both contribute to our moral importance, but that are directly opposed in minimization cases such as the MTC. To know whether we are all things considered more important if it is impermissible to push the massive man, we thus need to know how to trade off the effects of higher inviolability against those of lower saveability, and vice versa.

Kamm's idea that we are morally more important if it is impermissible to infringe our rights for minimization's sake seems to simply *presuppose* that inviolability has a more significant impact on our moral importance than saveability does. But it is not clear why this should be so; at the very least, further argument is required. At core, the saveability criticism thus accuses Kamm of arbitrarily favouring the importance captured by inviolability over the importance captured by saveability in her attempt to answer the questions that the inviolability approach raises.¹²

¹¹ The term is Kamm's (1996, p. 275).

¹² Kasper Lippert-Rasmussen (1996) criticizes Kamm's suggestions about moral importance in essentially the same way as I have interpreted Kagan's comments here. Lippert-Rasmussen introduces an aspect of our moral status that he calls *unignorability*. Our unignorability is higher the fewer the circumstances under which others may permissibly allow us to be harmed (Lippert-Rasmussen, 1996, p. 340). He then argues that in minimization cases such as the MTC, inviolability and unignorability are directly opposed, so that Kamm's answers to the questions that the inviolability approach raises depend on the undefended assumption that "our status as inviolable [is] more important than our status as unignorable" (Lippert-Rasmussen, 1996, p. 341).

3.7 Moral Status, Moral Importance, and the Validity of the Saveability Criticism

In this section, I present plausible interpretations of the terms ‘moral status’ and ‘moral importance’ that render the saveability criticism valid. I also raise doubts that there is *any* plausible interpretation of the two terms on which the saveability criticism can successfully be repudiated. The problem has the form of a dilemma. Either we choose a conception of moral status that does not have a built-in preference for the importance captured by inviolability over the importance captured by saveability. If we take this first route, we arrive at the conclusion that human persons would not seem to be morally more important creatures if their rights are not permissibly infringeable for minimization’s sake. Or we choose a conception of moral status where inviolability-related importance takes precedence over saveability-related importance. If we take this second route, we are stuck with the hopeless task of non-circularly explaining how one kind of importance can be more important than another kind of importance.

3.7.1 Moral Status and Moral Importance

I shall conceptualize moral status so that an entity enjoys a *high moral status* when and because it is a *source of weighty moral reasons for action*.

How can an entity be a *source* of moral reasons for action? One way¹³ is for it to have a good, so that things can be in its interest (Goodpaster, 1978, pp. 318-9). It then becomes possible to act for the entity’s sake, and as moral agents, we may have reason to do so (ibid.; see also Kamm, 2007, pp. 228-9). We act for the sake of an entity if we:

¹³ An entity may also be a source of moral reasons, and may thus according to my definition have a moral status, without having a good (Kamm, 2007, pp. 228-9). The basic idea here is that there can be appropriate and inappropriate ways of treating an entity *given the kind of thing it is* that are not determined with reference to a good it may have. In the context of this chapter, I will not investigate this possibility further.

- Do something—perform an action, abstain from acting, put up with something—so as not to affect it *adversely*.
- Do something—perform an action, abstain from acting, put up with something—so as to affect it *beneficially*.

If an entity has a *high* moral status, it is a source of *weighty* moral reasons for action. This makes it *morally important*. According to Nozick (1989), something is important if it has “relational strength” (p. 179) or “impact beyond itself” (p. 171). If an entity is a source of weighty moral reasons, then the moral reasons to treat that entity in certain ways have the capacity to override other considerations, and the entity has “relational strength.” Moreover, the reasons to treat it in certain ways have the capacity to determine how a moral agent should all things considered act, and the entity does in that sense have “impact beyond itself.” The moral agent’s action may in turn impact on third parties as well.

If an entity has a higher moral status than other entities, we are required not to interfere with its good even if others (including ourselves) would benefit from doing so, and we are permitted (or possibly even required) to promote its good even if others (including ourselves) are affected adversely. In other words, if an entity has a higher moral status than other entities, the reasons to act for its sake outweigh those to act for the sake of other entities. To give an example, suppose it is permissible to torture a dog in order to save one person from enduring similarly painful torture. For the dog, this means that its moral status is not as high as it could be: the fact that torturing the dog affects it adversely does not always make it wrong to torture the dog. For the person, it means that her moral status is considerable, and that in at least one respect, her good matters more from a moral point of view than does the dog’s.

According to my understanding of the two terms, ‘moral importance’ is a more encompassing concept than ‘moral status’. That is, while an entity is morally more important if it has a higher moral status, an entity may be morally important without having a moral status at all. Consider a child’s beloved teddy bear. The teddy bear has *moral importance*: *we have moral reason to treat it in certain ways, and we have*

moral reason to avoid treating it in other ways. As strangers, if we witness the teddy fall out of the child’s knapsack, we have reason to run after the child and hand it to him; as parents, we have reason not to use the teddy bear as a sponge when washing the family car. But even though the teddy bear is morally important, it does not have a moral status, as it is not itself a *source* of moral reasons. We ought to return the teddy to its owner, and we ought not to abuse it as a sponge for reasons that have their source in its *owner*, or because losing the teddy or finding it damaged would deeply upset the *child*. The higher the moral importance of the soft toy—the more stringent the constraints and requirements that it imposes on us as moral agents—the weightier the moral reasons that the child is a source of, and the higher is the child’s moral status.¹⁴

An entity’s moral status must be distinguished from what happens to that entity (Kamm, 2007, p. 227). Suppose the moral status of dogs is such that it is impermissible to torture a dog for fun. If someone then tortures a dog for fun, that dog does not lose its status as a being that must not be so tortured. Neither can its status ensure that it will never be so treated. What its status does ensure is that *if* the dog is tortured for fun, then it is thereby *mistreated*: a wrong is being done.¹⁵ Put differently, the pleasure that a dog torturer would derive from torturing a dog is at most a weaker reason in favour¹⁶ of torturing the dog than the fact that it affects the tortured dog adversely is a reason against it. But this does not guarantee that a person who considers torturing a dog will give morally appropriate weights to her different reasons for action.

3.7.2 Do Groups Have a Moral Status?

It is usually assumed that only individuals have a moral status (Kamm in Voorhoeve, 2009, p. 36; Kamm, 2007, p. 254). But isn’t it perfectly coherent to suppose that an entity composed of several individuals can give rise to moral reasons as well, and

¹⁴ The example of the child should make it clear that our moral status is not exhaustively described by our inviolability and our saveability. The moral reasons not to harm us and to save us from being harmed are just *some* of the moral reasons that we are a source of.

¹⁵ See Kamm’s reference to Nagel in Kamm, 1996, p. 278, and her discussion of Scanlon in Kamm, 2007, p. 232.

¹⁶ It may be no reason at all, or indeed a reason against it.

can thus have a moral status (cf. Goodpaster, 1978, esp. p. 310)? Why shouldn't a *composite entity*, understood as a *group of individuals*, have a good?

Consider the following example. Suppose there is a certain species of bug that infects dogs with a lethal disease. Further suppose that it is impermissible to eradicate the entire bug species to prevent a limited number of individual dogs from dying. This implies that the good of any individual dog is not a sufficiently weighty reason to justify pushing the bug species in question over the brink of extinction. But now suppose that the bugs multiply and become extremely widespread, so that the whole dog species will go extinct if we do not get rid of all the bugs. Suppose that under these changed circumstances, it becomes permissible to eradicate the bug species to prevent the dog species from going extinct. Doesn't this mean that the dog species—an entity composed of an indefinite number of individuals—has a certain moral status?

It seems to me that it does, at least if we grant that we act for the sake of a species if we promote its survival. In other words, as long as a composite entity has a good that we have moral reason to respect or promote, that entity would seem to have a moral status. This seems to me straightforward. But there is a difficulty.

In a nutshell, the difficulty is this. It is not clear that a composite entity can genuinely have an *own* good, i.e. a good that is not merely derivative of the good of its parts. Think about the dog example again. If we save the dog species, is there really a sense in which it is good for the *species* if it continues to exist? Isn't it simply good for individual dogs to exist, and because no individual dogs would exist if the dog species didn't exist, it is good *that* the species should exist, though this is neither good nor bad *for the species itself*?¹⁷ I am not entirely sure what we should say about the case of a species. For our purposes, it suffices to note that at least sometimes—and in all of the cases that we will in the following be interested in—a group has a good only in a derivative way. To say that we do something for the sake of the group is then simply shorthand for saying that we do something for the sake of the members of that group. Consider the five people trapped on the trolley track in

¹⁷ It may also be good for human beings that dogs exist, or the dog species may have impersonal value. If so, this may add to the *moral importance* of the dog species and of individual dogs, but it is not relevant to settling the question whether the dog species has a moral status. See section 3.7.1 and footnote 13.

the MTC. As a group, they have no good beyond the good of the individual members. In fact, the reason why we think of the five as a group in the first place is because the circumstances of the case are such that their interests happen to be aligned, so that we benefit each of them if we benefit one of them.

When a composite entity does not have its own good, it is not itself a source of moral reasons, even if its individual members are. According to my definition of moral status, the composite entity then lacks a moral status, even if its individual members have one. But the composite entity may still be morally important: we may have moral reason to treat it in certain ways, and to avoid treating it in others. If its individual members have a moral status, the composite entity may be morally important partly¹⁸ because the moral reasons that its individual members are a source of combine into reasons to treat the group in certain ways.

3.7.3 The Weights-Based View of Moral Reasons

The saveability criticism builds on the idea that the moral importance of a group may derive from the moral status of its individual members, and it takes this idea one step further. More precisely, it presupposes that the *saveability of a group is a strictly increasing function of the saveability of its members*, so that with regard to minimization cases, the *saveability of an individual is a strictly decreasing function of the minimum size of the group for which it becomes permissible to infringe another individual's rights*. In this way, proponents of the saveability criticism are able to conclude that *each* of the five people on the trolley track is more saveable if it is permissible to push the massive man for the sake of the *group* of the five.

Here is how a proponent of the saveability criticism might try to explain and defend the idea that the saveability of a group is a strictly increasing function of the saveability of its members: “To say that it is permissible to push the massive man for the sake of the group is to speak imprecisely. The group does not have its own good, so we cannot act directly for its sake. We can benefit the group only by acting for the sake of its individual members. To say that it is permissible to push the massive man

¹⁸ I say ‘partly’ because the composite entity may have moral importance for reasons that have nothing to do with its individual members. See section 3.7.1 and footnote 13.

for the sake of the group is thus to say that it is permissible to push him for the sake of the individuals that make up the group. The fact that it would greatly benefit one person on the trolley track is a first reason in favour of pushing the fat man. Taken by itself, it is not a sufficiently weighty reason to outweigh the consideration that pushing the massive man affects him adversely. However, the fact that it would greatly benefit a second person is a second and equally strong reason in favour of pushing the fat man. And the fact that it would benefit a third (and a fourth, and a fifth) person is a third (and a fourth, and a fifth) reason in favour of pushing him. There are thus five similar reasons that all speak in favour of pushing the massive man. Suppose that *taken together*, they outweigh the consideration that pushing the massive man affects him adversely. This means that each individual reason is weightier than it would be if they did not outweigh that consideration. Now, by our definition of saveability, an entity becomes more saveable as *more may be done to save her*, or as the *reasons to save that entity from harm become weightier*. It follows that each of the five people on the trolley track—and, by extension, each of us—is more saveable if it is permissible to push the massive man.”

Let us grant for the time being that the saveability of a group is a strictly increasing function of the saveability of its individual members. By granting this, we implicitly take up a *weights-based view of moral reasons*: we assume that each moral reason has a weight that is *measurable on a cardinal scale*, so that the weights of different reasons can combine to outweigh opposing reasons. On the weights-based view of moral reasons, “the image of a balance scale immediately suggests itself: each reason falls like a coin with a certain weight into one or the other of the two pans in the scale; the greater the combined weight in a given pan, the more it tips the balance in its own favor” (Kagan, 1988, p. 23).

On the weights-based view of moral reasons, is the basic charge of the saveability criticism correct? Does Kamm seem to arbitrarily favour the moral importance captured by our inviolability over the importance captured by our saveability? On the face of it, it would seem so. This is easiest to see once we attach numerical weights to the different morally relevant considerations. In the MTC, the massive

man is affected adversely if you kill him. Let us assign¹⁹ an ‘inviolability weight’ of 10 to this consideration. The fact that pushing the massive man affects him adversely then counts with a weight of 10 against pushing him. But you also have to take into consideration that each of the five people on the trolley track is affected beneficially if you push the massive man. If we assign a ‘saveability weight’ of less than 2 to each of these benefits, it will be impermissible to push the massive man, as the saveability reasons of the five will then add up to less than 10.²⁰ With regard to MTC-like cases, our moral status can then be represented by the (*inviolability*, *saveability*) vector $(10, 2 - \epsilon)$. If we favour neither the importance captured by inviolability over the importance captured by saveability nor the other way around, it seems that we can simply add up the two weights for an overall moral importance—or for an overall weight of the reasons for which each of us is a source of—of just under 12. Next, suppose we assign a saveability weight of more than 2. It then becomes permissible to save the five, as their reasons add up to more than 10. If we think that it should be just barely permissible to push the massive man to save a *single* person, then we maximize our moral importance if we assign a weight of $10 - \epsilon$ to the saveability reason.²¹ With regard to MTC-like cases, our moral status can then be represented

¹⁹ In this chapter, I assume that a human person has a certain moral status, or that there is a certain way in which it is most appropriate to treat her. In that sense, I do not believe we are at liberty to assign whatever weights we please to status-based moral reasons. But insofar as it is the relative weights that matter from a deliberative point of view, 10 is as appropriate a choice as any other number.

²⁰ In this numerical example, I have chosen an unweighted sum as a simple example of a strictly increasing function, but the discussion generalizes. Apart from its simplicity, it seems to me that the functional form of an unweighted sum aptly captures the idea that the weights of similar individual saveability reasons combine to form a stronger group reason.

²¹ Why should we not make the saveability weight larger than the inviolability weight? In short, because doing so would lead to contradictory moral imperatives. Suppose the consideration ‘doing X would save one person from being killed’ would count with a weight of 12 in favour of doing X, whereas the consideration ‘doing X would kill one person’ would count with a weight of 10 against it. Now suppose that you could push the massive man onto the trolley track to save just one person on the track from being killed. Should you go ahead and push him? It would seem so—the balance of reasons speaks in favour of pushing him. But suppose that a bystander was observing you, and noticed that you were trying to shove the massive man over the railing. To the bystander, the moral reasons would then speak in favour of trying to stop you. Stopping you would save the massive man from being killed, and this weighs with a weight of 12 in favour of trying to stop you. Even if stopping you might mean killing the one person on the track, this counts with no more than 10 against stopping you. But if a second bystander observed how the first bystander was trying to stop you, the second bystander would all things considered have to try and stop the bystander from stopping you. . . and so on. Such a system of moral rules would not be properly action-guiding.

by the (*inviolability, saveability*) vector $(10, 10 - \epsilon)$, which seems to give us an overall moral importance of just under 20. This is substantially higher than an importance of just under 12. On this way of looking at things, our individual moral importance therefore seems highest if the saveability and inviolability weights of a given right to non-interference are matched, or if the consideration ‘Doing X would save a human person from having a specific right violated’ is just as weighty a reason in favour of performing an action X as the consideration ‘Doing X would infringe the same right’ is a consideration against performing X. But if the saveability and inviolability weights of a given right to non-interference are equally weighty, it is always permissible to infringe that right for minimization’s sake. Hence it would seem that—*contra* Kamm—we are morally *more* important creatures if our rights to non-interference are *always* permissibly infringeable for minimization’s sake.²²

But on closer inspection, things are much more complicated. For one thing, it is one of the core claims of the saveability criticism that in minimization cases, saveability and inviolability are *directly opposed*, so that if one of the two goes up, the other goes down. Yet in the numerical example just presented, it seems that I held the numerical weight of an inviolability reason fixed while increasing the numerical weight of a saveability reason. Does this mean that the weights-based view of moral reasons is unable to adequately capture the saveability criticism? Not necessarily. In deliberation, we weigh different reasons *against each other*, so that it is their *relative weights* that matter. In that sense, our inviolability is high and ‘inviolability reasons’ are properly considered weighty if they weigh *heavier* than reasons that may oppose them. Holding the numerical weight of an inviolability reason fixed while increasing the numerical weight of a potentially opposing reason therefore decreases the extent to which we are inviolable. It follows that we are in fact less inviolable if our moral status with regard to the MTC is $(10, 10 - \epsilon)$ instead of $(10, 2 - \epsilon)$. But if that is the case, is it still true that our individual moral importance is maximized if rights to non-interference are always permissibly infringeable for minimization’s sake? One might

²² If we assign an inviolability weight of infinity to the consideration ‘Doing X would kill a human person’, then our moral importance is maximal no matter what saveability weight is attached to the consideration ‘Doing X would save a human person from being killed’. But on the weights-based view of moral reasons, it could then under no circumstances whatsoever be morally obligatory to kill a human person. This is implausibly strong.

think so. To the extent that permissibly infringeable rights weaken our inviolability relative to our saveability, they strengthen our saveability relative to our inviolability. So these two effects would seem to balance.²³ But there are innumerable other reasons for action that may sometimes be directly opposed to saveability reasons, but that do not have their source in our good as human persons. As long as the weights of these other reasons are held constant, our overall moral importance would seem to increase if our saveability increases. To give an example, if the consideration that ‘Doing X would kill 1,000 dogs’ counts with a weight of 40 against doing X, then it is not permissible to save the five people on the trolley track by killing 1,000 dogs as long as the group saveability of the five amounts to less than 40. But once the group saveability of the five exceeds 40—as happens when we raise individual saveability from just under 2 to just under 10—it becomes permissible to kill 1,000 dogs to save the five.

But are the weights of moral reasons in fact transferable across contexts in this way? Intuitively, whether or not one may push the massive man in the MTC does not have much bearing on the number of dogs that one may sacrifice to save the five people on the trolley track. Rather, we seem to be dealing with two largely independent problems that need to be considered and decided on their own terms. Kagan and Kamm both think that this intuition is right. They argue that any plausible moral theory has to make room for the fact that the weight of a moral reason can vary with the sort of considerations that it is opposed to or that it interacts with (Kagan, 1988, esp. pp. 16-8 & 23; Kamm, 1986, p. 10; Kamm, 1996, pp. 51-6; see also Philips, 1987). For one way in which the weights of moral reasons do not remain constant across contexts, consider the famous ‘headache example’. If the weights of moral reasons were to remain constant across all contexts, and if—as seems plausible—the consideration that ‘Doing X would cure someone’s mild headache’ sometimes counts as a moral reason in favour of doing X, then the balance of reasons would at some point have to speak in favour of killing an innocent person to cure a sufficiently high number of mild headaches. But it seems that it is *always* wrong to kill an innocent person to cure a number of mild headaches, no matter how many mild headaches

²³ I discuss this further in section 3.7.4.

we are talking about. The reasons that are generated by sufficiently trivial harms are simply *not relevant* to the reasons generated by sufficiently substantial harms (Voorhoeve, 2014, pp. 65-6). Importantly, the problem in the headache case is not one of aggregation: we can recognize perfectly well that it is worse if more people have to suffer a mild headache than if fewer people have to do so. The point is simply that the consideration ‘Doing X would cure a mild headache’ should receive no weight against the consideration ‘Doing X would kill an innocent person’. In minimization cases, the saveability and inviolability reasons are always generated by equally substantial harms, so that the different reasons can safely be assumed to be relevant to each other. What the headache example shows is that their weights in minimization cases need not be exportable to other contexts.

In light of the fact that the weights of moral reasons can vary across contexts, we should give up the weights-based view of moral reasons. The weights-based view strongly suggests that the weights of moral reasons remain constant across contexts (cf. Kagan, 1988, p. 23). In this way, it creates the misleading impression that settling the weight of a moral reason in one context settles the weight of that reason in all other contexts as well. By creating this impression, it may lead us to construct arguments that are not in fact sound. Further above, I have suggested that matching the weights of inviolability and saveability reasons in minimization cases would seem to maximize our moral status. But this suggestion depends on the assumption that the weights of moral reasons remain constant across contexts, and does therefore not go through. In the next section, I introduce an alternative to the weights-based view of moral reasons, and investigate its implications for the saveability criticism.

3.7.4 From Weights to Trumps

Earlier, I have argued that we enjoy a high moral status if we are a source of weighty moral reasons for action. On the weights-based view of moral reasons, we think of weighty reasons as *heavy* reasons, and we picture practical deliberation as a process of placing reasons into one of two pans on a balance scale.

But we need not picture weighty or important reasons as heavy. Instead, we can think of them in a more general way as reasons that *trump* other reasons. If we do this, we best compare practical deliberation to a *card game*. On this view of things, an important moral reason functions like a card in a game in which it beats most other cards and card combinations. The trump-based view of moral reasons does *not* create the impression that the strength of moral reasons remains constant across different contexts. To give an example, the facts that (i) a blue card beats any red card, and (ii) a red card beats any green card reveal nothing about the strength of a blue card against a green card, just as the fact that two sevens of two different colours are worthless tells me nothing about the strength of two sevens of the same colour. To play a card game right, we have to learn its—possibly very complex—rules. This is not to suggest that card games are always based on entirely arbitrary rules. There may well be a logic to the rules of any given card game, and this logic may help us grasp its rules more quickly, or infer some of its rules without being told what they are.

On the trump-based view of moral reasons, does the saveability criticism still go through? Do we arbitrarily favour the importance captured by inviolability over the importance captured by saveability if we claim that our moral status reveals us to be more important creatures the more inviolable we are? In a nutshell, the answer seems to be yes, but it will take us some time to get there.

First, there appears to be the following problem to the saveability criticism on the trump-based view. If the saveability reasons of the five people on the trolley track function like trumps, their strength need not combine. Depending on the nature of the game, five cards that are individually strong may jointly have no more force than taken by themselves. Alternatively, five individually very weak cards could jointly trump most other cards and card combinations. Hence if we should think of moral reasons as trumps, then it seems wrong to assume that we can infer anything about the moral importance of an individual member of a group from the moral importance of the group. Consider the following analogy. You introduce me to a game, and I

ask you about the strength of an individual card. Suppose you tell me: “If you have this card together with four other cards of the same colour, it beats combinations X, Y, and Z.” Though you have given me information that is relevant to the game, you have not really answered my question. From what you have told me, I cannot infer anything²⁴ about the card’s individual strength.

It seems to me that this problem to the saveability criticism can be countered. It is true that on the trump-based view of moral reasons, we cannot infer anything about the strength of an individual saveability reason from the strength of a group saveability reason *in the absence of further argument*. But we have further arguments at our disposal. In section 3.7.3, I *arrived* at the weights-based view from what it is *plausible to assume* about moral reasons in minimization cases. More specifically, I argued that in the MTC, we have moral reason to act for the sake of the group of the five to the extent that we have moral reason to act for the sake of the first of the five, the second of the five. . . and the fifth of the five. The five people on the trolley track are furthermore indistinguishable from each other, so that it is perfectly reasonable to assume that the strengths of the reasons that they are a source of are all equal. In the MTC—and in minimization cases more generally—it thus seems plausible to assume that individual saveability reasons combine to form a weightier group saveability reason, and that we can draw inferences about the strength of individual saveability reasons from the strength of the group saveability reason. To return to the card game analogy: in a card game, the strength of a given card combination *need not* derive from the strength of the cards that make up the combination, but the rules of the game can specify conditions under which it does.

On the trump-based view, it is correct that inviolability and saveability are directly opposed in minimization cases, so that if one of the two aspects of our moral status goes up, the other goes down. On my definition of moral status, our inviolability increases as the ‘inviolability reasons’ which we give rise to become weightier, and our saveability increases as the ‘saveability reasons’ which we give rise to become weightier. On the trump-based view of moral reasons, a reason becomes weightier

²⁴ This may be a slight exaggeration. We can expect that taken by itself, the card does not beat more combinations than it beats together with other cards.

as the constellations of reasons in which it beats the reasons that oppose it become more numerous. In each minimization case, if it is permissible to infringe a right for minimization's sake, this means that there is a constellation of reasons where saveability reasons beat opposing reasons. By contrast, if it is impermissible to infringe a right for minimization's sake, this means that there is a constellation of reasons where inviolability reasons beat opposing reasons. In short, either inviolability reasons trump saveability reasons or the other way around, so that either inviolability gains at the expense of saveability or the other way around.

Can we say that one of the two aspects of our moral status gains more than the other loses in minimization cases? I don't think so. Suppose someone argued as follows: "Surely some constellations are more significant than others. To return to the card game analogy, surely it is much more significant if a card beats four aces than if it beats a jack. But if not all constellations are equally significant, then we cannot conclude that in minimization cases, saveability gains *as much* as inviolability would lose, and vice versa." This argument is simply confused. Each minimization case describes *exactly one constellation of reasons*. Even if not all constellations are equally significant, any *given* constellation is of a *given* significance, and is 'won' either by saveability or by inviolability reasons.

But there is a better argument in favour of the view that one aspect of our moral status gains more than the other loses in minimization cases. Consider the following line of thought: "It seems to me that in minimization cases, inviolability gains *more* than saveability loses if inviolability considerations trump saveability considerations. Here's why. Suppose that doing X would kill one person, but would save one hundred people from being killed. This means that there is a constellation of 101 status-based reasons: one inviolability reason against doing X, and one hundred saveability reasons in its favour. Surely there is somehow much more at stake for the inviolability reason? If we think of reasons in terms of cards in a card game, surely the one card would be *extraordinarily strong* if it were to beat one hundred cards, whereas none of the one hundred would necessarily be very strong if their combined strength were to beat just one other card!" I believe that this argument is confused also, though in a somewhat subtler way. To show that the argument does not work, we can

just flip it on its head and observe that the one hundred cards would seem to be *extraordinarily weak* if just one other card were to beat them, whereas the single card would not necessarily be very weak if one hundred cards were to beat it. In other words, if one ‘inviolability card’ beats one hundred ‘saveability cards’, the effects on both inviolability and saveability would seem *similarly drastic*, whereas if one hundred saveability cards beat one inviolability card, the effects on both inviolability and saveability would seem *similarly limited*.

A third argument that someone might attempt is the following: “Even if the weights of reasons can vary across contexts, it seems to me that inviolability reasons are more exportable to other contexts, and are in that sense stabler, than saveability reasons. I grant that in any given minimization case, inviolability can be increased only to the extent that we decrease saveability. But because a change in inviolability translates to more contexts than a change in saveability does, our overall moral status would seem to increase if we raise inviolability at the expense of saveability in any given minimization case. Suppose we conclude that it is impermissible to kill one person in a given way to save 100 other people from being killed in the same way. Given the logic of our morality, we can with confidence say that this implies the following:

- It would also be impermissible to kill the one to save 99, or 98, or 97...or 1 other person from being killed;
- It would also be impermissible to kill the one to save 100 or fewer beings with a moral status lower than that of a human person from being killed or from a harm less bad than death;
- It would also be impermissible to kill the one to save 100 or fewer human persons from a harm less bad than death.

In short, high inviolability in one minimization case can safely be assumed to spread to a number of other contexts.” I don’t think this argument is successful, either, as saveability implications seem to spread similarly to inviolability implications. If it is permissible to kill one person in a given way to save one hundred from being killed, this would for example imply that:

- It would also be permissible to kill the one to save 101, or 102, or 103. . . persons;
- It would also be permissible to kill a being with a moral status lower than that of a human person in the given way to save 100 or more human persons from being killed;
- It would also be permissible to inflict a harm less bad than death on the one to save 100 or more from being killed.

In sum, on the trump-based view of moral reasons, inviolability and saveability are two aspects of our moral status that are directly opposed in minimization cases. Depending on whether it is permissible to infringe a right for minimization's sake, either an inviolability reason is weightier than a combination of saveability reasons, or the other way around, so that either inviolability is strengthened at the expense of saveability, or the other way around. As nothing we have said comes close to suggesting that the two effects do *not* balance, our overall moral status—the overall strength of the reasons that we are a source of—would seem to remain constant no matter how this trade-off is decided.

If all of this is right, then it is a well-founded charge that Kamm arbitrarily favours the moral importance captured by inviolability over the importance captured by saveability when she argues that our moral status is higher and we are morally more important beings to the extent that our rights to non-interference are not permissibly infringeable for minimization's sake. On at least *one* plausible interpretation of moral status, it is difficult to see how our moral importance (as captured by our moral status) could be higher or lower depending on what it is permissible to do in minimization cases.

But even if the conception of moral status presented in this chapter is a plausible conception, there might be *other* plausible conceptions on which Kamm's claims are indubitably correct. Kamm stresses that our moral status is something that we possess *as individuals*. If our moral status is higher, this means that *each of us considered independently of others* is morally more important. Does my conception of moral status adequately capture this idea? Suppose someone argued as follows: “If we understand moral status properly individualistically, then our saveability is

not actually at stake in minimization cases. Think about the MTC, and consider how it affects our inviolability if it is impermissible to push the massive man. If it is impermissible to push him, our inviolability will be higher, as the consideration ‘It would kill one of us’ trumps an opposing combination of reasons. As for saveability, the relevant question is the following: would it be permissible to push the massive man to his death if there were only *one* person stuck on the trolley track? Most of us do not think so, even if we think it permissible to push him for the sake of the five. That is, we tend to believe that if it is permissible to push the massive man for the sake of the five, this is the case because the five *outnumber* him. But on a properly individualistic conception of moral status, our saveability increases only if a person’s *individual* good is sufficient to ground the permissibility of harming others. If the consideration ‘It would save one of us from being killed’ does not trump an opposing consideration by itself, our saveability does not increase. It follows that our saveability remains fixed if we hold constant what may be done for one individual person while varying what may be done for a group of five. In the card game analogy, this point is actually very easy to see. Either a single card is decisive when played by itself, or it is not. There is no way in which a single card becomes *more decisive as a single card* if it is *not* decisive as a single card but beats other cards as part of a set of five.”

In response to this criticism, I can only reiterate what I said earlier. If it is permissible to save the five, this would seem to be the case because we have reason to act for the sake of the first of the five, the second of the five... and the fifth of the five. In the card game analogy, if we *know* that a set of five equal cards derives its strength from the strength of the cards that make up the set, there *is* a sense in which each individual card becomes stronger as the set becomes more decisive. In essence, a decisive set is evidence for the fact that the individual cards that make up the set are strong.

One might also reject the conception of moral status presented in this chapter by simply insisting that the moral importance captured by inviolability ought to be prioritized over the moral importance captured by saveability in determining how high our moral status is. In some passages, Kamm seems to be doing just that, e.g. when she asserts that higher saveability gained at the cost of lower inviolability means a lower moral status because it implies an “intrinsically less elevated conception of ourselves” (Kamm, 1996, p. 278). The problem with this sort of claim is obvious: in the absence of an independent argument in its favour, it begs the question against the saveability criticism.

It is possible that Kamm tries to sketch an independent argument in favour of the view that inviolability-related importance is more significant than saveability-related importance. Consider the following passage (Kamm, 1996, pp. 276-7):

“Suppose the high status of persons, making it important to save them, was held to stem from some other property than their inviolability. Then, to say that their status would go down if they were violable (in certain ways) even if they were more saveable is to claim that the properties that underlie one’s inviolability are more important than those that give rise to any other significance one has. Might it be to claim that having a rational will, whose consent we must seek when interfering with what a person has [...], will give a person higher status than being a complex, feeling creature who cares about whether it lives or dies?”

But I don’t think that what Kamm says in this passage can be turned into a valid argument in favour of the view that inviolability-related importance is somehow more significant than saveability-related importance. The passage seems to presuppose that we are morally important not because and insofar as others ought to treat us in certain ways, but that others ought to treat us in certain ways because and insofar as we are morally important. In this way, the passage blurs the distinction between being morally important and having moral worth.

To see what I mean, consider the following *museum analogy*. Suppose there are two paintings hanging side by side in a museum. Visitors are not allowed to touch the paintings, and they must not take flash pictures of them. On my understanding of importance, this makes both paintings *equally important*, as *similarly stringent rules of conduct apply with respect to each of them*. But now suppose that painting *A* is much more valuable than painting *B*. On my understanding of the terms, it is reasonable to assume that a more valuable painting will be protected by more stringent rules, and will therefore be more important. But to the extent that paintings *A* and *B* are protected by similarly stringent rules, they are on my understanding of the term similarly important.

Translated to the museum analogy, Kamm's argument seems to be the following. First, she refers to the more valuable painting *A* (i.e. to our property of having a rational will) as 'more important', thus blurring the distinction between importance and value (i.e. moral worth). Second, she suggests that the rules that protect the more valuable painting (i.e. the moral reasons that are 'inviolability reasons') are somehow more significant *simply because* they protect a more valuable object. But this seems to me both confusing and conceptually confused. Whatever rules are put in place to protect the two paintings, it should remain a conceptual possibility that the rules which protect the less valuable painting *B* are more stringent than those which protect the more valuable painting *A*, so that they make *B* more important than *A*, and are in that sense more significant than the rules protecting *A*. To return to the actual problem, even if the property that gives rise to our saveability is of lower moral worth than the property that gives rise to our inviolability, this should not automatically imply that our overall moral importance increases if inviolability gains at the expense of saveability. Rather, the notions of 'moral importance' and 'moral worth' should be kept apart. While the latter should be thought of as a factor that determines our moral rules, the former should be thought of as a consequence or an implication of them.

But whether or not the passage just quoted can be turned into a valid argument, it proposes an interesting way of thinking about moral status. The passage suggests that we may have *different morally relevant properties that give rise to different aspects of*

our moral status. This seems to me an intuitively plausible and interesting way of looking at things. Also, once we think about moral status in this way, we do not need to dwell on considerations of moral importance to determine how different aspects of our moral status ought to be traded off against each other. Instead we can ask what sort of moral status would be appropriate for us given our morally relevant properties.

In the next section, I pursue this line of thought further, and I argue that it allows us to come up with a promising *fittingness defence* of the inviolability approach to solving the paradox of deontology.

3.8 A Moral Status That Fits Us

In the last section, I have argued against a *moral importance* defence of the inviolability approach. But if an appeal to moral importance does not help justify the non-minimization logic of our moral rights to non-interference, then the question arises anew: what reason do we have to think of ourselves as highly inviolable in minimization cases? What is the point of a set of moral rules that on the one hand condemns certain types of harming, yet on the other hand does not allow us to minimize instances of them?

Kamm (1996, p. 273) suggests that “it is certain properties [...] that we have as individuals that would account for our status.” So far, I have argued that it is unhelpful to understand this claim in terms of moral importance, or to think that certain properties endow us with moral worth, and that our moral status should capture this worth by making us morally highly important. But it is not necessary to understand the relation between “certain properties we have as individuals” and our moral status in terms of moral importance. Instead, we can understand it in terms of *fittingness*. According to this second understanding, certain properties we have as individuals make some ways of treating us and of deliberating about us morally more appropriate than others (Darwall, 1977, p. 39; Velleman, 2005, pp. 91-2 & 103-4; Quinn, 1989a, p. 312). The primary function of our moral status is then not to express our high moral importance (though it may do that also), but to capture the ways in which it is most appropriate to deliberate about as well as treat us,

given our morally relevant properties. On this second understanding, the proper trade-off between inviolability and saveability is determined not by asking “what combination of the two represents high status” (Kamm, 1996, p. 276), but by asking what combination of the two best fits our morally relevant properties. On this way of thinking about our moral status, the inviolability approach to solving the paradox of deontology is vindicated if being highly inviolable fits well with the kind of moral creature we are.

3.8.1 Inviolability and Our Capacity for Rational Agency

A long line of thinkers inspired by Kant argue that our stringent rights to non-interference derive from and are justified by our capacity for rational agency.²⁵ Nozick (1974, pp. 49-50) proposes that the distinguishing feature of a rational being is its capacity to govern itself in the light of reasons. He argues that a rational being is “a being able to formulate long-term plans for its life, able to consider and decide on the basis of abstract considerations or principles it formulates to itself and hence not merely the plaything of immediate stimuli, a being that limits its own behavior in accordance with some principles or picture it has of what an appropriate life is for itself and others [...]” (Nozick, 1974, p. 49). This suggests that as rational beings, if our lives are to be meaningful and valuable, we need some degree of control over them; we need to be able to live them following our own plans, and in accordance with our own ideas of the good (see also Rawls, 1971a, pp. 126-30). Warren Quinn argues that stringent rights to non-interference *enable* us to lead such self-determined lives, and that it fits the kind of moral creature we are that we should be *entitled* to lead such lives:

²⁵ Kamm (1996, pp. 276-7) agrees that our “having a rational will” is likely to be the source of our high inviolability. According to my way of looking at things, our capacity for rational agency gives rise to our inviolability largely because it fits the rational aspect of our human nature that we should have primary say over what may be done to us. This is consistent with what Kamm says, but her focus is on the *high moral worth* that having a rational will endows us with. That is, Kamm seems to think that inviolability considerations should usually win out against saveability considerations not because it is inappropriate to interfere with a rational being against its consent, but because our inviolability derives from an especially valuable feature of our moral nature.

“A person is constituted by his body and mind. They are parts or aspects of him. For that very reason, it is fitting that he have primary say over what may be done to them—not because such an arrangement best promotes overall human welfare, but because any arrangement that denied him that say would be a grave indignity. In giving him this authority, morality recognizes his existence as an individual with ends of his own—an independent *being*. Since that is what he is, he deserves this recognition. Were morality to withhold it, were it to allow us to kill or injure him whenever that would be collectively best, it would picture him not as a being in his own right but as a cell in the collective whole.” (Quinn, 1989a, p. 309, original emphasis, footnotes omitted)

Succinctly put, the *fittingness defence* of the inviolability approach runs as follows. High inviolability is an important precondition for leading a self-determined life. If we are highly inviolable, we have a right to decide for ourselves what should happen to us; no one else may interfere with our bodies and minds against our consent. Figuratively speaking, high inviolability turns each of us into a *sovereign ruler over his or her own territory*. This *personal sovereignty* is something that it is right to accord us and that we value independently of its expected and actual effects on the general welfare. As rational beings, we care about deciding things for ourselves, and it is appropriate that we should be entitled to do so, even if some other arrangement would better promote the general welfare. But if this is right—as I suppose it is—shouldn’t we be maximally inviolable? Shouldn’t it always be impermissible to interfere with us against our consent?

No, it should not, mostly because rationality is not the only morally relevant aspect of human nature (Kamm, 1996, p. 277). Another relevant aspect is *sentience*, or the fact that each human person is “a complex, feeling creature who cares about whether it lives or dies” (ibid.). It is plausible that our sentience grounds a moral concern with what *actually happens to us*, thus underpinning utilitarian ideas according to which

we should try and do what best promotes the general welfare.²⁶ Conflicts between our rights to non-interference and between bringing about what a utilitarian would consider the ‘best possible outcome’ can then be understood to reflect an underlying conflict between the moral implications of our rationality and our sentience—the former giving rise to the non-consequential value of personal sovereignty, the latter grounding a consequentialist concern with our welfare. Depending on the circumstances of the case, either side may prove to be more weighty. In the MTC, if it is in fact impermissible to push the massive man, personal sovereignty ‘wins out’ against welfare considerations, in the sense that from a moral point of view, it is more important that the massive man ought to have the final say over what should happen to him than it is to enable a net saving of four lives. Once we regard the MTC as such a conflict of values, we can see that there is nothing puzzling (or possibly even self-defeating) about resolving the conflict in favour of personal sovereignty.

3.9 The Enforceability Criticism

Once we employ the idea of personal sovereignty to create a connection between our capacity for rational agency and the inviolability that our rights to non-interference secure, we are able to give plausible answers to a number of questions that the inviolability approach raises. This suggests that the inviolability approach can be fleshed out into a satisfactory solution to the paradox of deontology—all we have to do is shift our attention from the *moral importance* that high inviolability endows us with to the *fittingness* of being highly inviolable. But the fittingness defence of the inviolability approach has been criticized as well.

In the previous section, I argued that high inviolability puts us in charge of our lives by making us sovereign over the territory of our bodies and minds. This suggests that ‘inviolability’ and ‘personal sovereignty’ are more or less interchangeable terms. But do inviolability and personal sovereignty come to the same thing? Otsuka (2011, pp. 51-4) argues that the latter should properly be thought of as a more

²⁶ This fits well with Nozick’s (1974, p. 39) idea that we should base our moral principles on “utilitarianism for animals; Kantianism for people”. If their sentience suffices to morally characterize non-human animals, it may be appropriate to treat them in accordance with utilitarian principles.

encompassing concept than the former. More specifically, he claims that personal sovereignty includes not only rights to non-interference, *but rights to enforce such rights as well* (Otsuka, 2011, p. 54).²⁷

Otsuka's claim seems conceptually correct. Suppose I harass another person by shoving him around and punching him, thus disrespecting his rights to non-interference. Now suppose that the moral rules that governed us were such that my victim—let's call him Bob—had no right to fight back. If under such circumstances other people came up to me and told me that I ought to stop harassing Bob, then they could appeal to Bob's inviolability, but not to his sovereignty. They could coherently argue that it is wrong to harass Bob because he is not violable in this way; they could point out that Bob is a morally significant being who is owed respectful and considerate treatment. But they would be confused if they claimed that I was wrong to harass Bob because of the *sovereignty* he enjoyed over the territory that is his body and mind. Unlike inviolability, personal sovereignty at least partly consists of a moral authority to fend off intruders. If we were all like beautiful flowers—beings that it is wrong to violate, but that lack all ability to enforce their inviolability—it would seem a category mistake to refer to us as 'sovereign'. In other words, if an entity lacks the capacity for self-defence, then it also lacks the capacity for sovereignty, and only if we may rightfully use our capacity for self-defence are we actually sovereign. We should therefore agree with Otsuka that sovereignty includes rights to non-interference as well as enforcement rights.²⁸

²⁷ Note that according to the Will Theory of Rights, you only have a certain right to non-interference if you *also* have a power to choose whether to waive or enforce that right. For a Will Theorist, rights to non-interference are thus not complete without enforcement rights. See e.g. Cruft, 2004, p. 367. In this chapter, I do not want to enter the debate between Will and Interest Theorists of Rights. Since I agree with Otsuka that enforcement rights play an important part in our morality, the points I make in this chapter are consistent with different versions of both the Will and the Interest Theory.

²⁸ Otsuka does not clarify whether he understands enforcement rights as mere Hohfeldian liberties or as Hohfeldian claim-rights. Given that he insists on the possibility of enforcement rights held against uninvolved third parties, I suspect he understands them as mere Hohfeldian liberties. In the remainder of this chapter, I will argue that properly construed, our enforcement rights consist of Hohfeldian claim-rights we hold against liable attackers. For a discussion of the Hohfeldian analytical system, see Wenar, 2011, sec. 2.1.

If personal sovereignty consists of rights to non-interference as well as enforcement rights, our sovereignty increases not only with our inviolability, but also with our moral authority to enforce our inviolability. Otsuka proposes we construe the enforceability dimension of personal sovereignty as follows (cf. Table 1). First, our authority to defend ourselves consists of and increases with the enforcement rights we have against those who threaten to illegitimately interfere with us (Otsuka, 2011, p. 53). Second, it also consists of and increases with the rights we have against innocent bystanders (ibid.). Otsuka’s idea is that if we may permissibly harm an innocent bystander to protect ourselves against illegitimate interference, then the enforceability dimension of our sovereignty is strengthened because we have more means available with which we may enforce our inviolability. Lastly, Otsuka extends the enforceability dimension of personal sovereignty to others intervening on our behalf (Otsuka, 2011, p. 54). That is, he argues that our inviolability becomes more enforceable as others have more extensive rights to defend us against illegitimate interference, either by being permitted to interfere with the attacker, or by being permitted to use innocent bystanders to shield us from the attack (ibid.).

		Rights held against...	
		The attacker	Third parties
Rights held by ...	Victim	Increase overall sovereignty	Increase overall sovereignty subject to a proportionality constraint
	Third parties	Increase overall sovereignty	Increase overall sovereignty subject to a proportionality constraint

Table 1: The Enforceability Dimension According to Otsuka

Does Otsuka think that our personal sovereignty increases whenever its enforceability dimension is strengthened? Not quite. According to Otsuka, enforcement rights held against innocent third parties strengthen the enforceability dimension of personal sovereignty, but they weaken its inviolability dimension. Only if they

strengthen its enforceability dimension more than they weaken its inviolability dimension do they make us *overall* more sovereign. Put differently, enforcement rights held against third parties strengthen our overall sovereignty only subject to a *proportionality constraint* (Otsuka, 2011, p. 54).

With the help of the MTC, we can get a better grip on the details of Otsuka's conceptualization of personal sovereignty. To Otsuka, if it is permissible to push the massive man, then there is a sense in which our sovereignty increases even though our inviolability suffers. More specifically, if you are permitted to intervene on behalf of the five people trapped on the trolley track, you are authorized to enforce their rights not to be killed via the means of harming the massive man, which strengthens the enforceability dimension of our sovereignty.

But what happens to our sovereignty all things considered? How are the negative effects on inviolability to be traded off against the positive effects on enforceability? Otsuka claims that the inviolability dimension of personal sovereignty is slightly weightier than its enforceability dimension, so that when "other things [are] equal", inviolability "makes a greater contribution to our personal sovereignty" than enforceability (Otsuka, 2011, p. 54). His idea is that we are more sovereign if it is impermissible to kill one innocent bystander in order to enforce one other person's right not to be killed. But once "other things are not equal"—or once a fairly minimal principle of proportionality is respected—Otsuka thinks that a strengthening of the enforceability dimension can result in an increase in personal sovereignty even if that strengthening comes at the cost of lower inviolability (*ibid.*). By way of example, he thinks that the overall effect on our sovereignty would be positive if it became permissible to "non-lethally maim one innocent bystander" to enforce another person's right not to be killed (*ibid.*). Similarly, he would seem to think that the overall effect on our sovereignty would be positive if it became permissible to kill one innocent bystander to enforce *several* other people's rights not to be killed.²⁹

²⁹ Otsuka does not explicitly state this, but given that he wants to challenge the inviolability approach to solving to the paradox of deontology, he needs to be willing to commit to it.

In sum, Otsuka’s idea is that there are two dimensions to personal sovereignty, an inviolability dimension and an enforceability dimension. The two can come into conflict, in which case we can either increase inviolability at the cost of lower enforceability, or vice versa. To Otsuka, we maximize overall personal sovereignty if we resolve such conflicts in favour of whichever dimension makes the weightier contribution to personal sovereignty given the details of the case.

If we accept Otsuka’s conceptualization of personal sovereignty, then the fittingness defence of the inviolability approach is undermined. The fittingness defence depends on the assumption that it *strengthens* personal sovereignty—and, in that sense, fits our capacity and aspiration to lead a self-determined life—if our rights to non-interference do not follow a minimization logic. But if Otsuka’s ideas are correct, it *weakens* our sovereignty if our rights do not follow a minimization logic, as the negative effects this has on the enforceability of our rights outweighs the positive effects it has on our inviolability.

3.10 Why the Enforceability Criticism Fails

In this section, I argue that the enforceability criticism fails. While Otsuka is right to insist on an enforceability dimension to personal sovereignty, he goes wrong in extending that dimension to enforcement rights held against innocent third parties. Once this mistake is corrected, the enforceability dimension of personal sovereignty no longer undermines the fittingness defence of the inviolability approach.

To see why the enforceability dimension of personal sovereignty should not extend to enforcement rights against innocent third parties, we need a better understanding of the concept of personal sovereignty. Personal sovereignty is but a borrowed concept; we are speaking metaphorically when we refer to a person’s body and mind as their

'territory'. Fundamentally, 'sovereignty' is an attribute of *states*, not of persons.³⁰ It thus seems appropriate to get a better grip on the concept by inquiring into its meaning and implications in political philosophy.

According to Dan Philpott (2010, sec. 1), a sovereign state is a state that exercises "supreme authority" within the borders of its territory, "but also, by definition, with respect to outsiders, who may not interfere with the sovereign's governance." Philpott notes that "quintessential modern sovereignty" is absolute, so that the state's authority extends "to all matters within [the state's] territory, unconditionally" (*ibid.*). This fits well both with the general idea that there is an inviolability dimension to personal sovereignty, and with the more specific claim that our personal sovereignty increases with the stringency of our rights to non-interference. But what about the idea that enforcement rights form an integral part of sovereignty? In a *system* of sovereign states, does each state have a right to enforce its authority against other states that disregard it?

Hugo Grotius was the first political philosopher to provide a systematic account of the natural law between sovereign states (Rabkin, 2005, pp. 74-8). In *De Jure Belli ac Pacis* (published in 1625), Grotius argues that it is legitimate for a sovereign state to wage war against a state that has attacked it. However, the victim state must not involve neutral states in its war effort if they do not agree to being involved, as neutral states may "rightly look to their own interests and stay on the sidelines" (Rabkin, 2005, p. 77). It seems that all commentators, from Grotius onwards, agreed that victim states have certain enforcement rights against their attackers, but none against uninvolved third parties (Rabkin, 2005, p. 95; see also Walzer, 1977, pp. 233 & 236-7). Why is that so? In a nutshell, because enforcement rights against attackers, but not enforcement rights against third party states, are compatible with

³⁰ Michael Walzer argues that to work out a theory of aggression between states, we need to be able to rely upon a "domestic analogy", so that we can deduce the rights of states via "analogical reasoning" from the rights we ascribe to individuals. See Walzer, 1977, p. 58. Walzer might thus disagree with my assessment that 'sovereignty' is fundamentally an attribute of states, and it is possible that Otsuka would side with Walzer. For a discussion of personal sovereignty that is based on the assumption that 'sovereignty' is fundamentally an attribute of states, see Feinberg, 1989, ch. 19. Feinberg's focus is not with the enforceability dimension of personal sovereignty, but in relevant aspects, his discussion is consistent with the claims I make here. See Feinberg, 1989, esp. pp. 52 & 68.

the absolute sovereignty of states.

Consider a system of states in which each state is sovereign according to Philpott's definition of the term. In such a system of states, each state enjoys exclusive authority over all matters within its territory, and other states have a duty not to interfere with its governance. To argue that victim states have certain enforcement rights against third parties is to deny that states *always* have a duty not to interfere with each other's governance, which amounts to denying that states are *absolutely* sovereign. Suppose a victim state (or other states choosing to come to the victim's aid) ask a neutral state's consent to use some of the neutral state's resources in an effort to fight off an attacker state. Suppose the neutral state denies access to its resources—it does not wish to get involved. If the victim state (or other states choosing to come to the victim state's aid) may nevertheless go ahead and put the neutral state's resources to use, then the neutral state simply does not enjoy *exclusive* rightful control over *all* matters within its territory.

The situation is different when we are dealing with attacker states. An attacker state has violated (or is threatening to violate) its duty not to interfere with the authority that other states enjoy over their territory. This makes it *liable* to defensive action: because it shows insufficient respect for the authority of others, it *forfeits* some of its own rights so that order can be restored. Owing to the concept of forfeiture, there is thus room in the natural law governing sovereign states for the idea that a victim state (and other states choosing to come to the victim's aid) have a right to fend off an attacker state.

In sum, the absolute sovereignty of states leaves room only for certain enforcement rights, namely for those held against attackers. For our purposes, this is interesting in two different but related ways. First, it suggests that absolute or maximal sovereignty is not a matter of *extensive* enforcement rights. *Some* enforcement rights against attackers may well be necessary for sovereignty, but we can refer to a state as *absolutely* sovereign without having maximized the extent of its enforcement rights. As long as a state has sufficient enforcement rights, what determines the extent of its sovereignty is simply the exclusivity of its authority over its territory, so that a state is absolutely sovereign if its authority extends to all matters within its territory, without exception.

In other words, *rightful exclusive control over a territory* sits at the *core of sovereignty*; enforcement rights play a mere supporting role.

Second, the fact that absolute or maximal sovereignty is incompatible with enforcement rights against uninvolved third parties suggests that the existence of such rights would not actually promote sovereignty, but would rather detract from it. If the core of sovereignty is exclusive rightful control over a territory, no measure that encroaches on this core can boost sovereignty.

If we apply these insights to the concept of personal sovereignty, they suggest the following revised picture of its enforceability dimension (cf. Table 2):

	Rights held against the attacker
Rights held by...	Victim
	Consistent with core value of sovereignty because attacker forfeits some rights; to some extent necessary for sovereignty
	Third parties

Table 2: Personal Sovereignty’s Enforceability Dimension

According to the revised picture, our rights to non-interference sit at the core of personal sovereignty. Our sovereignty increases with the stringency of our rights to non-interference, or with the exclusivity of the authority we enjoy over our bodies and minds. If our sovereignty were absolute, our inviolability would be maximal—it would never be permissible to interfere with us against our consent. But rights to non-interference do not exhaust personal sovereignty. Unlike beautiful flowers, we are the kind of creature who has the *capacity* to defend itself against illegitimate interference, and our moral status as personally sovereign means that we are *morally entitled to make use of that capacity*. Those who illegitimately interfere with us forfeit some of

their rights to non-interference, and in defending ourselves, we may harm them to the extent that they have forfeited their rights.³¹

Thanks to the concept of forfeiture, enforcement rights held against attackers do not limit our rights to non-interference and do not, therefore, encroach on the core value of personal sovereignty. By disrespecting the moral status of others, attackers *make* themselves more violable; they *cause* their moral status to be diminished at least temporarily, so that the proper moral order may be restored. With enforcement rights against innocent bystanders, the situation would be different. Because innocent bystanders have done nothing to forfeit their rights to non-interference, enforcement rights held against innocent bystanders would impose limits on our rights to non-interference, and would thus encroach on the core value of personal sovereignty. It follows that rights to harm innocent bystanders need to be excluded from the enforceability dimension of personal sovereignty. To promote personal sovereignty is to give innocent individuals *more authority over their own bodies and minds*, not to restrict that authority in order to give each individual rights over the bodies and minds of *others*.

If personal sovereignty never increases with the permissibility of harming innocent bystanders, then personal sovereignty does not increase with the permissibility of infringing rights to non-interference for minimization's sake, and the fittingness defence of the inviolability approach is restored. According to the fittingness defence of the inviolability approach, our rights to non-interference do not have to follow a minimization logic because their purpose is to protect our inviolability, which increases with the stringency of our rights to non-interference. As rational creatures, we value inviolability independently of its effect on our welfare, as it entitles us to lead self-determined lives by making us sovereign over our bodies and minds.

³¹ Suppose I have a right not to be killed for fun. Further suppose that I have the following enforcement right over that right: if someone tries to kill me for fun, I may defend myself by imposing harm of size X on my aggressor, where X corresponds to roughly a punch in the stomach. This enforcement right would seem to fall short of an *appropriate* enforcement right for the right not to be killed. According to the picture of sovereignty sketched here, does this mean that I am merely inviolable, but not sovereign? Or are minimal enforcement rights sufficient for sovereignty, but my sovereignty would increase if I had 'more appropriate' enforcement rights (other things being equal)? I am not sure. But to rebut the enforceability criticism of the inviolability approach, we need not answer these questions.

Crucially, none of this is to say that we should be maximally inviolable, or that it should never be permissible to harm innocent bystanders against their consent. As I have suggested in section 3.8, other values may sometimes conflict with the value of personal sovereignty, in which case either value may win out. The only thing I have tried to show here is that we cannot promote *personal sovereignty* by making it permissible to harm innocent bystanders. If the core idea of some value is that individuals should have certain rights, we cannot promote that value by curtailing these very rights.

3.11 Effective Rights to Non-Interference

As sovereign beings, we are morally entitled not to be interfered with against our consent, and we are morally entitled to defend ourselves against illegitimate interference. But if personal sovereignty is simply a matter of our moral status, then it should not be our exclusive concern insofar as we are rational beings. In section 3.8, I argued—following a long line of thinkers inspired by Kant—that insofar as we are rational beings, we aspire to lead self-determined lives, and that personal sovereignty is an important precondition for leading such a life. But if personal sovereignty is merely an important precondition for leading a self-determined life, what else is needed for us to be able to actually lead such a life? Suppose someone argued as follows: “To lead an autonomous life, we must by and large be able to count on not being interfered with against our consent. If others frequently interfere with us against our consent, or if they inflict grave harm on us, our ability to decide things for ourselves is undermined. In fact, our ability to lead a self-determined life is threatened even if there is merely a high risk that we will be interfered with against our consent. It is thus an important aspect of leading a self-determined life that we have largely *effective* rights to non-interference. Rights to non-interference and enforcement rights are important not only because they give expression to the fact that we are sovereign beings who are morally entitled to decide things for themselves; they are important also because they make

it more likely that we will not be interfered with against our consent.³² All things considered, our lives may be more self-determined if we have less extensive rights to non-interference that we are more often permitted to enforce than if we have more extensive rights that we are less often permitted to enforce.”

The basic thrust of this argument seems to me correct. If it is, then our sovereignty should be limited not merely because as sentient creatures, we care about the general welfare, but also because as rational creatures, we care about the effectiveness of our rights to non-interference. Consider the following example. It is doubtlessly permissible to step on someone’s toes against that person’s consent to save five people from being killed by a vicious murderer. This implies that our sovereignty is less than absolute. Limiting our sovereignty in this way is justified because of the large net gain in welfare it allows us to bring about. But it is also justified for the sake of autonomy. By trying to kill the five, the vicious murderer arrogates to himself a decision about their fate. If we are permitted to stop him even though this involves trespassing on someone’s body, then we are able to put the five back in control over their lives, thus promoting their autonomy. Moreover, if the murderer manages to kill the five, he robs them of all future opportunities to pursue their own plans and projects. If we are permitted to stop him, we are able to preserve these future opportunities for them, thus promoting their ability to lead a self-determined life. It is true that if it is morally permissible to step on someone’s toes against that person’s consent to save the five from being killed, then our authority over our bodies is not as exclusive as it could be. But this limitation does not seem to detract significantly from our ability to lead a self-determined life. Having even just a slightly better chance of not being killed by a vicious murderer would seem to easily make up for it.

These considerations complicate the picture presented in section 3.8, which gave the impression that personal sovereignty neatly captures our moral concerns insofar as we are rational beings who aspire to lead self-determined lives. It clarifies that our capacity for rational agency grounds a concern not only with what rights we have, but

³² In fact, being able to lead a self-determined life depends on more than personal sovereignty and effective rights to non-interference. Accidents or natural disasters that harm us without violating our rights may also undermine our ability to lead a self-determined life. The same goes for addictions and other mental afflictions, or for extreme poverty. Since none of these aspects are directly relevant to our discussion, I leave them aside here so as not to complicate things.

also with what is done to us. When these two considerations come into conflict, and when the latter carries more weight given the circumstances of the case, we should think of personal sovereignty as limited in such a way that this limitation *fits* the rational aspect of our human nature. This more complex picture *need not* undermine the fittingness defence of the inviolability approach, but it has the *potential* to do so. More precisely, it creates a problem for the fittingness defence just in case it is true that our moral rules *best* respond to our capacity for rational agency if moral rights to non-interference are *always* permissibly infringeable for minimization's sake. But is this the case?

Consider again the MTC. Intuitively, it seems wrong to push the massive man. Can we make sense of this intuition if rights to non-interference should always be permissibly infringeable when this would on balance constitute the best possible response to the rational aspect of our human nature?

One striking feature of the MTC is that if you kill the massive man, you *use him as a means* to save the five people trapped on the trolley track. If we agree with Quinn that as a rational being, the massive man is not a mere “cell in the collective whole”, but a “being in his own right” who ought to be recognized as such, we may feel that sacrificing him for the sake of the five would be especially abominable, so that morality ought not to endorse it. But even if pushing the massive man would show an especially blatant disregard for his rationality, it is not clear that it should therefore be ruled out as impermissible. After all, it would save *five* people from having *their* rationality blatantly disregarded by a vicious murderer who is trying to kill them for his sick pleasure. It is true that if the five are killed, they are killed *impermissibly*—their moral status marks them out as beings who must not be killed for the thrill of it. But how significant a consideration is this?

The principled thing we can say is the following. Whether the fittingness defence of the inviolability approach to solving the paradox of deontology still goes through depends on the proper understanding of what it means to live a life that suits a rational creature, or what it means to self-determinedly shape one's life in accordance with one's own ideas of the good. If being self-determined is importantly a matter of one's moral status as a sovereign being, it still goes through. If being self-determined

is exclusively a matter of effectively controlling whether one is being interfered with, it does not. In the former case, it is important that morality *recognizes* our existence as beings who must not be interfered with against their consent. In the latter case, our moral rules ought to simply *advance* our not being interfered with against our consent, and personal sovereignty matters only because and to the extent that it promotes our not being interfered with. At first sight, this may look like a stalemate that cannot be decided without further investigation into concepts such as self-determination and autonomy. Wouldn't it be mere stipulation if at this point, we insisted that being sovereign is an important aspect of leading a life that suits a rational being?³³ On closer inspection, however, the situation is not undecided.

Recall how the paradox of deontology starts from the observation that our firmly held judgements about cases such as the MTC indicate that rights do *not* always follow a minimization logic. If the sole purpose of our rights were to advance our not being interfered with against our consent, then these judgements would have to be discarded. For if rights had this exclusive purpose, they would *always* and *necessarily* have to be permissibly infringeable for minimization's sake. Otherwise, they would defeat their own purpose. Yet if at least *part* of the purpose of our rights is to express the fact that we are morally entitled not to be interfered with in certain ways, then we can make sense of our intuitive judgements about cases such as the MTC. It follows that our intuitive judgements lend support to the idea that being morally entitled to non-interference is important in its own right, and therefore provide us with a reason to accept conceptions of the good life for a rational being that take the significance of moral status into account.

³³ Feinberg argues that personal sovereignty figures importantly in our understanding of autonomy. See Feinberg, 1989, chs. 18 and 19.

3.12 Conclusion

If our moral rights to non-interference are not always permissibly infringeable for minimization's sake, we are more inviolable than we would otherwise be. Being highly inviolable is good for us: it means that we are morally important creatures who are "worthy of respect, owed respect" (Kamm, 1996, p. 272). This suggests that we can appeal to considerations of moral importance to explain and justify the peculiar non-minimization logic of our rights to non-interference.

But on closer inspection, it is not clear that our moral importance goes up whenever our inviolability increases. In minimization cases, inviolability is directly opposed to saveability, so that if one of these aspects of our moral status goes up, the other goes down. On at least one plausible interpretation of moral importance, the effects of these changes on our moral importance would seem to cancel each other. In light of this, I have argued that we should move away from considerations of moral importance in our attempt to solve the paradox of deontology.

The fittingness defence of the inviolability approach is based on the idea that being highly inviolable fits the kind of moral creature we are. To the extent that others must not interfere with us against our consent, our moral rules put us in charge of our own lives. As rational beings who aspire to lead self-determined lives, this is something we value independently of its effects on our welfare.

The enforceability criticism proceeds from the idea that we are truly in charge of our own lives only if we are permitted to defend ourselves against illegitimate interference. From this, it infers that our capacity for rational agency gives rise to our moral status as *personally sovereign*, which consists of an inviolability as well as an enforceability dimension. I have argued that this basic contention is correct. As personally sovereign beings, it is wrong to interfere with us against our consent, and we may rightfully defend ourselves against illegitimate interference. Those who illegitimately interfere with us disrespect the fact that we are morally entitled to decide certain things for ourselves, and for that reason forfeit some of their own rights to non-interference.

The enforceability criticism contends that our overall personal sovereignty is higher if our rights are permissibly infringeable for minimization's sake. With this contention, it threatens to undermine the fittingness defence of the inviolability approach. For if it is true that our capacity for rational agency gives rise to our moral status as personally sovereign, and if it is true that our sovereignty is higher if our rights follow a minimization logic, then it becomes doubtful whether the presumed non-minimization logic of our rights to non-interference can be explained and justified with reference to the rational aspect of our human nature.

But I have argued that the enforceability criticism fails. The enforceability dimension of personal sovereignty is necessarily limited to enforcement rights held against liable attackers. To extend it to enforcement rights against innocent third parties is to assume that we can become *more* sovereign as the rightful control that each of us enjoys over their own territory becomes *less exclusive*. I have argued that this assumption is not tenable.

There is, however, a further complication. As rational creatures who aspire to lead self-determined lives, we care about rights to non-interference and enforcement rights partly because they allow us to *effectively control* whether we are being interfered with. This fact is consistent with the fittingness defence of the inviolability approach as long as there is independent value in our being entitled to non-interference. But do we—and, more importantly, should we—value being *entitled* to non-interference even when this may lead to more instances of unwanted interference?

I think so. I have ended the last section with a coherentist argument in favour of this view. If our moral status matters independently of its effects on what happens to us, this allows us to make sense of our intuitions about cases such as the MTC, and in this way gives us reason to trust them. Our considered intuitive judgements in turn provide us with a reason to endorse the idea that is able to explain these judgements.

But we need not stop at this narrow reflective equilibrium. The idea that a human person is a particular kind of moral creature whom we have reason not to treat in certain ways is independently plausible and powerful. It accurately captures at least one way in which we tend to think about ourselves and others from a moral point of view. If our rights to non-interference make us sovereign over our bodies and

minds because it fits the rational aspect of our human nature that each of us should have primary say over what may be done to him or her, then it is true—as Nozick suggests—that killings are bad and ought not to occur. But, contra Nozick, we have a right against being killed *not because* killings are bad and ought not to occur, but *because interfering with us against our consent is an inappropriate response to the rational aspect of our human nature*. Importantly, there is no conceptual incoherence nor threat of self-defeatingness if certain ways of interfering with a human person constitute an inappropriate response to her rational nature *even if interfering with her against her consent would save several others from being interfered with against their consent*. If the purpose of our rights is to express that each of us should have primary say over what may be done to him or her, their purpose does therefore not dictate a minimization logic. It follows that the fittingness defence of the inviolability approach provides us with a coherent and plausible way of thinking about rights to non-interference that leaves no room for the worries of self-defeatingness that underlie the paradox of deontology. On the assumption that we ought not to harmfully interfere with each other because each of us ought to be recognized as a sovereign ruler over the territory of her body and mind, the paradox simply does not get off the ground.

4 | The Redirection of Bombs

During World War II:

A Philosophical Investigation^{*}

“It is no use saying, ‘We are doing our best.’ You have got to succeed in doing what is necessary.”

– Winston Churchill (undated)

4.1 Introduction

In June of 1944, the German Wehrmacht started terror bombing London. The Germans were most likely targeting the area around Charing Cross in Central London, but the actual ‘mean point of impact’ (MPI) of their bombs lay in Dulwich, some four miles south-east of Charing Cross. With the help of double agents that the British authorities had recruited among German spies, it would have been possible to redirect the bombs further away from Central London. This would likely have saved a great number of civilian lives in Central London, while resulting in an increase of civilian deaths in certain areas in South London. The expected net effect was

^{*} Earlier versions of this chapter were presented at the MANCEPT Workshops in Political Theory, the London School of Economics and Political Science, the ‘Kritik und Krise’ Symposium der Schweizerischen Philosophischen Gesellschaft, and the Society for Applied Philosophy Annual Conference. I owe special thanks to Edward Carpenter, who conducted an MSc in Philosophy of the Social Sciences at the London School of Economics in the academic year of 2011-12. On a suggestion by Alex Voorhoeve, Carpenter wrote a formative MSc essay on the issues taken up in this chapter, and was kind enough to let me use his material as a starting point for my own investigations. I thank Alex Voorhoeve for introducing me to the historical case, and for helping me carry out archival research for the purposes of this chapter.

a significant reduction in the overall number of civilian deaths. While the British deception authorities were convinced that redirecting the bombs was the right thing to do, the War Cabinet's Ministers (meeting without the Prime Minister, Winston Churchill) were opposed, and issued an order to "confuse the enemy" instead. A battle of wills ensued. The deception authorities argued that while confusing the enemy was not a feasible alternative, it would be possible to try and stabilize the MPI at its current location in Dulwich, if this was deemed preferable to redirection. In the end, the War Cabinet's Ministers authorised the deception authorities to implement a cautious redirection strategy.

In this chapter, I investigate whether the Ministers were right to feel hesitant about the redirection strategy. Were there any considerations that spoke against it? If yes, did these considerations make it the case that confusing the enemy or stabilizing the MPI at its current location would have been morally preferable to redirection?

In section 4.2, I take a close look at the historical details of the case.¹ In section 4.3, I consider whether the Ministers' proposal to confuse the enemy would have been feasible. I conclude that it would have been feasible, but that implementing it would have jeopardized the double agents' excellent reputation with their German controllers. In section 4.4, I investigate how the historical case has been discussed in the philosophical literature. I reject both Jonathan Glover's (1977, p. 103) proposal that the Ministers' resistance to redirection stemmed from their belief that it makes a moral difference whether a given harm is the result of a positive intervention or an omission, as well as John Martin Fischer and Mark Ravizza's (1992, p. 18) proposal that the Ministers' resistance to redirection can be explained by an appeal to the Doctrine of Double Effect (DDE). In section 4.5, I defend the idea that the Ministers' resistance to redirection might have stemmed from a legitimate concern with treating all of their citizens equally. Unlike the Ministers' proposal to confuse the enemy, the redirection strategy was a protective measure that did not express equal concern for all citizens; instead it favoured some parts of the population over others. This

¹ In November 2013, I carried out archival research for the purposes of this chapter at the UK National Archives in Kew, Richmond, where an abundance of documentary material relating to the 1944-45 terror bombing of London can be viewed by the general public. The archival research for this chapter was carried out jointly with Alex Voorhoeve.

made the redirection strategy morally problematic, as a government owes its duties of protection equally to all of its citizens. In section 4.6, I argue that the moral appeal of the proposal to fix the MPI at its current location is best understood from a rights-based perspective. If the Ministers were worried that because of the unequal concern that it expressed, the redirection strategy might violate some citizens' right to life, they could to some extent have circumvented this worry by instead stabilizing the MPI at its current location. In section 4.7, I discuss whether a decision to adopt the redirection strategy would have been difficult to defend to the general public. I suggest that it should have been possible to secure the general public's approval of the redirection strategy by combining it with a credible commitment to a swift and decisive elimination of the overall bomb threat. I argue that if the British government had lived up to such a commitment, redirection would have been morally preferable to confusion, so that the redirection strategy could have formed part of a morally optimal response to the bomb threat. Section 4.8 concludes.

4.2 The Historical Case

In June of 1944, the German Wehrmacht started terror bombing London.² The 'V-weapons' ('Vergeltungswaffen', or 'retaliatory weapons') that the Wehrmacht fired at London were guided missiles that German engineers had started to develop in the late 1930s. When the terror bombing of London started, the V-weapons were at a stage of development where they performed reasonably well in test runs, but still had some flaws. Most importantly, their guidance system was not yet very accurate, so that the German engineers needed to know where their missiles were falling in order to make the necessary adjustments. Due to the fact that the British press was heavily censored, and because reconnaissance flights over London were not feasible, the Germans turned to their spies that they had stationed in London. They asked them to urgently report timing and location of bomb impacts as completely and as accurately as possible.

² Where not indicated otherwise, this section draws on Campbell (2012).

As it happened, many London-based German spies had been turned into British double agents by June of 1944. These double agents now inquired with their British controllers what information they should pass on to their German contacts.

The British thought it likely that the Germans were trying to make Charing Cross³ the target ‘mean point of impact’ (MPI) of their bombs, thus maximizing their bombs’ expected negative impact on critical infrastructure, civilian lives, and British morale. The British soon realized, however, that the V-weapons were on average falling short of their presumed target. While individual bombs fell scattered all over Greater London (and around South East England at large), it was evident that south-east London was getting the brunt of the attack. A few days after the bombings had started, the V-weapons’ actual MPI was calculated to lie somewhere in Dulwich, some four miles south-east of Charing Cross (Jones, 1978, p. 420).

While individual bombs had a limited blast radius and tended not to be very deadly, many bombs were fired at London each day. When the bombings stopped on 28 March 1945, just under 4,000 V-weapons had reached the Greater London area, killing almost 9,000 civilians, and injuring about 24,500 (Campbell, 2012, p. 445). The deadliest bomb fell on Saturday, 25 November 1944, in New Cross, south-east London. It hit a Woolworth’s department store full of weekend shoppers, killing 168 people and leaving 122 seriously injured (ibid.).

In late June of 1944, when the bombings had only been going on for a couple of weeks, the British authorities had to decide what information their double agents should transmit to their German controllers. Sir Samuel Findlater Stewart, chairman of the Home Defence Executive,⁴ came up with a deception policy that he sent to the Armed Forces Chiefs of Staff for approval. His proposal was to make the double agents selectively report actual incidents so as to “create the impression that the bombs are appearing to overshoot the target (which we assume to be Central London) in the hope that the range and deflection may be moved further to the east and south than it is at

³ Charing Cross is a busy railway station in Central London. On a roundabout right next to the station, the equestrian statue of Charles I marks the ‘official centre’ of London.

⁴ The Home Defence Executive, later renamed to Security Executive, was a committee set up in 1940 by the Prime Minister to organise the defence of Britain against enemy forces.

present” (CAB 113/35, 5 July 1944).⁵ In other words, Stewart’s idea was to redirect the bombs even further away from Central London and towards less built up areas by feeding the Germans *true but strategically selective* information. Stewart’s proposal was *prudent*: it protected the excellent reputation that the British double agents enjoyed with their German controllers by allowing them to pass on seemingly useful and in principle verifiable information in a prompt way. For the British authorities, safeguarding the reputation of their double agents as reliable German spies was a very important concern, as they wanted to ensure that their double agents would be available for further deception projects in case the war should drag on. Findlater Stewart’s proposal also promised to be *effective*, or to (i) result in a net saving of many civilian lives, to (ii) curb the number of bomb-related injuries, and to (iii) reduce the risk of damage to critical infrastructure. If—as seemed likely—the Germans had no choice but to rely almost exclusively on their spies’ reports to establish where their V-weapons were falling, the British were presented with a genuine opportunity to keep in check the expected negative impact of the enemy’s bombs.

Findlater Stewart pressed the Armed Forces Chiefs of Staff to come to a quick decision in the matter. The double agents had been ordered by their German contacts to make reporting bomb incidents their top priority, so that they could not credibly remain silent for very long. Given this, the Chiefs of Staff approved Stewart’s proposal and instructed the deception authorities to go ahead with it for the time being (CAB 121/214, 7 July 1944). But because civilian lives were involved, the Chiefs of Staff also felt that they were not authorised to decide the matter finally. Hence, they forwarded Stewart’s proposal to the War Cabinet, and asked for the Ministers’ backing of the redirection strategy (CAB 121/213, roughly 10 July 1944). On 28 July 1944, the War Cabinet (with Winston Churchill absent that day) discussed Findlater Stewart’s proposal and argued “that it would be a serious matter to assume any direct degree of responsibility for action which would affect the areas against which flying bombs were aimed” (CAB 121/214, 28 July 1944). To the War Cabinet’s Ministers, “it was not for the Government to authorise action which would result in certain areas

⁵ In the filing system of the UK National Archives, ‘CAB’ refers to records of the Cabinet office. The ‘PREM’ and ‘KV’ abbreviations that will come up later in this section refer to records of the Prime Minister’s office and to records of the Security Service, respectively.

in South London getting a heavier discharge of flying bombs and sustaining heavier casualties than they do at the present time” (PREM 3/111A, 3 August 1944). The Ministers felt that “if it came out that the Government had authorised such action their position would be indefensible” (ibid.). The Ministers concluded that Findlater Stewart’s proposed redirection strategy “could not be accepted” and “invited the deception authorities to arrange that the information conveyed to the enemy [...] was such as would create confusion in his mind and present him with an inaccurate picture” (CAB 121/214, 28 July 1944). Findlater Stewart’s proposal was thus deemed unacceptable despite the fact that neither its prudence nor its effectiveness were called into question.

This exasperated Findlater Stewart. In a second report (CAB 113/35, 31 July 1944), he argued that “confusing the enemy” was not an option for the following reasons:

1. “We are forced to give some information in order to maintain our deception machine in being.
2. Information cannot be given haphazardly since the enemy will draw an inference about his aim from whatever is said and we must therefore indicate some mean point of impact which would not be to our detriment.
3. If we tell the enemy or let him infer the truth we present him with valuable information which he can use to improve his aim, and [...] a shift to the north might be very serious.”

He then proposed a second deception strategy according to which “no attempt is to be made to shift the weight of attack in any direction, and material to be passed over should be selected so as to deny useful information to the enemy and to prevent him from moving the mean point of impact of his attack, particularly to the north” (CAB 113/35, 31 July 1944).

On 15 August 1944, the War Cabinet (with Winston Churchill now present) discussed Findlater Stewart’s second report. It seems that they were moved by Stewart’s arguments to the effect that “confusion must [...] be regarded as

impracticable” (CAB 113/35, 31 July 1944). They informed Stewart that they had reconsidered their original instructions, and that the deception authorities’ new “object should be to ensure that there is no deterioration in the position and that the enemy does not shift the pattern of his bombs towards the North West. With this in view you should continue to convey to the enemy information which will confirm his belief that he has no need to lengthen his range. You are also at liberty, within limits, to take such steps as you may judge safe to intensify this belief” (PREM 3/111A, 15 August 1944).

It is not entirely clear what the deception authorities made of these revised orders. They may have read them as a ‘go-ahead’ for Findlater Stewart’s second proposal, so that they tried to keep the MPI at its current position in Dulwich. Or they may have interpreted the revised instructions as consistent with Stewart’s initial proposal, so that they kept attempting to shift the MPI further south-east.⁶ What we know for certain is that they managed to induce the enemy to believe that he was on average on target when in fact he was not. German Intelligence appreciation of the fall of their V-weapons shows that the Germans were under the impression that they were hitting Central London with some precision (Howard, 1990, Fig. IA; Jones, 1978, Ill. 24), when the overall MPI for all the V-weapons that fell in the Greater London area between June 1944 and March 1945 in fact lay in Beckenham, eight miles south-east of Charing Cross (Campbell, 2012, p. 346; Howard, 1990, Fig IB).⁷ The fact that the eventual MPI lay several miles south-east of Dulwich (the area where the V-weapons had been falling most heavily in the first few days of the bombings) lends support to the hypothesis that the deception authorities went ahead with Findlater Stewart’s initial redirection strategy. Together with military

⁶ Howard (1990, p. 177, emphasis added) suggests that a sort of cautious redirection strategy was adopted, as the deception authorities felt they had been authorized “to prevent the enemy from moving his aim towards the north-west and, *to a slight extent*, to attempt to induce him to move it towards the south-east.”

⁷ This is true for the first generation of V-weapons, the V-1 flying bombs. Deception also took place with regard to a second generation of V-weapons, the V-2 rockets. The V-2 rockets at first had a tendency to overshoot London by many miles: in early October 1944, a few weeks after the first V-2 rocket had been launched with London as a target, most V-2’s fell around Norwich, around 100 miles north-east of London (CAB 121/215, 15 October 1944). The enemy subsequently fine-tuned his aim and was made to believe that he had managed to make Central London the MPI of his V-2’s once their actual MPI lay in Ilford, a London borough roughly 9 miles north-east of Charing Cross (Campbell, 2012, p. 445).

authorities, the deception authorities were at any rate extremely careful to avoid renewed involvement of the War Cabinet’s Ministers in deception issues. As General Ismay wrote to General Hollis on 26 January 1945, the Chiefs of Staff had “got [their] fingers badly burnt by bringing Ministers into this [deception] business”, and he, for one, should certainly not “want to risk doing so again” (CAB 154/49, 26 January 1945).

That some sort of deliberate deception went on throughout the bombings is furthermore suggested by a comment made by Wernher von Braun—a German engineer who took a leading role in the development of the V-weapons—in an interview with the *New Yorker* that appeared in its issue of 21 April 1951. Wernher von Braun there tells his interviewer (Lang, 1951, p. 88): “I found myself wondering [...] where our German agents in London had disappeared to. I never did find out, but one thing I know is that we had some good ones there. Our battery commanders on the French coast used to have reports on V-2⁸ effectiveness within an hour after a rocket had been launched. I’ve never heard in so many words how the reports reached them, but I assume it was by radio.”

4.3 The Feasibility of Confusing the Enemy

Findlater Stewart was incredulous at the Ministers’ resistance to his proposed redirection strategy. To Stewart, the Ministers’ instruction to ‘confuse the enemy’ must have seemed on all counts inferior to purposeful redirection. If confusion was a feasible option at all—something Stewart explicitly doubted—it was *less* suited to mitigate the harm that the V-weapons were bringing to London, and it seemed *more* likely to put Britain’s double agents at risk (to effectively create confusion, the agents might for example have to issue contradictory reports).

In the remainder of this chapter, I want to investigate whether Stewart was right not to have any sympathy for the Ministers’ ideas. To do this, I will suppose that ‘confusing the enemy’ was a viable option. Findlater Stewart doubted this. He argued that the Germans would calculate an MPI from *whatever* information they were given,

⁸ Von Braun here refers to the second generation of V-weapons. See footnote 7.

and that haphazardly reporting individual bomb impacts would not confuse anyone, but would merely amount to the throwing away of a valuable chance at strategically influencing the enemy. It seems to me, however, that Stewart was too hasty in his conclusion. He probably saw no compelling reason in favour of the confusion strategy, and for that reason did not think hard about possible ways in which to implement it. *Contra* Findlater Stewart, I find it plausible to assume that confusion would have been a viable option for the following reasons. First, it seems that different double agents could have issued reports that more or less cancelled each other. A first agent could have reported rumours of an impact at some given time and place; a second could have reported the same rumour, then saying that he went to check it out for himself, only to find no bomb damage. The agents could also have reported bomb damage in highly specific locations, but combined this with only very rough estimates of the times at which the respective bombs had fallen. This would have made their reports less valuable for the German engineers, who were interested in finding out where individual bombs had fallen. The double agents could quite credibly have argued that it made them look suspicious if they insisted on learning the precise time at which some bomb had exploded.⁹ A number of agents could furthermore have announced that they were not willing to stay in London and risk their lives, and that they would thus be retreating to the countryside for as long as the bombings went on. They could have turned the bombings into a personal issue by scolding their German controllers for not giving them an advance warning of the impending attacks.¹⁰

For the remainder of this chapter, I will thus suppose that ‘confusing the enemy’ (understood as ‘providing the enemy with largely useless information from which he should be unable to calculate an MPI’) would have been a feasible strategy if the British authorities had been willing to jeopardize their double agents’ reputation as reliable German spies. I will further suppose that if the British authorities had chosen

⁹ When the terror bombing started, double agent ‘Garbo’ used this latter strategy in order to play for time while waiting for instructions from his British controllers. Garbo even staged his own arrest, claiming that he was arrested at a bomb site because he had asked too many questions. His German contacts bought this story and told Garbo to lay low for a while. Cf. KV 2/69, 28 June 1944, and KV 2/69, 14 July 1944.

¹⁰ Some German contacts did in fact try to issue advance warnings to their favourite spies; others apologized for not having been able to do so. See Delmer, 1971, pp. 202-4. See also Howard, 1990, p. 169.

to adopt the confusion strategy, they would not have been able to predict where the MPI of the German V-weapons would subsequently have moved to (or whether it would have moved at all).

4.4 How the Historical Case Has Been Analysed in the Philosophical Literature

Sefton Delmer (1971, pp. 206-7), who worked for the British intelligence authorities during WWII, recounts the War Cabinet meeting in which the Ministers decided to reject Findlater Stewart's redirection strategy in favour of their confusion strategy as follows:

Churchill was not present at the meeting, and from the brief account which filtered down to the deception staff the plan [to redirect] was received with great distaste. Herbert Morrison was known to be against it. "Who are we", the Minister of Home Security was said to have asked his fellow Ministers, "to act as God? Who are we to decide that one man shall die because he lives in one place, while another survives because he lives in a different one?" No one seems to have pointed out that if the qualifications of the Cabinet for the status of God were thereby not immediately apparent, Hitler's were non-existent and he had started the whole business; or that in total war it might be necessary to put a batallion at risk to protect a brigade. Morrison's view won the day.

For my purposes, it does not matter whether Morrison really was worried about "acting as God". What I am trying to establish is whether there were *good reasons* against the redirection strategy. If Morrison's alleged refusal to act as God points to such reasons, the veracity of Delmer's memoirs need not concern us.

In the philosophical literature, Jonathan Glover (1977) as well as John Martin Fischer and Mark Ravizza (1992) take Morrison's alleged refusal to act as God as a starting point to think about possible reasons in favour of the Ministers' resistance

to redirection.¹¹ In the next two sections, I discuss their ideas.

4.4.1 The Difference Between Doing Harm and Allowing it to Occur

In the philosophical literature, the historical details are greatly distorted. Both Glover (1977, pp. 102-3) and Fischer and Ravizza (1992, p. 17)—the latter following Glover—represent the choice the British decision-makers were faced with as one between letting the bombs fall on Central London (where they were allegedly already falling in a very concentrated way), or redirecting them away from Central London. Given this misrepresentation, they then discuss whether the British authorities should have had their double agents *confirm that the German bombs were in fact hitting Central London*, or whether they should have used their double agents to *deceive the enemy into readjusting his aim to so far unaffected South London* by coming up with fake incidents of bombs falling in the north of the Greater London area.¹²

Given this misrepresentation, Glover (1977, p. 103) interprets Morrison’s alleged refusal to act as God as a refusal to *actively* redirect the bombs away from Central London when it was possible to *passively* let them fall where they were already falling. Glover then ascribes to Morrison the belief that “acts are more objectionably God-like than omissions” (ibid.), and subsequently rejects this belief as unfounded. Glover thus identifies a *possible* reason against redirection, but does not ultimately endorse it as a *valid* reason.

Not only does Glover get the details of the case wrong, but his analysis seems to me inadequate even if we accept his misrepresentation of the case. That is, even if the British authorities would have been faced with a choice between redirecting

¹¹ The bomb redirection case is also discussed by Michael Moore (2009, pp. 75-6) and David Edmonds (2013, pp. 3-8), but neither Moore nor Edmonds provide additional clues as to why the Ministers might reasonably have felt hesitant about redirection.

¹² Moore (2009, pp. 75-6) gives the same misdescription of the British decision situation as Glover and Fischer and Ravizza do. Edmonds (2013, pp. 5-6) describes the historical facts somewhat more accurately; he points out that the German bombs were in fact falling *short* of their presumed target. But he nevertheless misrepresents the choice that the British government was faced with as one where it could either “do nothing”, or where it could alternatively “try to change the trajectory of the [V-weapons]—through a campaign of misinformation—and so save lives” (Edmonds, 2013, p. 8).

the bombs away from Central London and letting them fall where they were already falling, both of these options would have required a *similar degree of activity*. No matter which option the authorities would have settled for, they would have had to instruct their double agents with similarly specific details regarding the information that the double agents were required to pass on to their German controllers. If we were to frame the British authorities' choice (as described by Glover) as a trolley case, we should thus not think of it as a standard 'side track' case where a runaway trolley will kill five people unless an agent decides to redirect it onto a side track where it will kill only one person (see Thomson, 1976, p. 207). We should instead think of the British choice as an amended side track case where an agent has to decide whether to *press a button* so that a runaway trolley continues to run towards five people (and will eventually kill them), or whether to *press another button* so that the trolley is redirected onto a side track, where it will eventually kill only one person. To put the same point differently, the British choice as described by Glover was actually a choice between imposing harm *in a given way* on more or on fewer people.¹³

I can make out one potentially morally relevant difference between the two courses of action that Glover thinks the British authorities considered implementing.¹⁴ Suppose the British authorities decided that they must not in any way get involved in the terror bombing of their own citizens. If so, what course of action should they have settled on? It seems to me that they should have *discontinued all communication*

¹³ Note that this makes the amended side track trolley case structurally similar to Philippa Foot's original trolley case. See Foot, 1967 and Thomson, 1976, p. 206. Foot and Thomson agree that in Foot's original trolley case, it is permissible to redirect the trolley away from the five and towards the one.

¹⁴ The two courses of action that Glover describes also differ in that one involves telling the truth about specific bomb incidents, whereas the other involves lying about them. But a double agent who tells the truth about a specific bomb incident is still hiding the fact that he is no longer working for the German side. Like a double agent who lies about specific bomb incidents, he is thus not committed to telling the truth. Having said that, I do not think that these considerations are of moral relevance in the case at hand. Throughout this chapter, I will simply assume that it is not morally wrong to lie to a culpable attacker in order to thwart his evil plans. Kantians argue that manipulative lies—i.e. lies told with an intention to get the addressee to act differently than he or she otherwise would—are morally reprehensible because they undermine the addressee's ability to make an autonomous choice (see e.g. Hill, 1991). Their idea is that the person who is lied to is used as a mere means to serve the deceiver's ends. To me, it seems plausible to assume that deceiving someone in this way involves using them as a mere means. But it seems equally plausible to assume that a culpable attacker *forfeits* his right not to be used as a mere means in the thwarting of his own attack.

with the Germans via the turned agents. I will thus refer to this course of action as *no involvement*. Now in Glover’s misrepresentation of the case, if the Germans would have been told the truth—namely, that their bombs were already hitting Central London—the bombs could have been expected to keep falling in an area where they were already falling (so that the government’s intervention would have kept things going as they were already going, and probably also as they would have kept going if the government had decided that it cannot get involved), whereas if the Germans would have been lied to, the bombs would have started to fall in an area where they had not so far been falling (and where they probably would not have started to fall in the absence of intervention). In other words, it seems likely that relative to a ‘no involvement’ baseline, redirection would have resulted in *different* people dying, whereas confirming that the bombs were hitting Central London would have resulted in the *same* people dying. While I have a hunch that some people will feel that this morally speaks against redirection, these people would have to explain why we should think of the ‘no involvement’ baseline as morally salient in the case at hand. Suppose a Minister of the War Cabinet argued as follows: “We should use the ‘no involvement’ baseline to spell out the distinction between actively inflicting harm and merely allowing it to occur. If we do this, we get the result that we would *actively harm* some of our civilians if we redirected the bombs away from Central London. Why? Because we would make some of our civilians worse off than they would have been had we decided that we must not get involved in the bombings. Contrariwise, we get the result that we would *merely allow harm to occur* to some of our civilians if we made our double agents tell the truth and had them confirm that the German bombs are hitting Central London. Why? Because we would make none of our civilians worse off than they would have been in the absence of our involvement, or if we had instructed our double agents to stop communicating with the Germans.”

But why should the ‘no involvement’ baseline allow us to distinguish between actively inflicting harm and merely allowing it to occur *in such a way that the former is somehow intrinsically morally worse than the latter*? Even in Glover’s fictional case, the British authorities did not seriously consider not getting involved at all, and rightly so. They were discussing only whether to strategically deceive the enemy, or

whether to tell him the truth. Relative to the option of strategically deceiving the enemy, it was an expected consequence of the decision to tell the truth that a group of civilians would die that would not otherwise have died. Relative to the option of telling the truth, it was an expected consequence of the decision to deceive the enemy that a different group of civilians would die that would not otherwise have died. It is difficult to see why ‘no involvement’ should be thought of as a significant baseline when under the circumstances of the case, ‘not getting involved’ was not a serious contender for the morally right thing to do.¹⁵

Does the distinction between actively inflicting harm and merely allowing it to occur do any work once we leave behind Glover’s misrepresentation of the historical details? If it were true that actively inflicting harm was worse than merely allowing it to occur, would this in some way speak in favour of confusing the enemy? I do not think so. It does not seem to me that compared to redirection, confusing the enemy was in one way or another a more passive strategy. Whichever strategy was adopted, the double agents had to be provided with similarly detailed instructions about what to report back to their German contacts. Also, whichever strategy was adopted, relative to the other strategy (as well as relative to a ‘no involvement’ baseline, where all communication with the Germans via the double agents was discontinued), it was reasonable to expect that some civilians would die who would not otherwise have died.¹⁶

What about the consideration that with the confusion strategy, the decision-makers were unable to anticipate which part of Greater London would have to bear the brunt of the attack? Does this render the confusion strategy more passive than the redirection strategy? If not knowing who your victims will be makes a harmful act more passive, then there is only a difference in degree between confusion and redirection. Whichever strategy was adopted, the decision-makers did not know

¹⁵ In section 4.6.2, I will return to the ‘no involvement’ baseline, and I will suggest (following Kamm) that it helps us spell out the distinction between the negative and the positive aspect of the right to life.

¹⁶ If the British authorities had adopted Findlater Stewart’s second proposal to keep the MPI at its current position in Dulwich, it would have been reasonable for them to expect that no civilians would be harmed by the German bombs who would not also have been harmed under a ‘no involvement’ scenario. In section 4.6.2, I will discuss whether there is something morally significant to this. See also footnote 15.

precisely which individuals would die as a result of it; with the redirection strategy, they simply had a clearer idea of the risks that they were exposing different parts of the population to. But I do not actually think that harming someone is a more passive act to the extent that you do not know whom you are harming. Suppose I have a choice between stabbing a victim whose identity is revealed to me, or stabbing a victim who is presented to me behind a curtain. Not only do both options strike me as similarly horrific, but it also does not seem to me that the second option involves a more passive way of harming someone than the first option does.

4.4.2 The Doctrine of Double Effect

Fischer and Ravizza (1992, p. 18) suggest that the Doctrine of Double Effect (DDE) captures some of the moral issues underlying the historical case as misrepresented by Glover.

They describe the DDE as follows (Fischer and Ravizza, 1992, p. 6, original emphases):

The **Doctrine of Double Effect** distinguishes between intended *ends* and necessary *means*, on the one hand, and unintended but foreseen *side-effects* on the other. The doctrine exploits this distinction to claim that whereas it is sometimes permissible to bring about a bad result as a merely foreseen side-effect of what you do, it is in general not permissible to bring about such an effect as an intended end or a necessary means to some intended end.

Fischer and Ravizza (1992, p. 18) then argue that according to the DDE, “the morally salient feature of the English decision was whether the deaths of the innocent citizens would have been either a necessary means or intended end of the missile diversion, or whether these deaths would have been merely a side-effect that was foreseen but unintended.” They think that there is reasonable disagreement on this issue (*ibid.*), so that the DDE can explain the disagreement that existed between the deception authorities who were in favour of redirection, and the Ministers who were against it.

Given Glover’s misrepresentation of the case, I do not see how the DDE can

explain the Ministers' resistance to redirection. If—as Glover implies—the Germans could simply have been lied to and fake incidents could have been reported to them in a credible way, I do not see how according to any interpretation of the terms, the deaths of innocent civilians were either a *necessary means* or else an *intended end* of the redirection strategy (see also Moore, 2009, p. 76). Surely everyone in favour of redirection would simply have been glad if the redirected bombs happened to consistently fall on uninhabited territory? Given Glover's misrepresentation of the case, civilian deaths seem to be no more than a foreseeable side effect of both the redirection strategy and the strategy to tell the truth.¹⁷

Once we consider the historical case as it actually presented itself, things become more complicated. Remember that Findlater Stewart wanted to safeguard the excellent reputation of the British double agents, and for this reason insisted that only actual bomb incidents should be reported. Now if the enemy ought to be induced to believe that he was at first hitting north-west London on average, and that he was hitting Central London on average after he had adjusted his aim, there had to be actual bomb incidents in north-west London and in Central London. But, it could be argued, you cannot realistically have a series of bomb incidents in north-west London and in Central London without having any dead civilians. If this is right, then it would seem that under the circumstances, the deaths of innocent civilians were *inextricably tied to a necessary means* of the redirection strategy, and in that sense formed a *contingently necessary* part of the redirection strategy. Similarly for Stewart's second proposal to fix the MPI at its current location. To fix the MPI at its current location, the enemy had to be induced to believe that he was hitting Central London on average. If only actual bomb incidents were to be reported, then there had to be some bomb incidents in Central London. But, it could be argued, you cannot have bomb incidents in Central London without having any dead civilians. Hence dead civilians were *contingently necessary* for Stewart's second proposal as well.

¹⁷ In the 'side track' trolley case, it is usually thought that the DDE would *not* rule it impermissible to divert a runaway trolley onto a side track (see e.g. Otsuka, 2008, pp. 93-4). Insofar as Glover's description of the British decision situation describes that situation as structurally similar to a 'side track' trolley case, we should thus not be surprised that the DDE does not speak against redirection.

By contrast, while it is not clear how the confusion strategy would have been implemented, it is certain that the double agents could not simply have reported a number of arbitrarily selected actual bomb incidents. As Stewart plausibly asserted (CAB 113/35, 31 July 1944), haphazard reports would not have confused the enemy; instead the Germans would have calculated an MPI from whatever bomb incidents that were reported back to them. To put the confusion strategy into practice, the double agents would thus for example have had to contradict each other, or to claim that they had left London, or to provide only very vague information (see section 4.3). At least with some adjustments, it should thus have been feasible to implement a confusion strategy entirely without reliance on any actual bomb incidents. Hence the confusion strategy did not *depend* on actual bomb incidents, and it is reasonable to assume that it would not as a matter of fact have made heavy use of them. With respect to the confusion strategy, bomb incidents in the Greater London area were thus more like a foreseeable side effect than a necessary means. Hence the DDE would seem to speak in favour of the confusion strategy over Stewart's deception proposals.

How successful is this line of argument? I believe that there is a decisive objection to it. The DDE means to rule out certain ways of *bringing about* bad results (Fischer and Ravizza, 1992, p. 6; see also McIntyre, 2011). In essence, it objects to intending bad results (McIntyre, 2011). But for each bomb incident that Stewart's deception proposals would have made use of, this bomb incident would *already have happened* by the time the deception authorities would have made use of it. To see what I mean, suppose the British authorities decided that the redirection strategy should be implemented. They would then have needed to report a bomb incident somewhere in north-west London. But they would *not* consequently have brought about a first such incident (and its accompanying harm) as a means to implement the redirection strategy and minimize civilian losses. Rather, they would either have reported a suitable incident that had already occurred, or they would have waited for such an incident to occur. For each bomb incident that the redirection strategy would subsequently have helped to bring about, this incident would have been brought about as a foreseeable side effect of the redirection strategy, even if it was afterwards used as a means to keep the strategy going. In other words, Findlater Stewart's

proposed deception strategies were built around the assumption that bomb incidents in north-west London and in Central London *would* materialize, and his strategies remained feasible only as long as such incidents *kept* materializing. But whatever information was passed on to the Germans was passed on to them in order to keep the weight of the attacks *away* from north-west London and Central London, not to cause incidents in those areas. Once we consider the matter carefully, the DDE does thus not speak against Findlater Stewart's proposals.

Kamm (2007, p. 101) argues that we sometimes pursue a course of action only because we believe that it will bring about a bad result, so that the bad result figures as a "condition of [our] action" (Kamm, 2007, p. 118). To Kamm, acting *only because* a bad result will thereby be brought about has to be distinguished from acting *in order to* bring that result about, as only the latter necessarily involves intending the bad result, and is in that sense especially morally problematic (Kamm, 2007, pp. 101-4).¹⁸ If a bad result figures as a condition of our action, we would not act unless we thought that we will thereby bring it about. This fact makes acting only because a certain result will be brought about similar to acting in order to bring that result about. But when we act only because we believe that a bad effect will be brought about, there is no rational requirement that commits us "to doing other things to bring [the bad effect] about" (Kamm, 2007, p. 101), whereas there is such a requirement when we intend the bad effect and act in order to bring it about.

I believe that the way in which report-worthy bomb incidents figured in Stewart's deception proposals fits Kamm's description of bad results that are a condition of action. Neither of Stewart's proposals would have been put into practice if the British authorities had not believed that putting it into practice would lead to report-worthy bomb incidents. But adopting either proposal would nevertheless not have rationally

¹⁸ As an example of acting only because a bad result will thereby be brought about, consider Kamm's *Party Case*: "I intend to give a party in order for me and my friends to have fun. However, I foresee that this will leave a big mess, and I do not want to have a party if I will be left to clean it up. I also foresee a further effect of the party: If my friends have fun, they will feel indebted to me and help me clean up. I assume that a feeling of indebtedness is something of a negative for a person to have. I give the party because I believe that my friends will feel indebted and (so) because I will not have a mess to clean up. These expectations are conditions of my action. I would not act unless I had them. [...] But I do not give the party even in part in order to make my friends feel indebted nor in order to not have a mess" (Kamm, 2007, p. 95).

committed the British authorities to do “other things” to help bring about report-worthy bomb incidents.

According to Kamm’s revised version of the DDE—i.e. her Doctrine of Triple Effect, or DTE—there is no especially strong presumption against pursuing a greater good in such a way that one acts only because one believes that some lesser evil will thereby be brought about (Kamm, 2007, p. 118). Like pursuing a greater good while foreseeing that some lesser evil will thereby be brought about, such action is sometimes morally justifiable (*ibid.*).¹⁹

If all of this is right, then DDE-based considerations do not provide a valid reason in favour of the Ministers’ confusion strategy over the deception proposals brought forward by Findlater Stewart. We can grant that with regard to the confusion strategy, bomb incidents in the Greater London area would have been a foreseeable side effect, whereas with regard to Stewart’s deception proposals, at least some bomb incidents would have been a necessary condition for the implementation of these proposals. But neither the DDE nor Kamm’s DTE tell us that it is worse to bring about bad effects that are a condition of action than it is to bring about foreseeable bad side effects.

In the next section, I argue that a valid reason against Findlater Stewart’s deception proposals can be found once we turn our attention to considerations of equality.

4.5 Showing Equal Concern

Each state has a responsibility to protect the lives of its citizens. This responsibility is fundamental: a state that does not adequately meet it is not a legitimate state (see e.g. Hobbes, 1994, ch. 21; Locke, 1990, ch. IX; Mill, 1998a, ch. 1). That there *exists* such a fundamental responsibility is not a contested claim. The contested issue is what this responsibility *amounts to*, or how it is *properly fulfilled*. In the case at hand, the British authorities did not doubt that they had a duty to protect their citizens; the question they were struggling with was how to live up to it in light of

¹⁹ Kamm (2007, pp. 121-2) does not endorse her DTE “as an adequate principle of permissible harm.” She merely thinks that the DTE constitutes a substantial improvement over the DDE that captures DDE-related considerations more accurately than the DDE itself does.

their citizens' competing claims to protection. Given the circumstances of the case, no matter what the British authorities decided to do, a significant number of innocent civilians was bound to be killed by the enemy. How should a state protect its innocent civilians when it cannot protect them all?

One answer to this question is that it should *minimize losses*. This answer has considerable intuitive appeal. If a state cannot protect the lives of all of its innocent civilians, what could possibly be required of it but to protect as many civilian lives as it is able to? If minimizing losses is the way to go, then it was wrong of the Ministers to oppose Findlater Stewart's redirection strategy. In a report compiled by Duncan Sandys—Chairman of the War Cabinet's 'Crossbow' Committee for the defence against V-weapons²⁰—it was estimated that (CAB 113/35, 2 August 1944):

- Moving the MPI from its current position in Dulwich six miles to the south-east would decrease average monthly bomb-related deaths by 1,600;
- Moving the MPI from its current position in Dulwich to Charing Cross (several miles to the north-west) would increase average monthly bomb-related deaths by 500.²¹

Compared to the redirection strategy, the confusion strategy thus had the potential to result in an average of up to 2,100 additional civilian deaths per month.²² If Findlater Stewart was operating under the assumption that it was the British authorities' task to win the war while minimizing civilian losses, it is no wonder that he grew exasperated when the War Cabinet's Ministers argued that confusion should be preferred over redirection.

²⁰ Duncan Sandys was the elected Tory MP for Norwood in South London. Winston Churchill, who was Duncan Sandys' father-in-law, made him the Chairman of the War Cabinet's 'Crossbow' Committee for the defence against German flying bombs and rockets when that Committee was set up in February 1943 (Campbell, 2012, pp. 61-2).

²¹ With regard to bomb-related serious injuries, it was expected that moving the MPI six miles to the south-east would lead to 4,600 fewer serious injuries each month, whereas a shift of the MPI to Charing Cross would lead to 1,600 additional serious injuries each month (CAB 113/35, 2 August 1944).

²² The confusion strategy also put the double agents' reputation with their German controllers at risk, and it exposed Central London's infrastructure to an increased risk of destruction. Given that both the double agents and Central London's infrastructure served as means to the protection of civilian lives, it was reasonable to expect that even more civilians would die if these means were compromised.

But there is a second answer to the question of how a state should protect its innocent civilians when their claims to protection compete. According to this second answer, a state should minimize losses to civilian lives *subject to an equality constraint*, or only to the extent that this can be done while showing a sufficiently balanced concern for the safety of all. The basic idea behind this second answer is that a state owes its duty of protection equally to all of its citizens. Only if the protective measures that it implements express a sufficiently equal concern for the safety of each of its citizens do they properly take into account that in a legitimate state, all citizens enjoy an equal standing (Rawls, 2001, p. 18; see also Pettit, 2012, pp. 77-81, and Locke, 1990, ch. IX).

If the British authorities had chosen to redirect the German bombs, they would have knowingly increased the risk of a bomb attack for some parts of its population in order to decrease that same risk for other parts of its population. It would thus appear that redirection was a protective measure that did not express an equal concern for the safety of all, as implementing it would have meant favouring some citizens over others. The same cannot be said of the confusion strategy. If the British authorities had chosen to confuse the enemy, they would merely have tried to keep their double agents in play while not favouring any part of their population over another. I thus suggest that the Ministers' instruction to create confusion sprang from a legitimate concern with equality.

This interpretation fits well with Morrison's alleged refusal to 'act as God'. It is a common adage that only God may decide who should live and who should die, and that it is not for any human being to make such a judgement. To protect everyone equally is to avoid judging. But I am not ultimately trying to find out whether the Ministers were *actually* moved by considerations of equality. What matters to me is whether considerations of equality provided a *valid* reason in favour of confusion over redirection, so that Findlater Stewart was wrong not to have any sympathy for the Ministers' opposition to the redirection strategy.

4.5.1 Did Redirection Show Equal Concern?

According to a view that I will refer to as ‘utilitarian’, implementing the redirection strategy would *not* have shown unequal concern. The utilitarian view centers around the idea that minimizing losses *coincides* with showing proper and equal concern for each citizen (see Mill, 1998b, ch. 3). An advocate of this view might defend it as follows: “At first sight, the redirection strategy may indeed appear to express unequal concern, as implementing it protects the people living in North London at the expense of those living in South London. But first appearances are deceptive in this case. Public decision-makers show an equal concern for the safety of all of their citizens if they assign the same positive value to each civilian life in their calculation about the right thing to do. If they proceed in this way, they assume that no one’s life counts for more than anyone else’s. This is what makes their stance impartial. From a suitable perspective—let’s say before the war started, or if citizens were deprived of their knowledge about where they lived—a general policy to minimize losses might moreover be in everyone’s best (expected) interest. This lends further support to the idea that minimizing losses coincides with showing proper and equal concern for the safety of all of one’s citizens.” The problem with this utilitarian view is that its implications about specific cases do not fit with our considered judgements. Suppose that an evil state E credibly threatens to terror bomb a number of major cities in state S unless S hands over one hundred of its citizens for E to execute. The statement that S were to treat all of its citizens equally if it followed E ’s orders is unacceptably counterintuitive. On top of that, it objectionably downplays the sacrifice that would be required of the one hundred if S were to turn them over to E (see e.g. Nozick, 1974, pp. 32-3). This provides sufficient reason to reject the utilitarian view.

Further above, I have argued that if the British authorities had implemented the redirection strategy, they would have knowingly increased the risk of death and injury for some parts of its population in order to decrease that same risk for other parts of its population. I have suggested that it is this feature which gives the redirection strategy the air of a protective measure that does not express equal concern for all citizens. In the example just given, if S ’s government were to hand over one hundred

of its citizens, it would similarly increase their risk of death in order to decrease that same risk for the remaining parts of its population. This suggests that a protective measure—i.e. a governmental intervention that reduces the average risk of death and injury²³ among the population—does not express equal concern for all citizens if it increases the risk of death and injury to some citizens in order to decrease that risk for others. Is this correct?

It is not. Consider the following protective measures that the British authorities implemented once the terror bombing of London started (CAB 121/213, 1944):

1. Sending out fighter aircraft to shoot down V-weapons;
2. Putting up barrage balloons to block and damage V-weapons;
3. Shooting down V-weapons from the ground with anti-aircraft guns.

All of these protective measures were implemented in (or over) sparsely inhabited parts of South East England. They therefore increased the risk of death and injury for the people who lived in these parts of the country. Moreover, they all aimed to reduce the overall deadliness of the German bombs by keeping some of them from reaching the Greater London area. It follows that all of these measures knowingly increased the risk of death and injury to some parts of the population in order to decrease that risk to others. The measures nevertheless do not strike us as expressing unequal concern. Why not? I believe that they *rightly* do not strike us as expressing unequal concern for the following reason. Both before and after their introduction, the brunt of the attacks was borne by the people who lived in the Greater London area. Each civilian who lived in the Greater London area was at a much higher risk of death and injury than those who lived somewhere else. Putting up barrage balloons, sending out fighter aircraft, and shooting down V-weapons from the ground all mitigated the relatively high risk that Londoners were exposed to. They therefore protected those civilians to whom the threat to life was most severe. They did this by knowingly

²³ Throughout this chapter, I think of the average risk of death and injury among the population, the risk of death and injury for individual persons, and the change in these risks if some governmental measure were adopted in terms of *estimates that it would have been reasonable to come up with at the time of decision-making given the then available information*. Whether some governmental measure is protective or whether it expresses an equal concern for the safety of all citizens is thus understood roughly in what Parfit calls an “evidence-relative sense”. See Parfit, 2011, p. 150.

increasing the risk of death and injury to civilians who were at a relatively low risk to start with, and who remained at a significantly lower risk than those living in the Greater London area even after the introduction of these measures.²⁴ Figuratively speaking, these measures obligated British citizens who lived in sparsely inhabited parts of South East England to step up and share in the burden that their fellow citizens in London were shouldered with. If we interpret the measures in this way, then it seems that they made reasonable demands on the solidarity of the people living in rural South East England: the measures did not put these people at a high risk of death and injury in absolute terms, and they burdened each person living in that area with only a fraction of the risk that each Londoner was burdened with. Their logic was thus to spread the burden of the German bombs from the most affected areas to a number of less affected areas while at the same time significantly reducing the average risk of death and injury among the entire population. It seems to me that a protective measure which makes reasonable demands on the solidarity of some citizens for the benefit of others is *not* thereby a measure that expresses unequal concern. When we show solidarity with others, we recognize them as our equals. A protective measure that makes reasonable demands on the solidarity of some citizens is thus very much in line with the idea that citizens are each other's equals, and that they ought to be treated as such. If this is right, then it follows that a protective measure need not show unequal concern even if it increases the risk of death and injury to some in order to decrease that risk for others.²⁵

²⁴ This is my conjecture; I have not been able to find any relevant numerical comparisons. In the literature, there are two main criticisms of countermeasures 1, 2, and 3: that they put soldiers' lives at too great a risk (especially measure 2), and that they were not very cost-effective (especially measures 1 and 3). See e.g. Campbell, 2012, pp. 312-4; p. 446. Based on this, I conjecture that very few—if any—civilians were killed or injured through countermeasures 1, 2, and 3, and that this is something the British authorities reasonably expected when they implemented these measures.

²⁵ 'Increasing the risk of death and injury to some in order to decrease that risk for others' is not only not a sufficient condition for a protective measure to express unequal concern, it is also not a necessary condition. Suppose that an evil state E has kidnapped one hundred citizens of state S . E announces that it will execute them publicly the next day. Suppose S can send in a Special Forces team to save 20 of its citizens with near certainty. Alternatively, it can send in a Special Forces team that tries to save all hundred, in which case S can expect to save one hundred with probability 0.2, and no one with probability 0.8. If S chooses the first alternative, then it fails to show an equal concern for the safety of all of its citizens despite the fact that its actions do not increase the risk of death to anyone. See also my discussion of the unequal concern that would be expressed by 'keeping the MPI in Dulwich' in the last paragraph of this section.

The logic behind the redirection strategy was different from the one just discussed. In essence, to redirect the bombs was to single out a specific part of the population and to make that part bear the brunt of the attack, thereby significantly reducing the average risk of death and injury among the entire population. The redirection strategy did not spread the burden from the most affected areas to a number of less affected areas. Instead it determined what should henceforth be the most affected area. Figuratively speaking, the redirection strategy did not ask for the *solidarity* of those living in south east London; it asked for their *sacrifice*. That is, it did not ask those living in south east London to step up and share in the burden; it more or less made the burden theirs. The redirection strategy therefore expressed unequal concern not because it increased the risk to some in order to decrease the risk for others, but because it singled out one group of citizens to bear the brunt of the attack.

If this is right, then considerations of equality spoke in favour of confusion over redirection: adopting the redirection strategy would have meant showing unequal concern, and this provided a reason against adopting it.²⁶ Did they also speak in favour of Stewart's second proposal—to keep the MPI where it is—over redirection? Did keeping the MPI in its current position express a less objectionably unequal concern than redirecting the German bombs further to the southeast of the Greater London area? It seems to me that compared to redirection, keeping the MPI in Dulwich was a similarly problematic intervention.²⁷ Admittedly, it did not *increase* the risk of death and injury to the people living in Dulwich, but it made sure that they would keep suffering the brunt of the attack. To me, making sure that some group keeps suffering the brunt of an attack is just as much a way of showing unequal concern as is making sure that another group will henceforth suffer the brunt of that

²⁶ I don't think that a government is *always* required to adopt a protective measures *only if* it expresses a sufficiently equal concern for all citizens. Depending on the available alternatives, adopting a measure that expresses unequal concern may well be the right thing to do. See sections 1.3 and 4.7.

²⁷ On my definition of the term, keeping the MPI in its current location was not a *protective* measure, as implementing it would not have reduced the average risk of death and injury among the population. Compared to not doing anything, keeping the MPI in Dulwich was simply a *prudent* measure that kept the British double agents in play and protected Central London's infrastructure.

attack.²⁸ But then why did Findlater Stewart come up with his second proposal? If it was just as unsuccessful at showing equal concern as his first, what made it morally attractive? After all, compared to keeping the MPI in Dulwich, redirecting the bombs was expected to result in 1,600 fewer civilian casualties *month after month* (CAB 113/35, 2 August 1944). Surely this spoke heavily in favour of redirection! In the next section, I offer a possible explanation of the moral appeal of Stewart's second proposal.

4.6 The Right to Life

So far, I have argued that when citizens' claims to protection compete, state actors should not under all circumstances try to minimize civilian losses, as they have a general duty to show a sufficiently equal concern for the safety of all of their citizens. In this section, I introduce a further constraint on minimizing losses: I argue that state actors must not minimize losses to civilian lives if doing so would violate some of their citizens' right to life.

4.6.1 Negative and Positive Aspects of the Right to Life

One way of thinking about a state's responsibility to protect the lives of its citizens is in terms of rights. According to such a rights-based view, each citizen has a *right to life against the state*. This right is taken to have both a *negative* and a *positive* aspect. Negatively, the state must not kill or injure its citizens without due process (Griffin, 2008, p. 212). Positively, it must protect its citizens against being killed or injured by other agents (Locke, 1990, ch. IX). The negative and positive aspects of different citizens' rights to life can come into conflict. Suppose that state *S* could save

²⁸ One could try to argue that keeping the MPI in Dulwich was a way of showing equal concern in the sense that *had the bombs fallen somewhere else, the MPI would have been kept in place in that other location*. For two reasons, I don't think this argument works. First, the deception authorities and the Ministers were willing to keep the MPI where it was only because it happened not to be in Central London. So it is not true that the government was adhering to a general 'laissez-faire' policy with respect to the bombs' MPI. Second, where governments have positive duties of provision—e.g. in the protection of their citizens from outside attack—a 'laissez-faire' policy seems to me simply a *failure* to provide some good, not a way of providing that good in a way that shows equal concern.

two of its citizens from being killed by an evil state *E* if it executed a third citizen. Should *S* protect the lives of citizens one and two or should it refrain from depriving citizen three of his life?

According to the rights-based view, the negative aspect of the right to life is significantly weightier than its positive aspect, so that there is a strong presumption against harming one's citizens even if doing so would be necessary to protect more numerous others from being harmed. The idea behind strong negative rights is that, contra Delmer (1971, p. 207), a citizenry is *not* like an army where it is usually right to "put a batallion at risk to protect a brigade." Rather, the idea is that each citizen deserves recognition as a separate person with "an inviolability founded on justice that even the welfare of society as a whole cannot override" (Rawls, 1971a, p. 3). The rights-based view stresses the fact that everyone has aspirations and interests of their own, and based on this argues that each citizen should have a general right against the state not to have his interests sacrificed for the interests of others. If state actors kill or injure some of their citizens, or if they expose them to an increased risk of death and injury, they *infringe* their citizens' rights to life. Under certain circumstances, such infringements are justified. Justified infringements are morally permissible; in the face of a state's duty to protect its citizenry, they may even be required. When a state infringes its citizens rights without sufficient justification, it *violates* its citizens rights to life: it *infringes them impermissibly*.²⁹

The logic behind stringent negative rights differs from the idea that a government ought to show a sufficiently equal concern for the safety of all of its citizens. Whereas the latter speaks against protective measures that make a specific group bear the brunt of an attack, the former introduces a general presumption against the redistribution of harm. If the War Cabinet's Ministers had been worried primarily about not infringing their citizens' rights, they might have opposed protective measures such as anti-aircraft guns or barrage balloons on grounds that these would increase the risk of death and injury among the people living in rural South East England. But if anti-

²⁹ In the context of this chapter, I will talk about *infringements* and *violations* of the right to life only with regard to its *negative aspect*. I reserve these terms for its negative aspect because my main interest lies with the constraints that this negative aspect of the right to life imposes on the protective measures that a government may permissibly adopt. This is not to rule out that similar concepts are applicable with respect to the positive aspect of the right to life.

aircraft guns and barrage balloons were reasonably safe, and if it could be expected that using them would significantly reduce the amount of harm that Londoners had to suffer, then the British government's duty to treat its citizens equally would not only have failed to speak against them; it would actually have spoken clearly in their favour.

None of this is to suggest that stringent negative rights rule out as impermissible most protective measures that considerations of equality would speak in favour of. Instead, it seems plausible to assume that considerations of equality will often provide the grounds for a *principled exception* to the general rule that each citizen has a right not to have his interests sacrificed for the interests of others. Further above, I have argued that a protective measure expresses equal concern for the safety of all citizens if it spreads the burden of harm from those who have to bear the brunt of it to those who are not as affected by it. Even if I am an individual with interests and aspirations of my own, I can still recognize that I should accept a setback to my interests if the sacrifice required of me is small, the benefit created to others is large, and those whom I benefit are my equals whose luck has been worse than mine. By contrast, if a protective measure shows unequal concern by singling out a group of citizens to bear the brunt of an attack, it makes the singled out group worse off than other citizens for the sake of those other citizens. This seems much harder to reconcile with the idea that each citizen has "an inviolability [...] that even the welfare of society as a whole cannot *override*" (Rawls, 1971a, p. 3, emphasis added), as it *submits* the fundamental interests of the individuals in the singled out group to the welfare of society as a whole. If this is right, then implementing a protective measure that shows equal concern, but that is expected to cause harm to some citizens, will often *merely permissibly infringe* these citizens' rights, whereas a protective measure that shows unequal concern and that is expected to cause harm to some citizens will frequently *violate* their rights.³⁰

The redirection strategy was a protective measure that showed unequal concern and that would have infringed the rights of the people who lived south-east of Dulwich.

³⁰ A government that implements a protective measure that shows equal concern and that permissibly infringes some citizens' rights may still owe compensation to the citizens whom it ends up harming. But this does not mean that it acted wrongly when it implemented the measure.

Was it this combination of factors that made the Ministers balk at it? Did they fear that because the redirection strategy combined these two factors, implementing it would probably violate some of their citizens' rights to life? If yes, then I think that the Ministers were right to be worried about violating their citizens' rights. But it is not clear that they could entirely have circumvented this worry by adopting either the confusion strategy, or the strategy to keep the MPI in Dulwich. Consider the confusion strategy first. Admittedly, this strategy did not express unequal concern. But it was likely to increase the risk of death and injury to *some* British citizens. It thus infringed some citizens' rights, and may therefore have violated them. Bear in mind that the confusion strategy was not a protective measure: it was not expected to reduce the average risk of death and injury among the population. Nor was it expected to redistribute the burden of the bombs from some of the more affected areas to some of the less affected areas. Rather, it was expected to reshuffle the burden of the bombs in an unpredictable way, possibly increasing the average risk of death and injury among the population, possibly decreasing it. In light of these facts, it is not clear that the British authorities could have justified the death and injury that adopting the confusion strategy would have caused. An average of up to 2,100 additional civilian deaths each month³¹ (CAB 113/35, 2 August 1944) might simply have been too high a price to pay for avoiding the expression of unequal concern. Unless the British authorities were *convinced* that confusing the enemy was *in any case* the right thing to do, they thus had to worry about violating some of their citizens' rights by adopting the confusion strategy. In sum, while considerations of equality spoke clearly in favour of confusion over redirection, both confusion and redirection would have infringed people's rights, and may thereby have violated them.

What about keeping the MPI in Dulwich? Would this alternative have not infringed—and, hence, definitely not violated—citizens' rights?

³¹ As well as putting Central London's infrastructure at risk and jeopardizing the reputation of the British double agents.

4.6.2 The Significance of the ‘No Involvement’ Baseline

So far, I have suggested that state actors infringe their citizens’ rights to life if they kill them, injure them, or if they expose them to an increased risk of death or injury.³² This is but a rough-and-ready description of what it means for a state actor to infringe his citizens’ right to life. Understood as an enumerative definition, its ambiguous terms are in tricky cases likely to raise more questions than they are able to answer. Alas, we may not easily be able to do much better. To keep the negative aspect of the right to life neatly apart from its positive aspect, a clean distinction between *doing harm* and *not protecting someone from harm* is required. But drawing such a distinction is a notoriously difficult task, and there is no generally agreed-upon way of going about it that we could employ for our purposes.

To get a flavour of the intricacies involved in this issue, consider Stewart’s proposal to keep the MPI in Dulwich. Would adopting this proposal have infringed citizens’ rights? Would it have *done harm* to some citizens? On the one hand, one would think that it would not have. Keeping the MPI in Dulwich was expected not to lead to any bomb-related deaths or injuries that would not also have occurred under a scenario where the British government discontinued all of its communication with the Germans via the double agents once the bombings started. Now if the government had chosen to discontinue communicating via the double agents once the bombings started, it seems that it would not have infringed any of its civilians’ rights, as rights infringements—or doing harm—presuppose activity.³³ But if a ‘no involvement’ strategy would not have infringed anyone’s rights, and if ‘keeping the MPI in Dulwich’ was not expected to make anyone worse off than ‘no involvement’ would have, then surely ‘keeping the MPI in Dulwich’ did not infringe anyone’s rights either? On the other hand, there seems to be a clear sense in which keeping the MPI in Dulwich would not have merely allowed the people in Dulwich to be hit by bombs. ‘Keeping the MPI in Dulwich’ was not a ‘look the other way’ strategy. Rather, it required purposeful and continuous governmental intervention. If I ensure that someone will be killed by channeling harm towards that person, my actions help bring about that person’s death in a perspicuous

³² See footnote 23.

³³ Maybe it would have violated its duties of protection; but that is another issue. See footnote 29.

way; I do not merely fail to protect that person. Surely a government infringes its citizens' rights if it *makes an effort to channel harm towards them?*

Kamm's discussion about the difference between killing and letting die provides a way out of this muddle. According to Kamm (1996, p. 31), it is a necessary feature of letting someone die that the person whom an agent lets die is not made worse off relative to a counterfactual situation in which we think away the agent. To Kamm (*ibid.*), this feature—this “definitional property” of letting someone die—mitigates the badness of letting someone die. In killing cases, this feature may or may not be present. When it is present, it mitigates the badness of the killing in question (just as it always attenuates the badness of letting someone die). But when it is not present—and where other factors are held constant—an instance of killing is worse than an instance of letting die.

Applied to the historical case, Kamm's ideas can be translated as follows. Even if 'keeping the MPI in Dulwich' infringed the rights of the people whom it channelled harm towards, the badness of these infringements was mitigated by the fact that they shared a significant structural feature with what would have been a government's mere failure to protect its citizens. Just like a mere protective failure, keeping the MPI in Dulwich would not have made anyone worse off than they would have been in the absence of governmental involvement.

It is important to bear in mind that Kamm proposes a feature that makes certain ways of letting harm come to people *less bad than others*, not a feature that makes certain ways of letting harm come to people *not so bad simpliciter*. Depending on the circumstances, not protecting one's citizens may be a very bad failure indeed. If the British authorities could safely have redirected the German bombs towards uninhabited territory, but failed to go ahead with this because they were preoccupied with building extravagant bomb shelters for themselves, they could not later have argued that their failure was “not so bad” because “no one was hurt who would not have been hurt in our absence anyway.”

One way of understanding Kamm's mitigating feature is that *other things equal*, its presence gives us sufficient reason to adopt one measure over another. Suppose that when the terror bombing of London started, the British authorities had a choice

between two similarly prudent deception strategies. On the one hand, they could have had their double agents pass on information that would have kept the MPI in Dulwich. Alternatively, they could have had them pass on information that would have shifted the MPI to Wimbledon. Further suppose that Dulwich and Wimbledon were similarly densely populated. ‘Keeping the MPI in Dulwich’ and ‘shifting the MPI to Wimbledon’ would then have been two similarly prudent measures showing similarly unequal concern that would both have kept the number of expected casualties constant. Following Kamm, it would then have been preferable to keep the MPI in Dulwich as opposed to shifting it to Wimbledon, as keeping the MPI in Dulwich would not have made anyone worse off than they would have been in the absence of governmental intervention. From a rights-based perspective, this idea makes sense. Compared to ‘shifting the MPI to Wimbledon’, ‘keeping the MPI in Dulwich’ has a feature that makes it less clearly a case of harmful interference and structurally similar to a mere protective failure. If we accept that other things equal, the negative aspect of the right to life takes precedence over its positive aspect, keeping the MPI in Dulwich would thus have been the more preferable option. What sort of rights-based justification could the British authorities have offered for keeping the MPI in Dulwich as opposed to shifting it to Wimbledon? They might have argued as follows. ‘It is true that by actively keeping the MPI in Dulwich, we channel harm towards the people living in Dulwich. This would usually be objectionable. But under the circumstances of the case, it seems that the people in Dulwich do not actually have a justified complaint. They could try to object to our *active interference*. But if we refrained from intervening, they would have to suffer the same amount of harm anyway—the only difference is that we would risk losing our valuable double agents! Hence there is no valid objection to the ‘active interference’ part of ‘keeping the MPI where it is’. Could the people in Dulwich demand that we should redirect the bombs to Wimbledon? Admittedly, if we redirected the bombs to Wimbledon, we would protect the people in Dulwich. This counts in favour of redirection. But for every civilian that we would protect in Dulwich, we would put one at risk in Wimbledon. This sort of exchange is precisely the thing that we are not entitled to authorize. As public decision-makers, we must not see our citizens as interchangeable. We have to

take seriously the fact that each citizen is a separate person with a life of his own, not a stand-in for every other citizen. Of course we have to try and do our best to relieve the suffering in Dulwich. We have to build more bomb shelters in the area, and we have to improve the local early warning systems. We have to locate the launch sites for the V-weapons, and destroy them. But we must not shift the primary burden of the bombs from one area to another.”

In the actual historical case, *other things* between the different alternatives were *not equal*. Even if its mitigating feature spoke in favour of keeping the MPI where it is (as I think it did), it is thus not at all clear that ‘keeping the MPI where it is’ would have been the right deception strategy to adopt. To recapitulate, I have argued for the relevance of the following considerations:

- **Prudence.** *Keeping the MPI in Dulwich* and *Redirection* were both prudent strategies. By contrast, *Confusion* would have jeopardized the reputation of the British double agents, and it would have put Central London’s infrastructure at risk.
- **Effectiveness.** Compared to *Redirection*, *Keeping the MPI in Dulwich* was estimated to result in an average of 1,600 additional bomb-related civilian deaths each month. *Confusion* was estimated to result in an average of *up to* 2,100 additional deaths each month.
- **Equal Treatment.** *Keeping the MPI in Dulwich* and *Redirection* both expressed unequal concern. They singled out a specific part of the Greater London area to bear the brunt of the attack. *Confusion* avoided the expression of unequal concern, but it was not expected to spread out the burden of the bombs more equally among the population.
- **Absence of rights violations.** *Redirection* and *Confusion* both infringed citizens’ rights, and may thus have violated them. *Keeping the MPI in Dulwich* may have infringed citizens’ rights as well, but it would not have made anyone worse off than they would have been in the absence of governmental intervention.

4.7 The Defensibility of the Redirection Strategy

When the War Cabinet's Ministers rejected the redirection strategy, they argued that "it was not for the Government to authorise action which would result in certain areas in South London getting a heavier discharge of flying bombs and sustaining heavier casualties than they do at the present time" (PREM 3/111A, 3 August 1944). The Ministers felt "that if it came out that the Government had authorised such action their position would be indefensible" (ibid.). One way of reading what the Ministers were saying is as follows: "We don't think that redirection would be *wrong*. But it would be difficult to convince the general public of its rightness. Hence we refuse to go ahead with the redirection strategy." If this is what the Ministers were in fact saying,³⁴ were they mere cowards? Did they not want to do what they thought was best for their people simply because they were worried about damaging their re-election prospects?

I don't think so. For one thing, deception was an extremely secretive business. Only a handful of people actually knew about it. On top of that, during the war the British Press was heavily censored. For a long time to come, there was thus only a negligible chance that the general public would find out about their government's deception, and the Ministers must have been aware of this. For another thing, public decision-makers *rightly* feel accountable to the general public: for each measure that they consider implementing, they *should* ask themselves whether they could defend their choice to the general public. Whenever a measure is deemed indefensible, this provides a reason against adopting it.

According to my understanding of the term, to *defend* a public decision is to *successfully demonstrate that it was based on a careful weighing of pertinent reasons*, and to *successfully refute any charges that self-interest or other unduly partial motives played a pivotal role in the decision-making process*. If a decision was arrived at in the right way, it will usually be a defensible decision (and vice versa), but the two considerations can come apart. Think of it in this way: if a measure has overt

³⁴ The overall evidence does not suggest that this is what they were saying, but I think that this interpretation describes a stance worth exploring.

advantages for the public decision-makers who support it, they may have a tough time defending it, even if they arrived at it by weighing the right reasons against each other.

When the defensibility of an option and its being the right thing to do come apart, public decision-makers find themselves in a tricky situation. Isn't the point of their accountability—i.e. of their being required to explain and defend their decisions to the general public—to ensure that they do their job well? And don't they do their job well if they balance the right kinds of reasons to implement measures that are in the public interest? To some extent, this seems right. Should public decision-makers thus bravely defy what is usually a sensible check on their power in cases where this check risks defeating its own purpose?

Not necessarily. By adopting an indefensible measure, public decision-makers risk undercutting public morale and stirring public unrest. In times of crisis (such as during a war), this is to be avoided not primarily because it may be detrimental to a politician's career, but because it may impede the government's ability to govern effectively. Yet the public's trust in its government and its support for the government are at stake only if an indefensible decision is made public. So if it is possible to hide a difficult-to-defend but justified course of action from the public, should government officials stealthily defy what is usually a sensible check on their power?

Again, not necessarily. First, there is the obvious risk that the public will nevertheless get wind of the government's decision, in which case the secrecy surrounding the difficult-to-defend decision will make it look even more suspect. But even if this risk is negligible, there is still an important downside to secrecy. Openness and transparency are important values for public decision-making not merely because the power of the executive needs to be kept in check; they are important also because a democratic political process is one that the general public has access to and is able to help shape. Openness and transparency don't just help bring about what are independently good political decisions. Rather, the goodness of the political decisions that they help bring about derives at least in part from the fact that these decisions were arrived at in an open and transparent way.

But in the historical case, the Ministers had no choice but to keep deception issues

secret from the general public. If the public knew what was going on, there was too great a risk that the enemy would find out, too. Under these special circumstances, was it simply irrational to worry about defensibility? I don't think it was, but we will have to consider some of the details of the historical case to see why.

If we assume that the Ministers judged that the balance of reasons weighed in favour of redirection, why might they still have been worried about its defensibility? For one thing, South London used to be where the poorer working classes lived; most politicians and their families would have lived north of the river. For another, by directing the bombs even further away from Central London, the British authorities were keeping Whitehall—their place of work—bomb-free. The general public might thus have felt that the authorities saw the redirection strategy in an unduly favourable light primarily because it allowed them to keep themselves, their families, and the 'important' upper classes safe from harm.

Was this worry justified? It might seem so. In the actual historical case, as the terror bombing dragged on for week after week and month after month, people living in the most affected areas started to become increasingly unnerved. Letters to newspaper editors grew full of complaints such as “if the weight of the attack had been falling on [...] Whitehall and Buckingham Palace, far more vigorous attempts would have been made to stop it”, and “if Whitehall were in [Kent], these attacks would cease” (Campbell, 2012, p. 420). The people living in South London thus felt that their government wasn't dealing with the bomb threat in an effective way because the bombs *happened* to be falling in South London. Had the people living in South London known that the authorities had in fact *actively kept the bombs away from themselves*, their dissatisfaction with their government would doubtlessly have been even more pronounced.

At core, it seems to me that the South Londoners were complaining that they were not taken seriously. They were confident that more *could* have been done to counter the bomb threat, and that more *would* have been done if the weight of the attack had fallen on a different part of London. As the bombs kept falling, the South Londoners grew more and more convinced that their government wasn't taking decisive action to eliminate the threat simply because from the government's perspective, the danger

seemed sufficiently contained. But for the South Londoners, there could not have been a more pressing issue than the bomb threat! Every day, people were dying and suffering terrible injuries, and the South Londoners had no way of going about their daily business without fearing for their lives. This made their government's perceived inaction insufferably insulting.

We have no way of knowing whether the South Londoners were justified in growing discontented with their government. It is conceivable that there simply was no way of stopping the attacks, short of winning the war.³⁵ But for our purposes, it suffices that the complaints that they brought forward point to an important problem with the redirection strategy that we have not so far considered. The redirection strategy was a *protective measure*. Even compared to other protective measures, it was an extremely effective way of *getting the bomb threat under control*. Redirecting the bombs was expected to substantially reduce both the average risk of death and injury among the population, as well as the threat to Central London's infrastructure. To implement the redirection strategy was thus to take the edge off the bomb threat—for everyone except the people living south-east of Dulwich. In that way, there was a real danger that once the redirection strategy was in place, the bomb threat would no longer be regarded as an urgent issue in need of further attention. But this seems utterly reprehensible. If things *would* have turned out this way, the people living south-east of Dulwich would have had their interests sacrificed for everyone else's benefit in a

³⁵ On 28 February 1945, Winston Churchill complained that not enough seemed to be done to counter the bomb threat: "I do not at all feel convinced that [the Air Ministry] are exerting themselves as they should, or as they would be forced to do if the bombs were falling in the Whitehall area. I am not at all satisfied with the efforts the Air Ministry are making in this matter. All this weight falls on such limited areas, and some allowance must be made for the feelings of the people living in them" (PREM 3/111A). Churchill requested that the matter be taken up at the next meeting of the War Cabinet, and that it be brought before the Armed Forces Chiefs of Staff (*ibid.*, 28 February 1945). On 2 March 1945, Archibald Sinclair, the Royal Air Force's Secretary of State for Air, laid out in some detail why the Air Ministry were doing everything they reasonably could be expected to do (PREM 3/111A).

To the general public, it was obvious that not everything was being done to stop the bomb threat because the Allied Forces had not yet started to terror bomb German towns in response to the attack on London. Many Londoners explicitly demanded that there be "reprisals for this illegal form of warfare" (Campbell, 2012, p. 285), and retaliatory terror bombing was in fact entertained as an option by the British authorities immediately after the terror bombing of London started (see CAB 154/49, 21 June 1944). It is not clear to me, however, that reprisals would have stopped or even slowed down the attacks on London.

way that is irreconcilable with their free and equal standing with all other members of society.

It seems to me that these considerations point to a proviso with respect to the moral and political permissibility of the redirection of threats. In short, the idea behind the proviso is the following: if you redirect a threat, then you ought to subsequently expend the same effort to protect the smaller number as you ought to have expended to protect the larger number in the absence of redirection. Whether you are a government or an individual, what makes redirecting a threat justifiable is partly a commitment to do just as much to stop the threat as you would and should have done in the absence of redirection. To liken the point at issue to a trolley case, suppose a runaway trolley is hurtling towards five people, and that you redirect the trolley onto a side track where it threatens to kill only one person. Further suppose that whether or not you redirect the trolley, there are certain things you can do at relatively modest cost to yourself and to otherwise uninvolved bystanders in order to try and stop the trolley before it reaches its victim(s). It then seems to me that if you redirect the trolley, this does not reduce what you ought to do to stop the trolley *simply* because the lives at risk have been reduced from five to one. In other words, even if lesser evil considerations make it permissible to redirect the trolley, you still have to accept that other things equal, your responsibility to try and stop the trolley remains unchanged. What is the idea behind this proviso? I think it has to do with proper respect for persons. To redirect a trolley from five to one is to take sides with the five against the one. There are certain problems with this both from a rights-based perspective and from the point of view that everyone should be treated equally. But the initial act of redirection is at least compatible with the idea that the life of the one was permissibly put at risk *only because and only insofar as* this served to significantly reduce a *comparable risk* to others. However, if less is being done to stop the trolley after redirection, this makes it the case that the life of the one was put at risk partly to reduce what are relatively trivial costs to others. This is deeply disrespectful towards the one.

To recapitulate, the War Cabinet's Ministers were aware of the fact that their decision to redirect the bombs would not be revealed to the general public. If they

felt that the balance of reasons weighed in favour of redirection, should they thus have endorsed the redirection strategy despite their worry that adopting it would be difficult to defend? Possibly, but only after giving it careful thought. Worries about the defensibility of an option may express important concerns that have not so far been articulated clearly; they therefore deserve diligent investigation. To investigate their worries, the Ministers might have engaged in the following back-and-forth: “What makes redirection difficult to defend?” “Well, in essence, it keeps the bombs away from the rich at the expense of the poor.” “Ok. What if the situation were the other way around? What if redirection meant keeping the bombs away from the poor at the expense of the rich? Would we still judge that we should go ahead with it?” “I think so. It would still save thousands of lives, and there is just no real alternative to it.” “What would the rich say if we consulted their opinion?” “Good question. What would *I* say if the bombs were redirected towards me? I guess I would insist that the threat be dealt with swiftly and decisively. I don’t think I could reasonably be opposed to redirection if there was a credible commitment that way.” “I tend to agree. But then doesn’t this imply that we should be able to defend the *actual* redirection strategy if we commit to doing everything we reasonably can to put a swift end to the attacks?”

This hypothetical back-and-forth shows that the Ministers might have arrived at my ‘proviso on redirection’ simply by thinking through their defensibility worries. Thinking through these worries would have brought out something that is easily overlooked in the assessment of different deception strategies, namely that the general public probably would have and rightly should have held its government accountable for the government’s *overall response to the attacks*, not for individual aspects of this response considered in isolation. Once this is kept in mind, it seems plausible that the redirection strategy could have formed part of a justifiable *and* defensible overall response to the attacks. If it had been combined with a commitment to a swift and decisive elimination of the overall threat (that would actually have been honoured), I believe that the redirection strategy could have formed part of a morally optimal response to the bomb threat, and would as part of such a response have been morally justifiable. To be sure, redirection was a protective measure that expressed

unequal concern, and this fact morally spoke against it. But the price to pay for avoiding the expression of unequal concern would have been extremely high: up to 2,100 additional bomb-related civilian deaths each month, possibly many more bomb-related serious injuries each month, a non-quantifiable increase in the risk of damage to critical infrastructure, and a high likelihood that the excellent reputation of the British double agents with their German controllers would be destroyed. In light of this high price, attempting to confuse the enemy *merely to avoid expressing unequal concern* seems foolish, and it does not strike me as a course of action that the general public would have insisted its government take if there had been a public debate on this issue. Redirecting the bombs *while ensuring that the elimination of the overall threat remains a top priority* would have been a more responsible—though necessarily imperfect—way for the British government to honour the equal standing of all of its citizens.

It is true that adopting the redirection strategy would have infringed the rights of those citizens whom it would have exposed to a higher risk of death and injury. But if redirection had formed part of an overall response to the bomb threat that would have been morally justifiable, then adopting the redirection strategy would not have violated citizens' right to life; it would merely have permissibly infringed them.

4.8 Conclusion

The redirection strategy was a prudent and effective protective measure. Adopting it promised to save many civilian lives, to prevent thousands of serious injuries, and to reduce the risk of damage to critical infrastructure—all while securing the excellent reputation that the British double agents enjoyed with their German controllers. But for all of these advantages, to redirect the German bombs was to show an objectionably unequal concern for the safety of different parts of the population. To redirect the bombs was to single out a specific part of the population to bear the brunt of the German bomb attack. This made the redirection strategy morally problematic, as a government owes its duties of protection equally to all of its citizens. If the fact that it expressed unequal concern meant that the redirection strategy was not morally justifiable despite its advantages, then adopting it would have violated the right to life of the Londoners living south-east of Dulwich.

If the British authorities were worried about violating their citizens' rights, they could to some extent have circumvented this worry by rejecting the redirection strategy in favour of Findlater Stewart's second proposal to keep the MPI in Dulwich. Just like redirection, keeping the MPI in Dulwich would have shown unequal concern. Unlike redirection, however, keeping the MPI in Dulwich would have less clearly infringed people's rights, as it would have let harm come to innocent civilians in a way that is structurally similar to a mere failure to protect them.

Having said that, I have argued that with suitable accompanying measures, the redirection strategy would have been morally justifiable despite its shortcomings, so that the Ministers could have adopted it *without* having to worry about violating their citizens' rights. First, the only alternative to redirection that avoided the expression of unequal concern was the Ministers' proposed confusion strategy. But compared to redirection, confusion might have resulted in thousands of additional civilian deaths each month, and it would have jeopardized the reputation of the British double agents, which was strategically valuable. There would thus have been a steep price to pay for avoiding the expression of unequal concern. Second, the British government could have taken steps to ensure that eliminating the overall bomb threat would have

remained a top priority even after the redirection strategy was adopted. In that way, the British government could have gone a fair way towards honouring the equal standing of all of its citizens even while adopting the redirection strategy. Given this, it seems plausible to assume that redirection could have formed part of a morally optimal response to the German terror bombing, and would as part of such a response have been morally justifiable.

At any rate, it is important to keep in mind that the general public would have held its government accountable for the government's *overall* response to the bomb threat, not for its choice of a deception strategy considered in isolation. While the British authorities seem to have overlooked this fact when they tried to settle on a suitable deception strategy, a careful investigation of their lingering defensibility worries would probably have brought it out.

5 | A Profiting Innocent Threat and You: Why It's a Toss-Up*

"I don't translate my own convenience into other people's duties."

– Farebrother (in George Eliot's *Middlemarch*, 1874, p. 146)

5.1 Introduction

Imagine that, through no fault of your own, you find yourself at the bottom of a deep well.¹ A gust of wind has picked up an innocent person—call him Bob—and has blown him down the well. Bob is now falling towards you. If you do nothing, your body will cushion Bob's otherwise lethal fall. This will guarantee his survival, but it will kill you. If you shoot your ray gun, you vaporize and kill Bob, thereby saving your life. Are you morally permitted to shoot your ray gun?

If you are, then it is permissible to kill a *profiting, non-responsible threat*. Bob is a *threat* (as opposed to an attacker) because he endangers your life through no immediate action of his own. He is *non-responsible* because he neither intended to become a threat, nor was becoming a threat a consequence of an intentional action of Bob's that he could reasonably have been expected to foresee.² Bob *profits* from you

* A version of this chapter will appear as 'The Toss-Up Between a Profiting, Innocent Threat and His Victim' in *The Journal of Political Philosophy*. For valuable feedback and discussions I am grateful to Brian Berkey, Alexis Elder, Jeff McMahan, Michael Otsuka, Jonathan Parry, Jonathan Quong, and an editor as well as two anonymous referees of the *Journal of Political Philosophy*. Earlier versions of this chapter were presented at the Warwick Graduate Conference, the Yale/UConn Graduate Conference, the London-Berkeley Graduate Conference, the Brave New World Postgraduate Conference, and at a Choice Group Seminar at the London School of Economics and Political Science.

¹ This hypothetical situation is modeled on an example in Nozick, 1974, p. 34.

² This follows the use of the term 'moral responsibility' in McMahan, 2009, pp. 167-8.

because he depends on your presence for his survival.³ There is no agreement among moral philosophers whether self-defence against a profiting non-responsible threat is permissible. In this paper, I argue that the right thing to do when faced with such a threat is to flip a coin. By flipping a coin to determine who should live, you ensure that the indivisible good of continuing one's life is distributed as fairly as is possible without wasting it.

The structure of this paper is as follows. In sections 5.2 and 5.3, I present the main arguments that have been brought forward in favour of and against the moral permissibility of killing a non-responsible threat in self-defence. In section 5.4, I argue that all of these arguments rest on a misconception, and that flipping a coin emerges as the right thing to do once this misconception is corrected. In section 5.5, I compare and contrast a number of cases to test the plausibility of the arguments brought forward earlier. In section 5.6, I argue against the idea that a potential victim has a prerogative to save her own life.

5.2 Arguments in Favour of Permissibility

Judith Thomson (1991, esp. pp. 300-3) argues that killing a non-responsible threat in self-defence is permissible because a non-responsible threat loses his right not to be killed once he starts threatening the life of his victim. Thomson's idea is that Bob loses his rights against you when he starts falling towards you in a life-threatening manner *without your having threatened his life first*. Given that Bob started falling towards you in this unprompted way, Thomson believes that Bob would *violate* your right not to be killed if he killed you. To Thomson, you have a right to enforce your right not to be killed. This is why she concludes that Bob must have lost his right not to be killed when he started falling towards you.

³ In this paper, my primary concern is with *profiting* non-responsible threats. Unless stated otherwise, all references to non-responsible threats should thus be read as references to *profiting* non-responsible threats. In the literature, the distinction between profiting and non-profiting threats is not usually drawn. One exception is Victor Tadros, who mentions the distinction, but thinks it irrelevant for his arguments. See Tadros, 2011, p. 248. For the remaining authors I mention in this paper, I will simply assume that they mean their arguments to apply equally to both profiting and non-profiting non-responsible threats.

According to Thomson's analysis, moral agency is not a necessary condition either for losing one's rights or for violating someone else's. Bob's falling towards you is not something he is in control of or responsible for; it is something that merely—and through no fault of his own—happens to him. Thomson argues that it nevertheless grounds Bob's liability⁴ to defensive harm by making it the case that Bob is threatening to crush you, which for her means that he is threatening to violate your right not to be killed.

Kamm (1992a, pp. 45-55; 2012, p. 61)⁵ argues that Bob would have to suffer greater harm to avoid harming you than an innocent bystander would have to suffer to keep Bob from harming you. To give an example, Kamm (1996, p. 96) thinks that if Bob could redirect himself away from you and save your life at the cost of breaking his leg, then he would have to redirect himself, even though an innocent bystander would not have to break one of his legs to keep Bob from harming you. To Kamm (*ibid.*), if Bob is unable to impose on himself the harm that he would be duty-bound to suffer for your sake, others may impose it on him on his behalf. But why does Kamm think that Bob's moral standing vis-à-vis you is weaker than that of an innocent bystander? Kamm (1992a, p. 47) claims that Bob's falling towards you means that he is "is headed to where [he] should not be, on the victim, who has a right not to be occupied." To Kamm (*ibid.*), "[one] simply has a right not to have someone on the body or property to which one is entitled, even if the wind put them there." In Kamm's view, Bob is thus liable to defensive harm not because he threatens to violate your right not to be killed, but because he threatens to violate your right not to be occupied. If either the victim or the threat will be harmed very seriously, Kamm (1992a, p. 54; 2012, p. 61) claims that it is permissible to side with the victim, presumably because the threat, but not the victim, is in a morally inappropriate position.

Like Kamm, Victor Tadros (2011, pp. 248-56) believes that Bob does not threaten to violate your right not to be killed, but that he would nevertheless have to suffer greater harm for your sake than an innocent bystander would have to suffer. To

⁴ The language of liability was popularized by McMahan. Thomson does not use it, though she is in effect relying on the same notion.

⁵ See also Kamm, 1987, and Kamm, 1996, esp. p. 96.

Tadros, this is so because we have a special responsibility for what our bodies do, even when what our bodies do is not a product of our agency. Tadros furthermore argues that you may impose more harm on a non-responsible threat than he has an enforceable duty to suffer. To Tadros (2011, p. 148), the harm someone has an enforceable duty to suffer is the harm that we may intentionally inflict on them when manipulatively using them as a means to promote some further end. But if you vaporize Bob, you do not use him as a means; you merely eliminate him. Tadros (2011, p. 244) thinks that eliminative harm is easier to justify than manipulative harm, as only the latter involves seeing people as “material for use to promote the greater good.”⁶ Given his idea that Bob would have to lose a foot to keep himself from crushing you, Tadros (2011, p. 251) then argues that once we “accommodate [the] difference between manipulative and eliminative harming by [increasing] the cost that would have to be borne by the threatening party”, it becomes plausible that you may kill Bob eliminatively.

5.3 Arguments Against Permissibility

In contrast, Jeff McMahan and Michael Otsuka argue that a non-responsible threat is not liable to defensive harm because he hasn’t done anything that would justify a diminution in his moral status.⁷ Because they think it impermissible to kill someone who is not liable to defensive harm in order to save a single person, they conclude that it is impermissible to kill a non-responsible threat in self-defence.⁸

McMahan (1994; 2002; 2005b) argues that we become liable to lethal defensive harm whenever we responsibly create a “forced choice between lives” (McMahan, 2002, p. 403), or bring it about that either we or someone else must die. Succinctly put, McMahan conceives of “the morality of self-defence as a matter of justice in the distribution of harm” (McMahan, 2002, p. 402), and maintains that it is just

⁶ Tadros notes that his distinction between manipulative and eliminative agency coincides with Warren Quinn’s distinction between opportunistic and eliminative direct agency. See Quinn, 1989b, p. 344.

⁷ By the ‘moral status’ of a human person I mean the moral standing she enjoys vis-à-vis others. Someone’s moral status is higher the less we may do to her against her consent, and the more we must do for her at some cost to ourselves or others. See section 3.7.1.

⁸ Kai Draper (1993) was an early defender of a view similar to McMahan’s and Otsuka’s.

to defensively impose harm on someone only if that person is morally responsible for making its distribution unavoidable. It follows that according to McMahan, it is unjust—and hence impermissible—to kill a non-responsible threat in self-defence, as such a threat by definition lacks moral responsibility for his threatening status, and is thus not responsible for making the distribution of harm unavoidable.

Otsuka (1994) starts from the observation that it is impermissible to “ride roughshod”⁹ over an innocent bystander to save one’s own life. To illustrate, imagine that a piano is falling towards you, and that the only way to prevent it from crushing you is to vaporize it with your ray gun. The problem is that if you shoot the piano, you will also end up vaporizing Paul, an innocent stranger who is standing at one of the windows of the building from which the piano is falling. May you vaporize the piano and ride roughshod over Paul to save your life? Otsuka thinks it intuitively highly plausible that you may not. He argues that if this is right, it should also be impermissible to vaporize Bob, as there is no moral difference between Paul and Bob. According to Otsuka, this can be seen by constructing an intermediate case in which Paul no longer stands at the window, but sits—through no action that would make him responsible for falling towards you—on top of the falling piano. If you vaporize the piano, Paul dies; if you refrain from using your ray gun, Paul survives unscathed, and you die. Otsuka would maintain that shifting Paul’s location from ‘standing at the window’ to ‘sitting on top of the piano’ cannot make Paul liable to defensive harm. Put differently, Otsuka denies that it can make a moral difference whether an innocent person is located next to a threatening object or somehow ‘tied up’ with it. Moving from the intermediate case to the non-responsible threat case, Otsuka removes the piano from under Paul, so that it is now Paul’s own weight that threatens to crush you. Otsuka would again asserts that no moral difference is involved in this move. To him, the fact that it is now Paul’s own body that threatens to crush you cannot render it permissible to vaporize Paul when before it wasn’t.

⁹ This is Thomson’s terminology. See Thomson, 1991, p. 290.

5.4 The Toss-Up Between a Profiting

Innocent Threat and His Victim

I believe that the arguments presented in the last two sections are misguided. Here I explain why, and introduce what I take to be the right way of looking at things.

5.4.1 The Misconception

The positions presented so far all identify a purportedly *decisive moral asymmetry* between Bob and you. Given this, it seems likely that they are based on considerations along the following lines: “It is either permissible to kill a non-responsible threat in self-defence, or it is impermissible. Either way, there has to be a moral asymmetry between the threatened person and the threat, an asymmetry that favours either the victim, in which case the killing is permissible, or the threat, in which case the killing is impermissible. It is our task as moral philosophers to identify and explicate this asymmetry.”

Thomson, Tadros, and Kamm identify an asymmetry that favours you. Though labeled ‘non-responsible’ for threatening you, Bob is nevertheless characterized as morally compromised in some way: he threatens to violate your rights, finds himself in the wrong position vis-à-vis you, or is responsible for the harm his *body* threatens you with. Bob’s inappropriate relation to you then grounds his liability to defensive harm, and makes it permissible for you—*the victim*—to defend yourself.

McMahan and Otsuka identify an asymmetry that favours Bob. His falling towards you merely happens to him; it is not something he is in control of. If he ends up harming you, he cannot reasonably be held responsible for doing so. The same cannot be said of you. If you shoot your ray gun, you do so voluntarily and deliberately, as a fully responsible moral agent. Given his innocence, Bob cannot be liable to defensive harm, hence you—*the morally responsible agent*—mustn’t shoot your ray gun.

Neither way of looking at things is wholly implausible. However, in explicating various asymmetries, both miss the fundamental *symmetry* that in my view characterizes the situation. Notice how similar Bob and you really are. Bob is

innocent; he cannot be blamed for falling towards you. But you too are not to blame for the predicament you find yourself in! You did not know, and could not possibly have known, what you were getting yourself into when you stepped into the well. Yet as things are, you can save yourself only if you vaporize Bob, and Bob will survive only if he ends up crushing you. Both Bob and you faultlessly find yourselves in perilous situations from which you will escape only if an innocent other dies through you. Given this fundamental symmetry, it seems promising to frame the problem of the permissibility of killing a profiting non-responsible threat in self-defence in *distributive terms*, and to think of the threat and his victim as individuals with competing claims to the important good of continuing one's life (cf. Levine, 1984).

Insofar as Bob and you are similar, your claims to the good of a continued life tie. But it would be rash to conclude that your claims are all things considered equal, as there are important differences between the two of you also. In the next section, I examine these differences, and discuss whether they strengthen or weaken the non-responsible threat's position vis-à-vis his potential victim.

5.4.2 Accounting for the Differences Between a Profiting Innocent Threat and His Victim

A first important difference between Bob and you is that you act towards Bob as a morally responsible agent, whereas whatever Bob does to you he does to you merely qua falling body. McMahan and Otsuka believe this asymmetry weighs crucially in Bob's favour. Their idea is that Bob is not liable to defensive harm because he hasn't done anything that would justify a reduction in his moral status. Given that Bob's threatening you is not something he is responsible for or in control of, Bob cannot sensibly be said to be under a duty not to threaten you, and you cannot accordingly have a right against him that he not threaten you. Bob, on the other hand, has a right against you that you not kill him, and you have a duty not to kill him. From Bob's perspective, this way of looking at things appears plausible: surely what Bob *ought* to avoid he has to be *able* to avoid?

Thomson disagrees. Her idea is that we can threaten to violate a right, and thus become liable to defensive harm, merely by being turned into a threat. From

Bob's perspective, this way of looking at things is less compelling than McMahan's and Otsuka's. Why should Bob accept a reduction in his moral status because of something he is neither responsible for nor able to control? We can make better sense of Thomson's view once we give up the perspective of the threat, and consider the problem from the victim's point of view. From your perspective, your rights to non-interference enable you to lead a self-determined life in accordance with your own interests by giving you "primary say over what may be done to [you]" (Quinn, 1989b, p. 309). Rights to non-interference make you *sovereign* over your own body: they make it the case that other people are *trespassing* on your territory, and may therefore be fought off, whenever they cross the protective boundary that your rights to non-interference draw around you. As far as his impact on you is concerned, Bob's lack of responsibility is irrelevant. His body threatens to damage yours, and unless you are permitted to fight Bob off, this will have a negative effect on your ability to lead your life as you please (Thomson, 1990, pp. 222-3). If we think of your rights to non-interference as drawing a protective boundary around you that you have a right to enforce, Bob is *unlucky* to be crossing that boundary entirely involuntarily and non-responsibly, but that doesn't change the fact that he is crossing it.

I believe that Thomson's view contains an important insight. From your perspective, if your rights to non-interference become ineffective against Bob while his remain effective against you, it does feel as if Bob's bad luck of being turned into a threat is passed on to you, and this seems unjust. But Thomson's view has the contentious implication that 'ought' need not imply 'can', or that Bob continues to be under a duty not to violate your rights even when he has temporarily lost control over his actions. This strikes me—and many others—as unacceptable. Kamm's view faces a similar problem. Like Thomson, Kamm focuses on the victim's rights to non-interference, and insists that these rights do not become wholly ineffective against a non-responsible threat. If we are committed to the idea that 'ought' implies 'can', do we therefore have to agree with McMahan and Otsuka that Bob is not liable to defensive harm?

Tadros does not think so. He suggests we instead drop the assumption that a person becomes liable to defensive harm only if he threatens to *violate* another's rights. To Tadros, Bob is liable to defensive harm because we are all "responsible for what [our bodies do], even when that is not a product of [our] agency", simply because our bodies are—at least to some extent—what *we* are (Tadros, 2011, p. 255).

What should we make of this idea? For McMahan and Otsuka, it might simply beg the most important question. McMahan and Otsuka stress the fact that Bob is a *non-responsible* threat; Tadros in essence denies that a human person could be such a thing. McMahan and Otsuka might well agree that we are responsible for what our bodies do to the extent that we are able to control them, or are morally responsible for not being able to control them. But why should—or how could—our responsibility for our bodies extend beyond that? Suppose I own a piano which, for reasons beyond my control, crushes you. How am I morally responsible for the harm the piano causes *merely* because the piano is mine? Given this, I don't think we should agree with Tadros that Bob is responsible for what his body does. But then what are we left with? Is the fact that Bob poses a threat to you as morally irrelevant as McMahan and Otsuka claim it is?

Kamm and Tadros both base their arguments on the following observation: *if* Bob *could* avoid harming you, he would have to suffer substantial harm to do so. This seems to me importantly right. If Bob could redirect himself away from you to avoid crushing you, he would have to do this even if it would cost him a foot. But why would Bob have to sacrifice his foot? In essence, because our morality is built around stringent rights to non-interference, or on a strong presumption against harming innocent others. As long as we are able to avoid harming innocent others, we generally have to go to great lengths to do so. The mere fact that we lack responsibility for finding ourselves in a situation where we pose a threat to innocent others does not make it permissible to harm them. This is why Bob would have to suffer substantial harm to avoid harming you if he could, or why Bob would be violating your rights to non-interference if he crushed you when he could have redirected himself away from you at reasonable cost to himself.

Moral rules that are built on a strong presumption against harming innocent others neither allow us to minimize the harm that comes to innocent persons, nor do they allow us to distribute harm equally among similarly undeserving recipients. But they shield us from harm that is imposed on us by other people as long as we do not inflict harm on others. In this way, they provide us with the peace of mind that other people must not, in general, threaten our physical integrity. Knowing that our bodies are largely off-limits for those around us allows us to live together in peace while pursuing our own interests.

Yet even if Bob *would* have to make great efforts to avoid harming you if he *could*, this seems to be no more than a hypothetical consideration. After all, we have assumed that Bob cannot actually avoid harming you. You, on the other hand, can avoid harming Bob. Given this, and given our morality's strong presumption against harming innocent others, doesn't it follow that given the way things actually are, you ought to refrain from shooting Bob?

To Tadros and Kamm, the reason why Bob would have to avoid harming you if he could applies also when he cannot in fact avoid harming you. If Bob were responsible for whatever his body did, or if you had a right against Bob not to have Bob's body on yours, this would hold true irrespective of Bob's ability to control his fall. But I have argued that Tadros and Kamm's ideas are not compelling. Contrary to what Tadros and Kamm suggest, moral rights to non-interference are best thought of as held symmetrically among moral agents. Qua moral agent, you have rights to non-interference against other moral agents, and they have the same rights against you. Only if one moral agent violates the rights of another is this symmetry disturbed so that balance can be restored. The protective logic of our moral rights to non-interference is thus built around—and for its functioning depends on—the idea of mutual recognition between moral equals.

But note that if this is right, Bob retains his rights to non-interference when he starts falling towards you not only because he lacks responsibility for falling towards you. Lack of responsibility for threatening another is a necessary condition for keeping one's rights to non-interference, but it is not sufficient. Bob also retains his rights because he retains his moral capacities, or because he remains the kind of creature

who would have to avoid harming you if he could. In other words, if you must not vaporize Bob's out-of-control body, this is partly because Bob is the kind of creature who can understand that if he could avoid crushing you, he would have a duty to do so. But if you must constrain yourself towards Bob and let his body harm you *partly because Bob would (right now!) have to suffer substantial harm to avoid crushing you if he could*, it would be unjust if you could not impose that harm on Bob on his behalf. By analogy, suppose that as a parent, you ask your son to share his Halloween candy with his sister. You tell him that sharing is appropriate partly because his sister would also have to share her candy if she could. If so, it would be unjust and illogical if you made your son share his candy, but forbade him to help himself to his share of his sister's candies that happened to be out of his sister's reach but well within his.

But what about the fact that whatever harm you impose on Bob, you impose on him as a morally responsible agent? Don't you violate Bob's rights if you harm him? As long as you inflict no more harm on Bob than he is duty-bound to suffer, you do not violate his rights to non-interference. Whatever harm Bob is duty-bound to suffer, he would—as a morally responsible agent—have to inflict on himself if he were able to. There is thus a certain amount of harm that may rightfully be assigned to Bob, given the situation he is in. Just as you may shove away a bystander who is unwilling or unable to move out of the way when you are running for your life, you can impose on Bob the more substantial harm he is duty-bound to suffer without violating his rights (cf. Øverland, 2011, p. 550). Having said that, burdening Bob with more harm than he is duty-bound to suffer does not seem justifiable, and would therefore violate his rights to non-interference.¹⁰

¹⁰ Tadros claims that you are permitted to kill Bob because you would not be using Bob as a means. I am not convinced by Tadros' argument in favour of this claim. His argument relies on the assumption that there are significant differences between two types of intentional agency, so that eliminatively killing someone is no worse than manipulatively cutting off their foot. At the same time, it is a consequence of his argument that liability to defensive harm does not vary with the type of responsibility that grounds it. According to Tadros, vicious threats, responsible threats, and non-responsible threats are all similarly liable to defensive harm. But if 'mode of agency' is a significant consideration in one context, why is it not significant when it comes to determining a person's liability to defensive harm?

So far, I have argued that there is a justice-based reason to make a non-responsible threat share in the costs that his out-of-control body threatens to impose on his victim. To the extent that a non-responsible threat would have to suffer harm to avoid harming his victim if he could do so, we may redistribute harm from the victim to the threat. This does not imply that Bob is liable to defensive harm. It merely means that his rights to non-interference are permissibly infringeable to the extent that there is a justice-based reason to make him bear part of the cost that he would otherwise impose on his victim.

If what I have argued so far is correct, where does it leave us with respect to the first asymmetry between Bob and you, i.e. the fact that whatever harmful effects you have on Bob are due to your responsible agency, whereas whatever harmful effects Bob has on you are not attributable to his responsible agency? Compared to a responsible threat, the asymmetry weighs in Bob's favour, but less decisively than McMahan and Otsuka have argued. Though you would be permitted to kill a responsible threat in self-defence, you are not similarly permitted to vaporize Bob to save yourself. Instead, you have to take Bob's lack of responsibility into account and let Bob's right to non-interference constrain you. But Bob's moral status is not thereby equal to that of an innocent bystander. If Bob could avoid crushing you, he would have to suffer substantial harm to do so. This gives you a justice-based reason to make Bob share in the harm that he is threatening you with.

But Bob differs from an innocent bystander not only because he is threatening you. Recall the innocent bystander Paul, who is standing at one of the windows of a building from which a piano is falling towards you. Imagine what would happen to Paul if we removed you from the scene. Without you present, Paul could stand at the window and watch the piano fall without any risk of harm. Bob's situation is different. If you weren't there to cushion his fall, Bob would die on impact. Paul is thus merely a bystander to a dangerous situation not only because he does not threaten you, but also because his life is not at risk unless you involve him in your problems. Bob, on the other hand, depends on you for his life. Now by assumption, Bob's depending on you for his life is not something he is morally responsible for. But this does not suffice to make it morally irrelevant. After all, Bob's threatening you is also not something

he is morally responsible for, yet it still serves to morally distinguish him from an innocent bystander. But why—and how—should Bob’s profiting from your presence affect his moral standing vis-à-vis you?

Thomson (1971, pp. 48-55) has famously claimed that if a world-class violinist’s life were to depend on his being hooked up to your circulatory system for the next nine months, you would be permitted to unhook yourself, thus killing him. To Thomson, you have no duty to sacrifice nine months of your life to avoid killing the violinist. Instead, she argues that his non-responsible dependence on you weakens his moral standing vis-à-vis you, so that you may do more to him than you would be permitted to do to an innocent bystander.¹¹ Kamm (1992a, pp. 20-45) argues that the violinist’s dependence on you weakens his moral standing vis-à-vis you because he cannot be worse off through you than he would have been in your absence. If you unplug the violinist, he dies; but in your absence, he would have died anyway. Because of this, unplugging the violinist resembles letting him die. While you may have to shoulder significant costs to avoid letting an innocent other die, these costs are not usually thought to be as substantial as the costs that you have to shoulder to avoid killing an innocent other (Kamm, 1992a, p. 71). Applied to the situation between you and Bob, if you vaporize Bob, you kill someone who would have died in your absence. By contrast, if you vaporize Paul, you kill someone who would have lived in your absence. This distinguishes Bob from an innocent bystander such as Paul.

The fact that a profiting threat depends on his victim not only distinguishes him from an innocent bystander; it also distinguishes him from his victim. Bob is dangerous—a threat—to you, whereas you usefully cushion Bob’s fall at best, or do not substantially alter his predicament at worst (if you vaporize him, he dies, but without you there, he would have died anyway). From a self-interested perspective, Bob thus welcomes your presence, whereas you cannot be said to welcome his. Following Thomson and Kamm, I think that this difference strengthens your moral standing vis-à-vis Bob. It does not give you license to treat Bob in any way

¹¹ This is not to say that responsibility is irrelevant to the assessment of dependence relations. If Bob were responsible for becoming dependent on you, his moral standing vis-à-vis you might well be more unfavourable than it is under the current circumstances, where he depends on you entirely non-responsibly.

you like, but it does seem to give you an advantage over Bob. Given that you are an innocent other who improves Bob's chances of survival, whereas Bob is an innocent other who worsens your chances of survival, Bob should have to make an extra effort to avoid harming you if he could.

5.4.3 Weighing Claims

So far, I have argued that there is a justice-based reason to impose on Bob the substantial harm that he would be duty-bound to suffer for your sake. Does this shed any light on the relative strengths of Bob's and your claim to the good of a continued life?

Suppose first that Bob was a non-profiting threat. I have argued that given our morality's strong presumption against harming innocent others, Bob would have to suffer *substantial* harm to avoid harming you if he could. But would he be required to subjugate his own interests entirely to yours? I don't think so. But why not? Given a strong presumption against harming innocent others, why shouldn't Bob have to do whatever it takes to avoid killing you? Recall the innocent bystander Paul, whose life you are required to spare at the cost of your own. It may be very demanding to let yourself be crushed by a falling piano when you could in principle vaporize it. Yet given our morality's strong presumption against harming innocent others, you are at most *excused* if you vaporize the piano and kill Paul. But if you have to avoid killing Paul at the cost of your life, why wouldn't Bob have to avoid killing you at the cost of his life?

The important difference is this. You avoid harming Paul as long as you do not harm anyone, whereas Bob avoids harming you only if he inflicts harm on himself. Because Bob has to inflict harm on himself to avoid harming you, Bob can at most be required to share equally in the harm that his body threatens to impose on you. If he were required to pick up the lion's share, this would deny him the right, qua moral agent, to regard harm that he imposed on himself as on a par with harm that he let come to you. Given Bob's innocence, this would be an alienating requirement. Even if our morality is built on a strong presumption against harming innocent others, when avoiding harm to an innocent other means inflicting that harm on our innocent

selves, we must not be required to treat ourselves as below par. Such a requirement would be inconsistent with a proper sense of self-worth.

Should Bob be permitted to give *more* weight to his own interests? If Bob harms you, he merely *allows* himself to harm you, and he merely *foresees* that harm will come to you. Allowing harm is often thought less bad than actively inflicting it, and merely foreseeing harm is often thought less bad than intending it. Does this mean that Bob should not have to intentionally inflict as much harm on himself than he lets come to you? Suppose you are the front passenger in a car whose driver starts fainting just as the car comes up behind a pedestrian. If you do nothing, the car will hit the pedestrian. Is the harm that you are required to suffer to avoid hurting the pedestrian reduced by the fact that whatever harm that you let come to him is neither intended nor actively inflicted? If you had already taken the wheel and started steering when you noticed the pedestrian, would this increase the harm that you would be required to suffer to avoid hurting the pedestrian? My hunch is: *not by much*. Why could that be so? It seems to me that our duties to avoid harming others are requirements to regard their bodies as territory that we need to steer clear of in the pursuit of our own interests. The main failure of respect involved in harming an innocent person is *not regarding that person's body as off-limits for one's purposes*. This holds true for all modes of harming, and it makes them all similarly bad. In other words, even if Bob's mode of harming you is less disrespectful than other modes, Bob's attitude towards you is nevertheless very inappropriate if he fails to redirect himself away from you at moderate cost to himself.

To what extent might Bob's relatively respectful mode of harming you work in his favour? Out of two comparable options, Bob should probably be allowed to pick that option which slightly favoured his interests over yours. If one option gave both of you an equal chance of survival, and a second option improved Bob's chances (while reducing yours) by a couple of percentage points, Bob should probably be permitted to go with the second option. But note that we have not yet factored in the consideration that Bob is a *profiting* non-responsible threat. Taken by itself, the fact that Bob benefits from your presence cannot be of overriding importance. Whether other people's mere presence happens to be beneficial or burdensome to us

is mostly a matter of our good or bad luck. Having said that, whether a person's presence is beneficial or burdensome is not entirely morally insignificant, as it can change the extent to which harming that person is disrespectful. In this way, it seems to me that a profiting threat would act unduly disrespectfully if he did not treat his victim's interests as on a par with his own. Suppose that as a profiting threat, Bob could again choose between two options, the first of which gave both of you an equal chance of survival, and the second of which was slightly skewed in his favour. Further suppose that Bob argued as follows: "I must not be required to pick the first option. That option is maximally demanding. It is the upper limit of what can be required of me without undermining my self-worth. Given that I do not trespass on my victim's territory in an especially disrespectful way, I should be cut some slack." It is not a profiting threat's place to argue in this manner. For even if Bob has to pick the maximally demanding option, he still has a 50% chance of survival, whereas if it weren't for you, he would simply die. Also, even if Bob has to pick the maximally demanding option, there is still a 50% chance that you will die, whereas in Bob's absence, your life would not be at risk. Given that circumstances have put Bob in such a favourable situation vis-à-vis you, Bob's attitude towards you should be one of humility. At any rate, he should not appeal to otherwise legitimate but relatively weak considerations to discount your interests relative to his own. It follows that a profiting non-responsible threat is morally required to treat his victim's interests as on a par with his own. From a distributive perspective, this allows us to conclude that Bob and you have similarly strong claims to the good of a continued life.

5.4.4 The Distributive Perspective

When different people have similarly strong claims to a good, it is fair to assign equal amounts of it to each of them (Broome, 1984; Broome, 1991a). In non-responsible threat cases, however, the good in question is *indivisible*. You cannot give half of a continued life to Bob and keep the other half to yourself. The only way in which you can assign equal shares is by vaporizing yourself, thus letting Bob fall to his death. That way, both of you end up with a share of zero. But this maximally fair solution

is unacceptably wasteful.¹² If it is possible to strike a balance between fairness and efficiency, ‘continuing one’s life’ is too valuable a resource to sacrifice for maximal fairness.

Broome suggests that when we are distributing an indivisible good between candidates with equally strong claims, we should hold a lottery in which all candidates are given an equal chance of winning the good. This avoids wastefulness, and the requirement of fairness is met to a considerable degree (Broome, 1991a, pp. 97-8; original emphasis):

“[If the good in question is indivisible and we don’t want to let it go to waste], the candidates’ claims cannot all be equally satisfied, because some candidates will get the good and others will not. So some unfairness is inevitable. But a sort of partial equality in satisfaction can be achieved. Each person can be given a sort of surrogate satisfaction. By holding a lottery, each can be given an equal *chance* of getting the good.”

Following Broome, we may thus conclude that as a potential victim of a non-responsible threat, you should flip a coin. You—a morally responsible agent—are charged with the distribution of an indivisible good. By ‘charged’ I mean: It just so happens that an important good needs distribution, and that you will end up distributing it, in that your action or inaction will inevitably determine how the good is distributed. Given this, you should aim at a distributively just solution, and should thus give an equal chance of getting the good to Bob and yourself. If you do this, your action properly reflects the equal moral standing Bob and you enjoy as individuals (cf. Broome, 1984, p. 38; Levine, 1984, p. 70; Davis, 1984, pp. 919-2; Lazar, 2009, p. 715) as well as the great value of the good in question (because you refuse to let it go to waste). If you shoot Bob without flipping a coin first, you rob Bob of his rightful chance to continuing his life. This chance is valuable: it represents the ‘surrogate satisfaction’

¹² Note that it is not Pareto efficient.

of Bob's claim to the good of a continued life, given the scarcity and the indivisibility of the good in question. It follows that you must not shoot Bob without holding a lottery first.¹³

5.5 Comparing Innocent Threat Cases

So far, I have argued for the following claims:

1. Compared to an innocent bystander, Bob non-responsibly depends on you for his life. This weakens his moral standing vis-à-vis you.
2. Compared to an innocent bystander, Bob non-responsibly threatens your life. This weakens his moral standing vis-à-vis you further.

Here, I want to test whether our intuitions are in line with these claims. To do this, I will alter the Original Case presented in the introduction by removing the factors that I have argued work to Bob's disadvantage. If (1) and (2) are correct, we should judge that Bob's moral position is strengthened as we move from one case to the next.

Case A

(α) *Bob does not profit from you;*

(β) *Bob is a threat to you*

The setup here is the same as in the Original Case, except for the bottom of the well now being laid out with protective padding. If you do nothing and let Bob crush you, you still die, but it is now the protective padding that cushions Bob's fall.

Given the setup of Case A, Bob is still intimately tied up with what poses a threat to you, but now your presence is no longer necessary for his survival. Rather, your

¹³ May you let yourself be crushed without flipping a coin first? It seems to me that while such a heroic act of self-sacrifice is not morally required, there is nothing wrong with your freely choosing it. Thomson (2008) would disagree. She argues that "a willingness to give up one's life *simply* on learning that [someone else] will live if and only if one dies is a sign of a serious moral defect in a person" (ibid., p. 366, original emphasis), as it is indicative of the fact that one values one's own life insufficiently. While I agree that proper respect for persons has to include proper self-regard, it does not seem to me that such self-regard is incompatible with a willingness to sacrifice oneself for an unspecified stranger. In the case at hand and given your innocence, I would think proper self-regard incompatible with a *denial* that you have a *rightful claim* to continuing your life, but not with the fact that you may *rightfully choose* to waive such a claim.

absence would now guarantee his safety. If given a choice, Bob would thus want you gone. Without you there, Bob would not have to deal with the prospect of crushing (or else being vaporized by) an innocent stranger, and the protective padding would safely cushion his fall.

Intuitively, I feel that Bob's moral position vis-à-vis you is strengthened as we move from the Original Case to Case A. This means that my intuitions are in line with the idea that Bob's depending on you works to his moral disadvantage.

In addition to the arguments already presented, there is one further theoretical consideration that supports this intuitive judgement. Suppose we follow Warren Quinn (1989b) and (more recently) Jonathan Quong (2009) in spelling out the Kantian idea of 'using someone as a mere means' in a counterfactual way, so that *X uses Y as a mere means whenever X harms Y without Y's consent, thereby achieving something X would not have been able to achieve in Y's absence*, Bob uses you as a mere means in the Original Case, but not in A. Under such a counterfactual definition of the term, my intuitive judgements thus suggest that in the Original Case, Bob's moral status is weakened by the fact that he uses you as mere means.¹⁴

In Case A, would it be wrong of you to flip a coin to determine who should live? As discussed, flipping a coin seems to me very much an upper limit of what could be required of Bob if he were able to control his fall; accordingly, a 50% chance of death seems to me very much an upper limit of what it would be permissible to impose on Bob in Case A. If you saw an option to give Bob a slight advantage over yourself, I think that you ought to go with that option instead of flipping a coin.

¹⁴ Quinn (1989b, p. 344) refers to using someone in this way as "direct opportunistic agency" and claims that it is especially hard to justify. Quong (2009, p. 525) agrees that there is something "particularly wrong" with harming someone in this way. When it comes to profiting non-responsible threats, there is the complication that the threat lacks control over his action, and can thus not be accused of "direct opportunistic agency". But if a threat *harms his victim without the victim's consent, thereby securing his survival (that he would not otherwise have been able to secure)*, this seems to me sufficient to conclude that he—entirely non-responsibly—uses his victim as a mere means in a counterfactual sense. As I have argued in section 5.4, our moral standing vis-à-vis each other does not depend exclusively on what we do to each other as morally responsible agents; the effects we have on others merely in virtue of our existence as embodied creatures can be morally significant also.

Case B

(α) *Bob does not profit from you;*

(β) *Bob is not a threat.*

This case is like Case A, except that the wind blows a spear down the well before Bob falls in. If you do not shoot your ray gun, your heart will be pierced by the spear, and Bob will then land on your dead body. You can either vaporize Bob and the spear at one go, or you can let your heart be pierced by the spear. The spear poses no threat to Bob. Moreover, if there were no spear, the impact of Bob's body on yours would cause you no harm.

In Case B, Bob and what's threatening you are entirely separable. That is, we can remove Bob from the scene without removing the threat, and we can remove the threat without removing Bob (so that Bob's presence is no longer sufficient for your life to be at risk). In Case B, if anything can be said to differentiate Bob from an innocent bystander, it seems to be mere spatial location—Bob is falling towards you, whereas innocent bystanders are usually assumed to stand some distance apart.

I feel that the Bob of Case B is morally indistinguishable from an innocent bystander whom you mustn't ride roughshod over. Intuitively, such a bystander has to incur moderate costs to move out of the way so that you can save yourself, but his duties are much more limited than those of a non-responsible threat. In B, it would therefore be impermissible to flip a coin to decide who of you should live.

5.6 A Prerogative to Save One's Own Life?

I shall now consider an important objection to the idea that when faced with a profiting non-responsible threat, one ought to flip a coin. The objection can be put as follows: "From an impartial point of view, the claims of a non-responsible threat and his potential victim may indeed tie. But it does not follow from this that the potential victim has to treat both claims equally, as he is not required to take up an impartial perspective. Given that his own life is at stake, the victim may give more weight to his own interests, and may permissibly treat his own claim as the more significant one." Nancy Davis (1984) and Jonathan Quong (2009) have both presented such a

prerogative-based argument in favour of the permissibility of killing non-responsible threats in self-defence.¹⁵ In the following, I focus on Quong’s argument, and explain why I think it fails.

Quong (2009, p. 518) comes up with and defends a general principle called the *Principle of Defensive Killing* (PDK). The PDK states that “[you] can permissibly kill [another person] X if X will otherwise kill you”, provided that “(i) killing X is the only reasonable course of action that can save your life, and (ii) you have not waived or forfeited your permission to act in self-defense.” (ibid.). According to the PDK, you are allowed to vaporize Bob because unless you do so, Bob will kill you, and because your vaporizing him is in compliance with (i) a proportionality and (ii) a ‘no forfeiture’ clause.

Like Thomson, Kamm, and Tadros, Quong argues unambiguously in favour of the permissibility of killing a non-responsible threat. His approach nevertheless differs from theirs in at least two important respects. Firstly, while Thomson, Kamm, and Tadros claim that you may kill Bob because he threatens to *violate your rights* or because he is *liable to defensive harm*, Quong thinks you may kill Bob *simply* because he will kill you unless you kill him. Secondly, Quong (2009, p. 519) argues (contra Thomson and Tadros) that responsible agency is necessary for violating another’s rights as well as for becoming liable to defensive harm. To Quong, when Bob starts falling towards you, he neither threatens to violate your rights, nor does he become liable to defensive harm. Unlike Thomson and Kamm, Quong is therefore not committed to the counterintuitive claim that ‘ought’ need not imply ‘can’. Moreover, unlike Thomson, Kamm, and Tadros, he is not committed to the revisionist claim that responsible agency is not a necessary condition for liability.

Instead, Quong appeals to the idea of a prerogative to justify the PDK. When an agent’s own life is at stake, Quong (2009, pp. 516-7) thinks that the “agent-relative value each person’s life has for them” grounds a “powerful agent-relative permission to avoid sacrificing or significantly risking their own life for the sake of others [...]” While Quong does not argue that such a prerogative implies a general permission to kill an innocent person whenever doing so would serve to save one’s own life, Quong

¹⁵ Levine (1984, esp. p. 74) presents a similar argument, but rejects it.

does believe that an agent may permissibly kill an innocent person to save his own life just in case he does not use that person as a mere means. Whether an agent uses another person as a mere means is decided through consideration of a counterfactual: if the agent would not be able to secure his own survival if the other person were not present, the agent uses that person as a mere means. Quong (2009, p. 525) appeals to a Kantian idea of respect to argue that there is something “particularly wrong” with harmfully using someone as a mere means, so that our otherwise powerful prerogative to save our own life cannot extend to killings that involve such use. Quong then claims that according to his definition of using as a mere means, no one is used as a mere means in the cases that the PDK captures, simply because where X is what threatens to kill you, removing X does not make it impossible for you to secure your survival—rather, it guarantees it.

At first sight, Quong does not seem to miss the fundamental symmetry that I have argued characterizes non-responsible threat cases. Quong would agree that from an impartial point of view, Bob’s and your claim to the good of a continued life tie. In fact, Quong insists that neither the non-responsible threat nor his potential victim lose their respective rights not to be killed, and that both threat and victim retain their right to self-defence, in that according to the PDK, they may permissibly infringe the other’s right not to be killed (Quong, 2009, p. 525; see also Davis, 1984, p. 193). In sum, Quong seems to offer a clear and elegant solution to the problem of self-defence against non-responsible threats that is consistent with the view that the situation between a profiting non-responsible threat and his victim is roughly symmetric from an impartial point of view.

On closer inspection, however, our Original Case points to a weakness in Quong’s account. Succinctly put, the problem is that Quong’s PDK does not match up with the rationale he provides in its favour. According to the PDK’s *wording*, the situation between Bob and you is indeed symmetric as I have sketched it in the above paragraph. But according to its *rationale*, the situation is asymmetrical: given that Bob would be using you as a mere means if he were to kill you, he would—according to the PDK’s rationale—have no right to kill you. If he could smash your ray gun, he would thus be disallowed from doing so, even if he knew that you will shoot him if he doesn’t.

The letter of the PDK thus allows certain killings that its justification forbids.¹⁶

To make matters worse, there are killings that seem intuitively impermissible, but that according to the PDK's justification we ought to endorse. Consider again Paul, the innocent bystander. If you vaporize the piano, you save your life without using Paul as a mere means, as Paul's presence makes no difference to your ability to vaporize the piano (Otsuka, 1994, p. 77; Hanna, 2012). In fact, you would welcome Paul's absence, as you could then vaporize the piano without also killing an innocent person. It follows that according to the PDK's rationale, it is permissible to vaporize Paul. I agree with most other moral philosophers that this is an unacceptable implication.¹⁷

Why does the PDK run into these problems? The underlying issue seems to me the following. Quong thinks that the agent-relative value each person's life has for them makes it sometimes permissible to kill an innocent person in order to save one's own life. He tries to substantiate what he means by 'sometimes', both by formulating a rule-like principle (the PDK), and by providing a rationale in its favour. While it may be possible to come up with a rule-like principle that matches our intuitions about specific cases, the search for an adequate rationale with genuine explanatory power may well be a futile one. Killing an innocent person *simply because* one puts one's own interests first sounds like a gravely disrespectful thing to do, and I doubt that there is a convincing rationale why doing so should sometimes be thought morally justifiable (as opposed to merely excusable). For any proposed rationale, I thus conjecture that it will be possible to construct counterexamples to it, i.e. cases in which the rationale permits killings that we find intuitively wrong.

I therefore propose a different approach. I suggest we take as a starting point cases where substantial harm needs distribution among innocent individuals and where one may clearly put one's own interests first, and that we then investigate whether a plausible constraint on agent-relative prerogatives emerges. What are cases where substantial harm needs distribution among equally undeserving individuals, and where

¹⁶ This assumes that intentional agency is not a necessary condition for 'using as a mere means'.

Quong might want to disagree with this. See footnote 14.

¹⁷ Quong is aware that the PDK's rationale sanctions killing Paul, and is willing to accept this implication of the PDK (that is, he thinks that we ought to reconsider our intuitions about the permissibility of killing innocent bystanders such as Paul). See Quong, 2009, pp. 529-30.

we do not doubt that the special value each person's life has for them grounds a prerogative to put one's own interests first? Suppose that you are out hiking in the woods with an innocent stranger when you chance upon a hungry bear (Boorse and Sorensen, 1988, p. 115). The bear starts charging, and it is evident that he will eat either you or the innocent stranger depending on who manages to run faster. Do the two of you have to flip a coin in order to determine who may run and who has to stay put? Of course not. You may simply run as fast as you can. It is not morally required that either of you be given a fair chance to survive; even if you are confident that you are able to run much faster than the stranger, you may still run for your life, thus leaving the other behind (Boorse and Sorensen, 1988, pp. 115-6). What it intuitively seems you must not do is stab the stranger with your hunting knife in order to slow him down.

To return to our Original Case, suppose the bottom of the well were large enough for you to step aside, so that you could save your own life without killing Bob, but by merely letting him die. In such circumstances, it seems to me again obvious that you would be permitted to secure your own survival, despite the fact that Bob is innocent, and despite the fact that stepping aside would mean his certain death. What makes the Original Case tricky is the fact that you cannot save your own life without *killing* Bob.¹⁸ Can these intuitions be said to capture a moral principle? At the very least, they are consistent with the following constraint on agent-relative prerogatives (CAP):

CAP. It is impermissible to depart from an impartial point of view and give more weight to your own interests if you can do so only by inflicting harm on another person that this person would not be duty-bound to suffer for your sake.

To violate the CAP is to make an exception of yourself in a morally objectionable way: it is to pursue your own interests in precisely the way you are morally asked not to pursue them. If you inflict harm on another person that this person is not required to suffer for you, you trespass on another person's sovereign territory that morality clearly flags as 'off-limits' for you. But why is each of us morally entitled

¹⁸ Elsewhere, Quong agrees that the distinction between harming an innocent other and allowing him to be harmed is morally significant. See Quong, 2012, p. 63, fn. 29.

to territory of their own? Why is our morality based on such a strong presumption against harming innocent others? Note first that we are social beings. As such, we lead our lives surrounded by other people. Second, we are rational creatures. As such, we have our own plans and projects, and we aspire to lead our lives in accordance with them. Third, we are embodied creatures: if someone harms our bodies, they harm us. Stringent rights to non-interference enable us to live together peacefully while still pursuing our own interests. Only if there is a general—as well as generally accepted—rule that we must avoid harming each other can we go about our lives without constantly seeing each other as potential threats.

If the CAP did not exist, or if people constantly ignored it, this would undermine the function of our stringent rights to non-interference. To violate the CAP is thus to pursue your own good in a way that is not universalizable. The situation presents itself differently if you are able to pursue your own good without trespassing on someone else's sovereign territory. In that case, a permission to put your own interests first need not undermine our general ability to pursue our own interests while living together peacefully. It follows that if you can pursue your own interests without trespassing on anyone else's sovereign territory, then you are in luck: an agent-relative prerogative may well apply to you.

Suppose this is roughly right, and something like the CAP does in fact apply. There are then no circumstances under which you have a prerogative to kill an innocent other in order to save your own life. Also, the only reason why you don't have to let Bob crush you, but may instead flip a coin to determine who of you should live, is because flipping a coin is the impartially right thing to do. When it comes to Bob and you, we can take up whatever perspective we want—the situation between the two of you is always a toss-up.¹⁹

¹⁹ Given our morality's strong presumption against harming innocent others, wouldn't it be better for an otherwise uninvolved bystander not to intervene? I don't think so. A strong presumption against harming innocent others means that innocent people will not usually be duty-bound to bear much harm. But when in exceptional circumstances they *are* bound to bear a certain amount of harm, we should intervene to ensure that the right people are harmed to the right extent. Also, even if flipping a coin is the impartially right thing to do, an uninvolved bystander may still have an agent-relative prerogative not to intervene. That is, our morality may not be as demanding as to require that a bystander run the risk of having to kill an innocent person in order to give another innocent person a fair chance to survive.

Bibliography

- Boorse, C. and Sorensen, R. A. (1988). Ducking Harm. *The Journal of Philosophy*, 85(3): 115–134.
- Bradley, B. (2004). When Is Death Bad for the One Who Dies? *Nôus*, 38(1): 1–28.
- Bradley, B. (2007). How Bad is Death? *Canadian Journal of Philosophy*, 37: 111–127.
- Bradley, B. (2009). *Well-Being and Death*. Oxford University Press, Oxford.
- Bradley, B. (2013). Instrumental Value. In LaFollette, H., editor, *International Encyclopedia of Ethics*. Wiley-Blackwell, first edition.
- Broome, J. (1984). Selecting People Randomly. *Ethics*, 95(1): 38–55.
- Broome, J. (1990-1991a). Fairness. *Proceedings of the Aristotelian Society, New Series*, 91: 87–101.
- Broome, J. (1991b). *Weighing Goods: Equality, Uncertainty, and Time*. Basil Blackwell Ltd., Oxford and Cambridge MA.
- Broome, J. (1993). Goodness is Reducible to Betterness: The Evil of Death is the Value of Life. In Koslowski, P. and Shionoya, Y., editors, *The Good and the Economical: Ethical choices in Economics and Management*, pages 70–84. Springer, Berlin.
- Broome, J. (2004). *Weighing Lives*. Oxford University Press, Oxford.
- Broome, J. (2013). The Badness of Death and the Goodness of Life. In Bradley, B., Feldman, F., and Johansson, J., editors, *The Oxford Handbook of Philosophy of Death*, pages 218–233. Oxford University Press, Oxford.

- CAB 113/35 (1944). *War Cabinet and Cabinet: Home Defence Executive: Secretary's Files. Operation CROSSBOW: deception.* The National Archives, Kew, Richmond.
- CAB 121/213 (1944). *Cabinet Office: Flying bomb and Rocket attacks on UK Vol. III.* The National Archives, Kew, Richmond.
- CAB 121/214 (1944). *Cabinet Office: Special Secret Information Centre: Files. B/Defence. Flying bomb and Rocket attacks on UK Vol. IV.* The National Archives, Kew, Richmond.
- CAB 121/215 (1944-6). *Cabinet Office: Special Secret Information Centre: Files. B/Defence. Flying bomb and Rocket attacks on UK Vol. V.* The National Archives, Kew, Richmond.
- CAB 154/49 (1943-5). *War Cabinet and Cabinet Office: London Controlling Section Correspondence and Papers. Plan CROSSBOW and other "V" Weapons.* The National Archives, Kew, Richmond.
- Campbell, C. (2012). *Target London. Under Attack From the V-Weapons During WWII.* Abacus, London.
- Cruft, R. (2004). Rights: Beyond Interest Theory and Will Theory. *Law and Philosophy*, 23(4): 347–397.
- Darwall, S. L. (1977). Two Kinds of Respect. *Ethics*, 88(1): 36–49.
- Davidson, D. (1974). On the Very Idea of a Conceptual Scheme. *Proceedings and Addresses of the American Philosophical Association*, 47: 5–20.
- Davis, N. (1984). Abortion and Self-Defense. *Philosophy & Public Affairs*, 13(3): 175–207.
- Delmer, S. (1971). *The Counterfeit Spy.* Coronet, London.
- Draper, K. (1993). Fairness and Self-Defense. *Social Theory and Practice*, 19(1): 73–92.
- Draper, K. (2004). Epicurean Equanimity Towards Death. *Philosophy and Phenomenological Research*, 69(1): 92–114.

- Draper, K. (2013a). Death and Rational Emotion. In Bradley, B., Feldman, F., and Johansson, J., editors, *The Oxford Handbook of Philosophy of Death*, pages 297–316. Oxford University Press, Oxford.
- Draper, K. (2013b). Epicurus on the Value of Death. In Taylor, J. S., editor, *The Metaphysics and Ethics of Death*, pages 71–79. Oxford University Press, Oxford.
- Edmonds, D. (2013). *Would You Kill the Fat Man? The Trolley Problem and What Your Answer Tells Us about Right and Wrong*. Princeton University Press, Cambridge MA.
- Eliot, G. (1994 (1874)). *Middlemarch* (Introduction and Notes by Doreen Roberts). Wordsworth Classics.
- Epicurus (1926). Letter to Menoeceus. In Bailey, C., editor, *Epicurus: The Extant Remains*, pages 83–93. Oxford University Press, Oxford.
- Epicurus (2006). *Letter to Menoeceus*. Available online at www.epicurus.info/etexts.html (accessed June 18, 2012).
- Feinberg, J. (1989). *The Moral Limits of the Criminal Law Volume 3: Harm to Self*. Oxford University Press, New York NY.
- Feldman, F. (1991). Some Puzzles About the Evil of Death. *The Philosophical Review*, 100(2): 205–227.
- Feldman, F. (1992). *Confrontations With the Reaper. A Philosophical Study of the Nature and Value of Death*. Oxford University Press, New York NY.
- Feldman, F. (2006). Actual Utility, The Objection from Impracticality, and the Move to Expected Utility. *Philosophical Studies*, 129: 49–79.
- Fischer, J. M. (2006). Epicureanism About Death and Immortality. *The Journal of Ethics*, 10: 355–381.
- Fischer, J. M. and Ravizza, M. (1992). Introduction. In *Ethics. Problems and Principles* (edited by J. M. Fischer and M. Ravizza), pages 1–57. Harcourt Brace Jovanovich College Publishers, Orlando FL.

- Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, 5: 5–15. Reprinted in her *Virtues and Vices* (1978), Oxford: Basil Blackwell.
- Glover, J. (1977). *Causing Death and Saving Lives*. Penguin Books, London.
- Goodpaster, K. E. (1978). On Being Morally Considerable. *Journal of Philosophy*, 75(6): 308–325.
- Griffin, J. (2008). *On Human Rights*. Oxford University Press, Oxford.
- Hanna, J. (2012). The Moral Status of Nonresponsible Threats. *Journal of Applied Philosophy*, 29(1): 19–32.
- Hanser, M. (2008). The Metaphysics of Harm. *Philosophy and Phenomenological Research*, 77(2): 421–450.
- Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., and Mikhail, J. (2007). A Dissociation Between Moral Judgments and Justifications. *Mind & Language*, 22: 1–21.
- Hetherington, S. (2013). Where is the Harm in Dying Prematurely? An Epicurean Answer. *Journal of Ethics*, 17: 79–97.
- Hill, T. (1991). *Autonomy and Self-Respect*, chapter ‘Autonomy and Benevolent Lies’, pages 25–42. Cambridge University Press, Cambridge.
- Hobbes, T. (1994). *Leviathan* (edited by Edwin Curley). Hackett, Indianapolis IN.
- Howard, M. (1990). *British Intelligence in the Second World War. Volume Five: Strategic Deception*. HMSO, London.
- Jones, R. V. (1978). *Most Secret War. British Scientific Intelligence 1939-1945*. Hamish Hamilton, London.
- Kagan, S. (1988). The Additive Fallacy. *Ethics*, 99(1): 5–31.
- Kagan, S. (1991). Replies to my Critics. *Philosophy and Phenomenological Research*, 51(4): 919–28.

- Kamm, F. (1986). Harming, Not Aiding and Positive Rights. *Philosophy & Public Affairs*, 15(1): 3–32.
- Kamm, F. (1987). The Insanity Defense, Innocent Threats, and Limited Alternatives. *Criminal Justice Ethics*, 6(1): 61–76.
- Kamm, F. (1992a). *Creation and Abortion: An Essay in Moral and Legal Philosophy*. Oxford University Press, New York NY.
- Kamm, F. (1992b). Non-consequentialism, the Person as an End-in-Itself, and the Significance of Status. *Philosophy & Public Affairs*, 21(4): 354–389.
- Kamm, F. (1993). *Morality, Mortality Volume I: Death and Whom to Save From It*. Oxford University Press, New York NY.
- Kamm, F. (1996). *Morality, Mortality Volume II: Rights, Duties, and Status*. Oxford University Press, New York NY.
- Kamm, F. (2007). *Intricate Ethics. Rights, Responsibilities, and Permissible Harm*. Oxford University Press, New York NY.
- Kamm, F. (2012). *The Moral Target: Aiming at Right Conduct in War and Other Conflicts*. Oxford University Press, New York NY.
- Korsgaard, C. M. (1996). Personal Identity and the Unity of Agency: A Kantian Response to Parfit. In *Creating the Kingdom of Ends*, pages 363–397. Cambridge University Press, Cambridge.
- KV 2/69 (1944). *Juan PUJOL-GARCIA, codename GARBO: copies of communications by letter and WT*. The National Archives, Kew, Richmond.
- Lamont, J. (1998). A Solution to the Puzzle of When Death Harms its Victims. *Australasian Journal of Philosophy*, 76: 198–212.
- Lang, D. (1951). A Reporter At Large. A Romantic Urge. Interview with Wernher von Braun. *The New Yorker*, 27: 75–93.
- Lazar, S. (2009). Responsibility, Risk, and Killing in Self-Defense. *Ethics*, 119(4): 699–728.

- Levenbook, B. B. (2013). Welfare and Harm after Death. In Taylor, J. S., editor, *The Metaphysics and Ethics of Death*, pages 188–209. Oxford University Press, Oxford.
- Levine, S. (1984). The Moral Permissibility of Killing a “Material Aggressor” in Self-Defense. *Philosophical Studies*, 45(1): 69–78.
- Liao, M. S. (2010). The Basis of Human Moral Status. *The Journal of Moral Philosophy*, 7: 159–179.
- Lippert-Rasmussen, K. (1996). Moral Status and the Impermissibility of Minimizing Violations. *Philosophy & Public Affairs*, 25(4): 333–351.
- Locke, J. (1990). *Second Treatise on Civil Government* (edited by C. B. Macpherson). Hackett, Indianapolis IN.
- Luper, S. (2004). Posthumous Harm. *American Philosophical Quarterly*, 41: 63–72.
- Luper, S. (2009a). Death. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2009/entries/death/>, Summer 2009 edition.
- Luper, S. (2009b). *The Philosophy of Death*. Cambridge University Press, Cambridge.
- Luper, S. (2013a). Exhausting Life. *Journal of Ethics*, 17: 99–119.
- Luper, S. (2013b). Retroactive Harms and Wrongs. In Bradley, B., Feldman, F., and Johansson, J., editors, *The Oxford Handbook of Philosophy of Death*, pages 317–335. Oxford University Press, Oxford.
- Masters, E. L. (2007 (1915)). *Spoon River Anthology* (edited by Lisa M. Miller and Mary Grimes). Prestwick House, Clayton DE.
- McIntyre, A. (2011). Doctrine of Double Effect. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/fall2011/entries/double-effect>, Fall 2011 edition.
- McMahan, J. (1988). Death and the Value of Life. *Ethics*, 99(1): 32–61.

- McMahan, J. (1994). Self-Defense and the Problem of the Innocent Attacker. *Ethics*, 104(2): 252–90.
- McMahan, J. (2002). *The Ethics of Killing. Problems at the Margins of Life*. Oxford University Press, New York NY.
- McMahan, J. (2005a). Our Fellow Creatures. *The Journal of Ethics*, 9: 353–380.
- McMahan, J. (2005b). The Basis of Moral Liability to Defensive Killing. *Philosophical Issues*, 15(1): 386–405.
- McMahan, J. (2009). *Killing in War*. Clarendon Press, Oxford.
- Mill, J. S. (1998a). *On Liberty and Other Essays* (edited by John Gray), chapter ‘On Liberty’, pages 5–130. Oxford University Press, Oxford.
- Mill, J. S. (1998b). *On Liberty and Other Essays* (edited by John Gray), chapter ‘Utilitarianism’, pages 131–204. Oxford University Press, Oxford.
- Moore, M. S. (2009). *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*. Oxford University Press, Oxford.
- Nabokov, V. (1975). Translator’s Introduction. In *Eugene Onegin*. Princeton University Press, Princeton NJ.
- Nagel, T. (1974). Subjective and Objective. *Philosophical Review*, 83: 435–50. Reprinted in Nagel, T., *Mortal Questions* (1979), pp. 196–214. Cambridge University Press, Cambridge.
- Nagel, T. (1979). *The Limits of Objectivity. The Tanner Lectures on Human Values*, chapter III. Ethics, pages 119–139. Available online at <http://tannerlectures.utah.edu/lectures/documents/nagel80.pdf>.
- Nagel, T. (1986). *The View From Nowhere*. Oxford University Press, New York NY.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books, New York NY.
- Nozick, R. (1989). Importance and Weight. In *The Examined Life*, pages 170–181. Simon & Schuster Paperback, New York NY.

- Nussbaum, M. (1999). Reply to Papers in Symposium on Nussbaum, The Therapy of Desire. *Philosophy and Phenomenological Research*, 59(3): 811–819.
- Olson, E. (2013). The Epicurean View of Death. *Journal of Ethics*, 17: 65–78.
- Otsuka, M. (1994). Killing the Innocent in Self-Defense. *Philosophy and Public Affairs*, 23(1): 74–94.
- Otsuka, M. (2008). Double Effect, Triple Effect and the Trolley Problem: Squaring the Circle in Looping Cases. *Utilitas*, 20: 92–110.
- Otsuka, M. (2011). Are Deontological Constraints Irrational? In Bader, R. M. and Meadowcroft, J., editors, *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*, pages 38–55. Cambridge University Press, Cambridge.
- Overland, G. (2011). Dividing Harm. *Journal of Moral Philosophy*, 8: 547–566.
- Parfit, D. (2011). *On What Matters. Volume I*. Oxford University Press, Oxford.
- Pettit, P. (2012). *On The People's Terms. A Republican Theory and Model of Democracy*. Cambridge University Press, Cambridge.
- Philips, M. (1987). Weighing Moral Reasons. *Mind*, 96(383): 367–375.
- Philpott, D. (2010). Sovereignty. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2010/entries/sovereignty/>.
- Pitcher, G. (1984). The Misfortunes of the Dead. *American Philosophical Quarterly*, 21: 119–134.
- PREM 3/111A (1944-5). *Prime Minister's Office: Operational Correspondence and Papers. CROSSBOW post-attack*. The National Archives, Kew, Richmond.
- Quinn, W. S. (1989a). Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing. *The Philosophical Review*, 98(3): 287–312.
- Quinn, W. S. (1989b). Actions, Intentions and Consequences: The Doctrine of Double Effect. *Philosophy & Public Affairs*, 18: 334–351.

- Quong, J. (2009). Killing in Self-Defense. *Ethics*, 119(3): 507–537.
- Quong, J. (2012). Liability to Defensive Harm. *Philosophy & Public Affairs*, 40(1): 45–77.
- Rabkin, J. A. (2005). *Law Without Nations? Why Constitutional Government Requires Sovereign States*. Princeton University Press, Princeton NJ.
- Rachels, J. (1990). *Created From Animals. The Moral Implications of Darwinism*. Oxford University Press, Oxford and New York NY.
- Rachels, J. (2000). Naturalism. In *The Blackwell Guide to Ethical Theory*, pages 74–91. Blackwell Publishers, Oxford.
- Rawls, J. (1971a). *A Theory of Justice (Original Edition)*. The Belknap Press of Harvard University Press, Cambridge MA.
- Rawls, J. (1971b). *A Theory of Justice (Revised Edition)*. The Belknap Press of Harvard University Press, Cambridge MA.
- Rawls, J. (2001). *Justice as Fairness. A Restatement*. Harvard University Press, Cambridge MA.
- Scanlon, T. (1998). *What We Owe To Each Other*. The Belknap Press of Harvard University Press, Cambridge MA.
- Scheffler, S. (1982). *The Rejection of Consequentialism*. Clarendon Press, Oxford.
- Silverstein, H. S. (1980). The Evil of Death. *The Journal of Philosophy*, 77(7): 401–424.
- Singer, P. (1980). *Practical Ethics (3rd edition)*. Cambridge University Press, Cambridge.
- Tadros, V. (2011). *The Ends of Harm. The Moral Foundations of the Criminal Law*. Oxford University Press, Oxford.
- Thomson, J. (1985). The Trolley Problem. *The Yale Law Journal*, 94(6): 1395–1415.

- Thomson, J. J. (1971). A Defense of Abortion. *Philosophy & Public Affairs*, 1(1): 47–66.
- Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist*, 59: 204–217.
- Thomson, J. J. (1990). *The Realm of Rights*. Harvard University Press, Cambridge MA.
- Thomson, J. J. (1991). Self-Defense. *Philosophy and Public Affairs*, 20(4): 283–310.
- Thomson, J. J. (2008). Turning the Trolley. *Philosophy and Public Affairs*, 36(4): 359–374.
- Velleman, J. D. (2005). *Self to Self. Selected Essays*, chapter ‘Love as a Moral Emotion’, pages 70–109. Cambridge University Press, Cambridge.
- Voorhoeve, A. (2009). Frances Kamm. In Search of the Deep Structure of Morality. In *Conversations on Ethics*, pages 15–42. Oxford University Press, Oxford.
- Voorhoeve, A. (2014). How Should We Aggregate Competing Claims? *Ethics*, 125: 64–87.
- Walzer, M. (1977). *Just and Unjust Wars*. Basic Books, New York NY.
- Warren, J. (2004). *Facing Death. Epicurus and His Critics*. Oxford University Press, Oxford.
- Wenar, L. (2011). Rights. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/fall2011/entries/rights/>, Fall 2011 edition.
- Williams, B. (1973a). A Critique of Utilitarianism. In Smart, J.C.C. and Williams, B., editors, *Utilitarianism: For and Against*, pages 77–150. Cambridge University Press, Cambridge.
- Williams, B. (1973b). The Makropulos Case: Reflections on the Tedium of Immortality. In *Problems of the Self*, pages 82–100. Cambridge University Press, Cambridge.