

The London School of Economics and Political Science

# **MATTERS OF LIFE AND DEATH**

Todd Nugent Karhu

A thesis submitted to the Department of Philosophy, Logic and Scientific  
Method for the degree of Doctor of Philosophy on 31 May 2019

## **Abstract**

This thesis is about four ethical problems concerning death. The first two chapters are about the reasons we have to prolong our own lives. In Chapter 1, I argue that we can defensibly maintain that it would not have been in our interests to have been brought into existence earlier than we were, while also holding that it is often against our interests to die earlier than we otherwise would. In Chapter 2, I argue that one can have a reason to continue one's conscious life even when the particular contents of it will not be good on balance. The last two chapters are about the morality of taking life. In Chapter 3, I argue against the common-sense view that the strength of the moral objection to killing a person is unaffected by certain facts about her, chief among them what she stands to lose by dying. In Chapter 4, I discuss the nature and extent of the obligations that are incurred by those who wrongfully kill, or who mistreat people in other ways that seem, on the surface, impossible to redress.

## Declaration

I certify that this thesis is my own work. I confirm that Chapter 3 is forthcoming in the journal *Utilitas* and that most of section 1 of Chapter 4 is published in *Analysis* as Karhu (2019). This dissertation contains 51,206 words.

## Acknowledgments

Since discovering philosophy some eight years ago, I have learned a great deal from a great many people. From many others I have benefitted in ways that are even more important than philosophy. In part because I could not hope to name them all, with apologies to the many teachers and friends I am excluding, I shall confine myself to listing those from whom I have benefitted the most in writing this thesis.

For valuable discussion, correspondence, or written comments on my work, I am grateful to Susanne Burri, Goreti Faria, David Kinney, Chris Marshall, Max Muir, Bryan Roberts, Tom Rowe, and Aron Vallinder.

For *invaluable* discussion, correspondence, *and* written comments, I am particularly indebted to Tomi Francis, Kacper Kowalczyk, Michal Masny, Jeff McMahan, Aidan Penn, Bastian Steuwer, and Alex Voorhoeve.

For camaraderie, love, and encouragement, I am grateful to many of those enumerated above and especially to Rebecca Bowen and Nils Reimer.

I am most of all grateful to Michael Otsuka. If his quality as a supervisor is surpassed, it is only by his quality as a friend.

# Contents

Prelude: Four Lucretian Arguments .....	1
Summaries of the Chapters .....	19
Chapter 1: Was I Born Too Late? .....	28
Chapter 2: Is Conscious Life Itself a Good? .....	78
Chapter 3: Not All Killings Are Equally Wrong.....	103
Chapter 4: Non-Compensable Harms .....	145
Bibliography .....	216

## Prelude: Four Lucretian Arguments

Many of us are distressed by the prospect of death, and Lucretius thought he could offer some philosophical therapy:

Look back again to see how the past ages of everlasting time, before we are born, have been as naught to us. These then nature holds up to us as a mirror of the time that is to come, when we are dead and gone.<sup>1</sup>

Philosophers have reconstructed Lucretius' argument in a number of ways, only some of which have been adequately distinguished. In this short, prefatory chapter, I will examine four incarnations of the Lucretian argument and conclude that only one of them has a chance of being successful. That one is the version that aims to establish that an earlier death cannot be bad for its subject, on the grounds that an earlier birth would not have been good for her. In the first chapter of the thesis that follows, I argue that even this Lucretian argument fails.

---

<sup>1</sup> Bailey (trans.), 1963, Book III, §ii, lines 973–5.

## FOUR LUCRETIAN ARGUMENTS

Lucretian arguments are a species of those arguments that aim to dispel or undermine some attitude or belief X by showing that it ought to conform to another attitude or belief Y. There are two ways in which an argument like this might, in the way I will use the phrase, “fail on its own terms”. The first is when no actual person holds X. The second is when the reason people hold X is manifestly because of properties of it that are not also properties of Y. I shall attempt to show that all but one version of Lucretius’ argument fails in one of these two ways. The other versions either target an attitude or belief that no one actually holds, or else rely for their success on pointing to properties of prenatal nonexistence that are not shared by nonexistence due to death.<sup>2</sup>

We can begin by isolating two broad kinds of Lucretian argument. The first kind of argument targets the negative emotional responses we have toward death: fear, dread, anxiety, dismay, and so forth. At the risk of oversim-

---

<sup>2</sup> It is implausible, of course, that an individual begins to exist precisely at birth. I simply use the word as a placeholder for whatever event does mark the beginning of one’s existence.

plification, I will limit my arguments to fear and its cognate emotions. This version of the argument, then, holds that because we do not, as a matter of fact, have emotions or attitudes comparable to fear toward the nonexistence that preceded our births, we are rationally required not to have them toward death, either. I shall call these “attitudinal arguments”.

The other kind of argument targets not an attitudinal or emotional response to death but rather a normative proposition about it. Conceived this way, the Lucretian argument is meant to show us that death is not *bad* for us, or alternatively that we have no prudential reason to want not to die. These arguments hold, roughly, that because the nonexistence that preceded our births was not bad for us, neither is that which will follow our deaths. I shall call these “normative arguments”.

There are also two targets against which either kind of argument can be levelled, depending on how we conceive of death. The first comes from conceiving death as the event-type that is the cessation of one’s own existence, whenever it happens. For simplicity’s sake, I will henceforth call this conception of death “death itself”. Alternatively, a Lucretian argument can be levelled against an attitude or belief we have about a particular token death as compared to some alternative death that we would otherwise have had in-



stead. I assume that in the context of the Lucretian argument the relevant alternative death is a later one, such that the argument is meant to change some attitude or belief we have about dying earlier than we otherwise would.<sup>3</sup> In short, we have four possible versions of the Lucretian argument—some of which, we shall see, have more nuanced variations of their own.

Of these four arguments, the only version which has a chance of being successful, on its own terms, is this last one. The others either target an attitude or belief that no one actually holds or else rely for their success on features of the nonexistence that precedes birth that are clearly not present when it comes to the nonexistence that follows death.

### **Attitudinal arguments**

I begin with those Lucretian arguments that aim to show that fear is an irrational emotional reaction to death, whether conceived as death itself or as an earlier death. Note that the sense of “irrational” at play here is not the sense in which an emotion is irrational whenever it is bad for us and it is in our control to rid ourselves of it. Our emotional attitudes toward death are often

---

<sup>3</sup> For a comprehensive discussion of how to individuate these alternatives, a difficult issue from which I will prescind entering into here, see McMahan (2002), pp. 98–107.

negative experiences in themselves, and they may prevent us from doing things that would be good for us. Lucretian arguments that the fear of death is irrational use the word in its other sense, according to which an attitude is rational if and only if it is a fitting or appropriate response to its object. Still, even those who think that the fitting-attitude conception of rationality is not action-guiding should care about whether fear is an appropriate response to death. Coming to believe fear is not a fitting attitude to have toward something may be the most effective way to disabuse ourselves of it.

#### DREAD OF DEATH ITSELF

Before continuing, we should distinguish two distinct senses of the word “fear”. The failure to draw this distinction can lead to mistakes. Thus Shelly Kagan has argued that, although it can be perfectly rational to fear an *earlier* death, which for all one knows might happen, it is never rational to fear death itself.<sup>4</sup> His reasoning is that fear is an attitude that is appropriate only toward things about which we are reasonably uncertain, of which death itself is not one. But Kagan’s argument is unsuccessful because it presupposes that the sense of “fear” in which we might fear an early death is the same sense of

---

<sup>4</sup> Kagan (2012), pp. 292–297.

“fear” in which people fear certain death. But there are, it seems to me, two relatively distinct emotions that are sometimes called “fear”. The first is indeed an emotion that we typically have only toward events which we perceive to be bad and are uncertain will occur; to avoid ambiguity I will call this emotion “trepidation”. The other is an attitude that we have toward events which we perceive as bad and are reasonably certain they will occur; I will call this sense “dread”.

To see the difference, imagine driving down the highway at night, in the middle of nowhere and realising that your car is nearly out of fuel. In the sense of “fear” which means “trepidation”, you will probably fear running out of fuel. But once you discover that there is no chance at all of passing a filling station before your tank runs dry, this emotion would disappear. You would not continue to feel trepidation about the prospect running out of fuel. But you might nonetheless dread the coming night, which you recognise will inevitably be passed in your car in the bitter cold.

I grant that, in the sense of “fear” that means “trepidation”, it is indeed irrational to fear one’s inevitable death. But Kagan’s intention is to disabuse us of an attitude we actually have, and no one, I submit, feels trepidation about her inevitable death. Rather, those who fear their inevitable death fear

it in the sense that means “dread”, in that it fills them with terror when they think of it. Suppose a white-hot iron is certain to be held against my skin tomorrow. It does not seem irrational to dread this event. Indeed, a person who greeted the prospect with equanimity would seem bizarre, or at best superhumanly unflappable—an unreasonably high bar for rationality for mortals. Since, presumably, most people who profess to be afraid of death are under no illusions that they will escape it in the end, it is safe to assume that if the object of their fear is simply death itself, then the emotion they have is dread.

In the Woody Allen film *Whatever Works*, the protagonist’s wife awakens to his panicked screams.

—“I’m dying! I’m-I’m dying!”

—“Should I call an ambulance!?”

—“No, not now. I mean eventually!”

Insofar as we dread death, we might dread it whenever it may happen. With the distinction between trepidation and dread in hand, we can ask whether there is a Lucretian argument to show that the dread of certain death is irrational.

As a rally against the dread of death, the Lucretian argument appears to be this:

- (1) We do not dread the nonexistence before our birth, but
- (2) There is no difference between that and the nonexistence that follows our death that could make the latter more worthy of dread (“nature’s mirror”), and
- (3) It is irrational to dread one thing but not another if there is no difference between them that could make the one more worthy of dread than the other, so
- (4) It is irrational for us to dread death.

Of course, it may be that, in the end, it is our attitude toward the nonexistence before our birth that should be brought into line. Then, of these four claims, we would conclude that we should revise the attitude cited in (1) instead of accepting (4). I shall not dwell on this possibility here, which may be a third way in which a Lucretian argument fails “on its own terms”.

Now, there is a sense of “dread” and some of its synonyms whereby to dread something is, roughly, to anticipate it with horror. Because it may be a

psychological fact that we cannot anticipate what we believe has already happened, this argument may seem to fail trivially on the grounds that (2) is false. But because we will not experience being dead, the only sense of “anticipation” in which we can anticipate death with horror is the sense that means that we believe it to be in the future. Though as a matter of fact we may dread only those events we believe to be in the future, it does not follow that these are the only events that are *worthy* of dread, or some comparable emotion. It may well be that not only do we not dread the nonexistence that preceded our birth, we also think it of as *unworthy* of dread, and so reflection upon it will lead us to conclude the same about death.

Insofar as we dread death itself, we might dread either the event itself—the cessation of existence—or else the *period* of postmortem nonexistence that follows death. The Lucretian argument against the dread of the event itself fails because there is no comparable event when it comes to our prenatal nonexistence. The relevant analogue to the event of death is the event of coming into existence.

The most natural reading of Lucretius’ poem is that his target is not the reader’s dread of the event of death but rather of the period of nonexistence that follows it. The Lucretian argument against the dread of the period

of postmortem nonexistence fails for another reason: this dread is directed at a property not shared by the period of nonexistence that precedes our birth. As Frances Kamm has pointed out, unlike the nonexistence that follows one's death, the nonexistence that precedes one's existence will eventually end—it will end when one begins to exist.<sup>5</sup> Had he been certain of his resurrection four days after his (first) death, Lazarus would not, I think, have dreaded his midlife nonexistence. Indeed, he may well have regarded it with all the calmness Lucretius recommends. This seems to show that what we dread about postmortem nonexistence is not that it is a state of affairs in which we do not exist, but that it is one in which we will never exist at a later time. The very feature we might dread about it is the one that is absent when it comes to the nonexistence that preceded our lives.

#### DREAD OR TREPIDATION ABOUT AN EARLIER DEATH

Although the Lucretian argument cannot undermine the dread of death itself, it might be levelled instead against the dread of dying an earlier death than one could have. As I have said, while the sense of fear that means “trepida-

---

<sup>5</sup> Kamm (1993), pp. 42–44.

tion” is rational only toward events we are reasonably uncertain will occur, the sense that means “dread” does not have that condition. We can begin, then, by asking whether the Lucretian argument can show that it is irrational to dread an earlier death—even when it is certain—and then ask whether it can show that it is irrational to feel trepidation about dying early.

Does anyone actually *dread* an early death, such that the attitude they have toward it would be present even if that early death became reasonably certain to occur? There is some reason to think so. Suppose I am extremely confident that I will die in 2030. I then come to realise that that belief was justified but false. A new mountain of evidence leads me to become just as confident that I will instead die in 2025. My dread of dying in 2025 would be much greater than was my dread of dying in 2030. Perhaps this shows that we do in fact dread dying an earlier death.

But the explanation of this is not that I dread dying earlier but rather that I dread death itself, and if an event is already an object of one’s dread, then that dread intensifies as the event draws nearer to the present. That is, I now dread dying in 2025 more than I dreaded dying in 2030 not because in the former case my death is earlier, but because it is sooner.

To see this, compare:



(a) My dread, in 2020, of my death in 2025 rather than 2030.

(b) My dread, in 2025, of my death in 2030 rather than 2035.

(c) My dread, in 2020, of my death in 2030 rather 2035.

The intensity of (a) and (b) would not be that of (c). And (a) would be roughly of the same intensity and character as (b). This suggests that insofar as I dread an earlier death I dread it not because it is earlier, in the sense that it is (more) pre-mature, but because it is sooner. Because we do not feel a distinct sense of dread for early death, the Lucretian argument against that emotion would not be on target.

Recall Shelly Kagan's argument that fearing death is irrational:

(1) It is rational to fear something only if it is reasonably uncertain it will occur, but

(2) Death is not reasonably uncertain, so

(3) It is not rational to fear death.

I said earlier that this argument is not useful, since insofar as (1) is true, the sense of “fear” it employs is not the same sense impugned by (3). The negative attitude we have toward certain death is one that is appropriate toward something we will undergo which seems to us very severe, regardless of how likely it is. But, as Kagan allows, in the sense of “fear” I have called “trepidation”, it can be rational to fear early death. Is there a Lucretian argument to the effect that trepidation about the possibility of dying earlier than we might is irrational if we do not have a similar feeling toward the possibility of being born later than we might have been?

I do not think there is an independent Lucretian argument to that effect. If the prospect of an earlier death gives rise to trepidation, that is because we feel trepidation about the loss of good life of which dying earlier will deprive us. And whether someone feels trepidation about the loss of something depends on whether they believe it would have been good for them to have it. Simply as a matter of fact, most people do not feel that being born earlier than they were would have been good for them, at least not to anything like the degree to which they believe that a later death would be good. The success of the Lucretian argument that trepidation toward the prospect of an earlier death is irrational depends on the success of the Lucre-

tian argument (or another argument to the effect) that dying a later death is not good for its subject. Reflection on our actual emotion toward the possibility of a later birth can do nothing on its own to show that trepidation about the prospect of an earlier death irrational, for that trepidation would be irrational only if we have already come to believe, independently, that early death is not bad for us.

### **Normative arguments**

#### THE BADNESS OF DEATH ITSELF

The other target of a Lucretian argument is not an emotional response to death but a normative proposition: roughly, that death is bad for the person who dies. On this type of argument, we begin not with an empirical observation about how we in fact feel about the time before our birth, but rather with the premise that nothing about that time was bad for us. As before, I begin with death itself. Insofar as the Lucretian argument is meant to show that the *period* that follows death is not in itself bad for a person, it seems to me useless; no one believes that it is in itself bad for a person not to exist.

For reasons akin to those we considered when discussing the dread of death itself, there is also no viable Lucretian argument against the claim that the event-type “ceasing to exist” is bad for a person. There are some reasons to think that it is in itself bad for a person to cease to exist; for example, that death takes from her the continuation of the goods of life that she already possesses, or that death means that everything is “all over” for her.<sup>6</sup> But there is no comparable consideration when it comes to birth or what precedes it.

#### THE BADNESS OF AN EARLY DEATH

We have covered three Lucretian arguments so far: those against the fear of death itself, the fear of an early death, and the badness of death itself. Let us turn, then, to the final Lucretian argument, which holds that early death is not bad for a person because later birth was not. Of course, there is one sense in which it would be more-or-less straightforwardly good for a person either to have had an earlier birth or to have a later death. Provided the extra life it would afford her is on balance good, and provided it does not, for

---

<sup>6</sup> Kamm (1993). I will discuss the possibility that death is in itself bad for its subject at length in Chapter 2.

derivative reasons, lower the quality of the life she would *anyway* have had, then by having an earlier birth or a later death the person would have had a longer, better life. Indeed, a number of philosophers accept what has been called the “Life Comparative Account” (LCA) of the badness of death, according to which death is bad for a person just in case it makes her lifetime wellbeing lower on balance than it would otherwise have been.<sup>7</sup>

Because the LCA seems, at first blush, to have as a corollary that an earlier birth would have benefitted someone if the life she receives thereby is good, a proponent of the account who is unhappy with that result must try to show that the LCA doesn’t have it.<sup>8</sup> But the Lucretian has the opposite problem. She needs to be able to say why the LCA is not the right account of the badness of death.

A Lucretian might argue as follows: “I would not have had a better life by being born earlier. Therefore, the LCA is false”. But it is hard to see how being born earlier could not have given me a better life. Surely the length and quality of a life largely determine how good a life it is.

---

<sup>7</sup> The name is McMahan’s (2002), p. 105. Defenders of the LCA, or of a relevantly similar view, include Thomas Nagel (1970), Fred Feldman (1992), John Broome (1993; 2004); and Ben Bradley (2008; 2009).

<sup>8</sup> As Feldman (1992, p. 155) does. At least one defender of the Life Comparative account is willing to accept that it would have been better for a person to be born earlier when that would have increased her lifetime wellbeing, however—see Bradley (2009), p. 64.

A better response on the part of the Lucretian is this: “There is an important sense of ‘good’ according to which having a better life is not always *ipso facto* good for a person. Therefore, the LCA is at most an incomplete account of when and why death is bad for its subject.”

In what sense of “good” could it be true that having had a better life might not have been good for its subject? I suggest that it is in the sense captured by what one has a present prudential reason to want.<sup>9</sup> We can coherently maintain that by having been born earlier a person’s life would have contained a greater sum of wellbeing or value, while denying that at every given time within it the person has a prudential reason to want that to be the case. The Lucretian opponent of the LCA can therefore say the following. “Although having been born earlier than I was might have made my life better, in the sense that it contained less total wellbeing or value, it is not *now* in my interest to have in that way achieved a better life.”

This fourth and final Lucretian argument has considerable force. It does not fail on its own terms—most of us do not regret that we were not born earlier, but do regret that we will not live longer with a life that is good; and there is at least no *obvious* feature of deprivations of life due to death that is

---

<sup>9</sup> Alternatively, it is the sense captured by what we have a present time-relative interest in having. I will discuss this notion in the next chapter.

not also a feature of deprivations of it due to a later birth. After all, death seems bad in large part because of all the good things it takes from us, and similarly good things might have been ours, had we only begun to exist earlier.

How death can be bad for its subject is something like the central organising question of this thesis. In the first chapter that follows, I will attempt to refute this last and most significant incarnation of Lucretius' argument. In the chapters which follow that, I will try to show how the ways in which death can be bad for its subject bear on some fundamental ethical issues—some familiar and others largely untrodden—regarding how we should think about our own lives and how we should respect the lives of others.

## Summaries of the Chapters

### Chapter 1: Was I born too late?

Most if not all of us have the asymmetrical attitudes toward birth and death identified by that fourth and final Lucretian argument. Provided we believe the life we would have gone on to live would have been overall good, it seems to us a deep misfortune that we must die sooner rather than later. But upon considering the possibility that we came into existence earlier than we did, few of us would consider this a comparable misfortune. In the first chapter of this thesis, I will argue that these asymmetrical attitudes toward birth and death are just what they should be.

I first consider the idea that the difference in our attitudes is justified on the ground that it is impossible to have come into existence earlier than when one did. Though I will argue that this argument ultimately fails, I will also show how it is more interesting than has been assumed by its opponents.

I then turn to the mainstream explanation for why we care more about dying later than about being born earlier: that we are biased toward the fu-



ture. It is widely believed that this difference in attitudes arises from a more general phenomenon according to which human beings more-or-less systematically prefer positive experiences to be in the future rather than the past, and negative ones to be in the past rather than the future. It is also widely thought that the source of that bias, in turn, is an implicit conviction in the objectivity of time's passage.

But we are not, I argue, biased toward the future, so that cannot be the explanation. Moreover, our preferences about birth and death have nothing to do with the belief that time passes, and truths about the metaphysics of time are irrelevant to whether or not those preferences are defensible. Instead, I claim, our actual attitudes track a concern for what happens to those stages of ourselves that are causally downstream from the present moment.

Finally, I defend the novel position that the relation which underwrites rational prudential concern is itself asymmetric. More specifically, I argue for what I call the "One-Directional View" against the prevailing "Two-Directional View". The One-Directional View tells us that although we should often have prudential concern for what happens to those stages of ourselves that are causally downstream, those stages ought to have no intrinsic prudential concern for what is happening to us now. If I am right, then what previ-

ously appeared to be a temporal bias is justifiable on the basis of the rational prudential concern relation itself. Because a person's death is downstream but her birth is not, the One-Directional View implies that our attitudes toward them are justifiable.

## **Chapter 2: Is Conscious Life Itself a Good?**

I then turn to the question of when one has a prudential reason to go on living. I defend the position that conscious life itself has non-instrumental prudential value for its subject. Accordingly, a person can have a reason to go on living even when the particular constituents of her conscious life are not themselves good on balance.

I begin with a brief clarification of that position and a reply to a criticism of it. I then introduce my central case, in which one can choose between death now and living for a while longer with a conscious life that is otherwise neutral overall in the elements of wellbeing it contains. It seems to me that most people would choose to go on living, which provides at least a *prima facie* case for my conclusion.

Some philosophers would say that I have a reason to go on living in this case because the end of conscious experience is in itself bad for me. I offer some objections to the arguments that have been advanced for that position, a direct argument against it, and several reasons to think that, even if the end of my conscious life really is in itself bad for me, this fact is still unlikely provide a reason for me to continue living.

I then consider some more direct objections to my central argument. I attempt to show why the intuition it evokes cannot be dismissed either on the ground that we tend to be overly optimistic about how good our future will be or on the ground that we are subject to some other bias that renders it epistemically unreliable.

Finally, I return in more detail to the idea that conscious life has positive value for its subject over and above its particular contents, and suggest a deeper reason why this might be the case.

### **Chapter 3: Not All Killings Are Equally Wrong**

Many people would balk at the idea that it could be less seriously wrong to kill someone simply because he is particularly old, sick, or unhappy. This is

curious, since these factors typically affect how harmful death is, and the strength of the moral objection to an act usually correlates with its harmfulness, other things being the same. The view that the wrongness of killing someone does not depend upon factors like these, a view whose most prominent contemporary proponent is Jeff McMahan, is known as the “Equal Wrongness Thesis”.

Kasper Lippert-Rasmussen has argued that we should reject the Equal Wrongness Thesis on the basis of the moral equivalence of killing someone and knocking someone unconscious. Several authors have raised objections to his arguments, however, and there are other issues with his approach that have not yet been noted. In this chapter, I argue that by drawing on some of his insights we can, nonetheless, show that killing someone and knocking her unconscious are morally equivalent in the way that is sufficient to refute the Equal Wrongness Thesis.

I begin with an analysis of Lippert-Rasmussen’s own argument. I identify what I take to be its key premise, discuss the various ways this premise can be attacked and defended, and suggest that his own defence is at best incomplete. The rest of the chapter is an extended argument for the following claim:

(C) It is as wrong to deprive someone of a certain amount of conscious life by knocking her temporarily unconscious in the middle of her life as it is to deprive her of the same amount of conscious life by killing her, other things being the same.

Together with the Equal Wrongness Thesis, (C) implies that it is as wrong to knock someone temporarily unconscious for (say) a month in the middle of her life as it is to kill someone who would otherwise have had decades of happy life ahead of her. Because that result is absurd, I conclude that we should reject the Equal Wrongness Thesis if (C) is true.

#### **Chapter 4: Non-Compensable Harms**

In the final chapter, I turn from questions about the wrongness of killing to the question of what we should do when we have killed. Killing someone, at least on the face of it, is a paradigmatic example of what I call a “non-compensable harm”. A non-compensable harm, as I use the phrase, is any harm whose victim cannot be fully compensated by its perpetrator. This interesting class of actions has received little attention from philosophers. I will try

to show how it raises a number of intriguing questions whose answers have important and far-reaching consequences both practical and theoretical.

I begin the chapter by sorting non-compensable harms in two categories. The first category comprises those harms whose perpetrators can do nothing to compensate their victims, at least beyond a certain point. The second category comprises those harms whose victims can be compensated further, but cannot be compensated fully. This chapter is divided into two extended sections, each corresponding to one of these two categories of non-compensable harms.

In the first main section, I defend what I call the “Non-Compensable Harms Thesis”. According to that thesis, to the extent that he cannot compensate his victim, the perpetrator of a wrongful harm has an undirected duty to promote the impartial good. I offer a defence of that thesis and close the first section by discussing two practical consequences of it.

The second section of the chapter focuses on the largely unexplored question of the limits of proportionality in the liability to compensate. Because in cases of non-compensable harms full compensation is impossible, we have to ask how much compensation, if any, is enough.

I proceed by discussing some structural similarities between compensation and self-defence. I then turn to an examination of what I call the “Received View” of proportionality. To oversimplify, in the context of *defence*, this view tells us that the amount of defensive harm an aggressor is liable to bear to avert the threat he poses is a function of the magnitude of the harm he threatens and the magnitude of the harm which averting the threat will cause him. In the context of compensation, the Received View says, analogously, that the amount of cost a perpetrator is liable to bear to compensate his victim is a function of the cost to him of paying it and the benefit to the victim of receiving it.

The remainder of the chapter is a set of arguments against the Received View as it applies to both compensation and defensive harming. My primary criticism of the Received View is that it cannot account for an important interpersonal dimension of harming; namely, how well off a perpetrator is relative to his victim. I defend the principle that when a perpetrator is culpable for the fact that his victim is worse off than he is, he is liable to compensate her at least until he is no better off than she is—a principle which conflicts with many of the Received View’s judgments about what amount of compensation is disproportionate.

After arguing against the Received View as it applies to the liability to compensate, I argue for what I call the *substantive symmetry* of defensive harm and compensation. This is the thesis that whenever a person is liable to bear some cost to avert a harm that he threatens, he is also liable to bear that cost to counterbalance the harm should it eventuate, and vice versa. That conclusion is, I hope, significant in its own right. But, more importantly for my purposes, together with my arguments about the limits of the liability to compensate, it implies that the Received View of proportionality in self-defence is false.



## Chapter 1: Was I Born Too Late?

As a general characterisation of our actual attitudes, it is hard to deny that we feel very differently about the life we might lose out on by dying earlier and the life we might have lost out on by being born later than we could have been.<sup>1</sup> Imagine you wake up in the hospital and temporarily cannot remember whether you are forty or fifty years old. What you do know is that in either case your expected lifespan is sixty years; if you are forty, you can expect to live another twenty years, if you are fifty, you can expect another ten. Most of us would hope to turn out to be forty. And the salient difference between the two prospects is whether an additional ten years in one's life comes from being born earlier or from dying later.

This indicates that we care less about the additional life we could have had by being born earlier than we care about additional life we could go on

---

<sup>1</sup> As before, I will generally use the term “birth” loosely, to refer to the beginning of one's existence.

to have by dying later. But Lucretius claimed not just that we care *less* about having had additional life by being born earlier, but that we do not care *at all* about the additional good life an earlier birth could have afforded us. Suppose, instead, that regardless of whether you turn out to be forty or fifty, you will die in six months. If we are truly indifferent about any additional past life, one would expect us in this case to be indifferent between being fifty and being forty. But several authors have reported, and I can add myself to their ranks, that in a case like this they would rather be fifty. Does this not indicate that we place at least *some* value on the life we could have had by being born earlier than we were?

Frances Kamm and Jeff McMahan have both proposed that the reason we place at least some value on having been born earlier is that we will then be more likely to have accomplished more in our lives.<sup>2</sup> This seems plausible. For suppose I were to learn that, regardless of whether I am forty or fifty, I will have accomplished the same things in my life. In this case I would care much less whether I am forty or fifty. More generally, several philosophers report that, like Lucretius, they are unconcerned about past *experiential* goods and bads, like pleasures and pains,<sup>3</sup> but when it comes to *non-experiential*

---

<sup>2</sup> Kamm (1993), pp. 36–37; McMahan (2005a), p. 219.

<sup>3</sup> Parfit (1984), pp. 173–4, 180–81; Brueckner and Fischer (1986); Hurka (2003), p. 61.

goods like successes and failures, a number of philosophers have also reported having no intrinsic preference about their temporal location relative to the present moment.<sup>4</sup> And the fact that a longer life is *ceteris paribus* more likely to contain more and greater non-experiential goods can explain why the asymmetry in our attitudes toward birth and death is at least a bit weaker than Lucretius proclaimed.

I shall return to the issue of the temporal location of non-experiential goods much later. In the meantime, to make progress, let us grant Lucretius his stronger claim as it applies to experiential goods. To the extent that the prudential value of the additional life in question is measured in experiential goods only, I care about having additional good life by dying later, but not about having had it by coming into existence earlier.

So far I have spoken only of our actual attitudes, having said nothing about whether or not these attitudes are justifiable. But whether we have a *reason* to want to die later, or lack a reason to want to have been born earlier, is another matter altogether. That is the question I am concerned with answering in this chapter.

Let us label the thesis that this pair of attitudes is, in fact, justifiable “the Asymmetry”.

---

<sup>4</sup> Hurka (2003), pp. 61–62; Kamm (1993), pp. 36–37; McMahan (2005a), p. 219.

*The Asymmetry:* One has a prudential reason to want more good life by dying later, but no prudential reason to want more good life by having come into existence earlier, at least when the life in question is good only in virtue of the experiences it contains.

It might sound odd to speak of a prudential reason to “want” to have been born earlier, since perhaps it seems we can have a reason to want only those things that can actually still happen. I use this phrase in its loose sense, whereby one has a prudential reason to want something just in case it would be in one’s present rational self-interest to have or have had it. Another way to put it is in terms of what McMahan has called an individual’s “time-relative interest”.<sup>5</sup> The Asymmetry, as I will construe it, is the claim that at a given time an individual has a time-relative interest in those good experiences she would secure by postponing her death, but not in those she would have secured by having come into existence earlier than when she did. Let me add to this clarification a few other preliminary remarks.

First, I will generally assume that any additional life one could have had, whether by having been born earlier or by dying later, would not have increased the number or quality of the non-experiential goods or bads in her

---

<sup>5</sup> McMahan (2002), pp. 105–6, 170–74.

life. I will also assume, unless I specify otherwise, that the additional period of life in question is of the same length and quality whether it is appended to the beginning of one's life or its end, and that it is on-balance good life to live. Lastly, I will take for granted that, insofar as a period additional life is worth wanting, it is worth wanting for its own sake. If I would not have my current back pain had I been born earlier than I was, then that itself would generate a reason to want to have been born earlier, but the reason is merely instrumental and thus it is not a counterexample to the Asymmetry as I am conceiving it.

Now, if we reject the Asymmetry, we will then think that we have as much reason to want to have been born earlier as we have to want to die later. (I won't consider the possibility that we have a reason to want to have been born earlier but not to die later, which seems to me unmotivated.) We might therefore reconcile our pre-theoretic attitudes in one of two directions, concluding either that we have as *much* reason to want to be born earlier as we previously took ourselves to have to want to die later, or else that we have as *little* reason to want to die later as we previously took ourselves to have to want to have been born earlier.

Lucretius himself would have wanted us to take the latter path, thereby accepting that we cannot be made worse off when we die earlier than we otherwise would have.<sup>6</sup> It is very hard to believe that a death which deprives its subject of decades of wonderful existence is not bad for her. But it is also highly intuitive that early death is worse for its subject than late birth. Unless we can show how the Asymmetry is defensible, then, we may well be forced to accept one strongly counterintuitive conclusion.

I shall argue that the Asymmetry is true. In the following section, I will begin by discussing, but ultimately rejecting, a well-known argument to the effect that we cannot have a reason to want to have been born earlier because to have been so is metaphysically impossible. Following this discussion, I will consider the prevailing explanation for *why* we prefer later death to earlier birth: that we are in general biased towards the future. I will consider and reject several arguments both for and against the Asymmetry that proceed from this claim. Ultimately, I will concede that if the bias toward the future

---

<sup>6</sup> As I noted in the Prelude, the famous “nature’s mirror” passage in Lucretius has sometimes been interpreted as a rally against the *fear* of death, as opposed to an argument that death is not *bad* for us. But elsewhere in his poem Lucretius is clear that he endorses the latter view. He argues for example: “For, if by chance there is to be grief and pain for a man, he must himself too exist at that time, that ill may befall him” (1963, Book III, §ii, lines 862–4).

really is the source of our preference for later death over earlier birth, then it is more likely than not that the Asymmetry is false.

But the bias toward the future is not, I claim, the source of that preference, because we are not biased toward the future. I will show that our preference for later death over earlier birth is independent of our attitudes toward time, and, consequently, that truths about the metaphysics of time are irrelevant to whether or not that preference is justifiable. I then offer an account of what our *actual* attitudes are, which I call a bias towards the “successive self”.

I then defend the novel view that the grounds of egoistic or prudential concern (what is sometimes styled “what matters in survival”) are asymmetric. In other words, it can be true that a person should now have prudential concern with what happens to her later but that she should not later have prudential concern with what is happening to her now. More specifically, I defend what I call the “One-Directional View”, according to which one ought to be prudentially concerned for only those stages of oneself that are causally downstream, but those downstream stages themselves should have no prudential concern with what is happening to one now.

Discussants of whether temporal bias is justified have approached the question by assuming that a certain type of good is *prima facie* prudentially valuable in the absence of any specification of its temporal location relative to the present, and then asking whether the temporal properties it does have can affect the extent to which it is in one's interest to have it. In contrast, I suggest that what was previously assumed to be a bias toward the future is justifiable on grounds that are internal to the rational prudential concern relation itself. Obviously enough, because death is causally downstream and birth is not, the One-Directional View implies the Asymmetry. I conclude by showing how my position can also account for our relative indifference as to when in our lives non-experiential goods occur.

### **1.1. Could I have come into existence earlier?**

I will begin with a well-known argument for the Asymmetry. In a celebrated article, Thomas Nagel argued that although we have a reason to want to die later when doing so will secure us more good life, we have no reason to want to have been born earlier, on the grounds that it is metaphysically impossible



for someone to have begun to exist earlier than when she in fact did. In his words:

[A]side from the brief margin permitted by premature labor, [one] *could* not have been born earlier: anyone born substantially earlier than he was would have been someone else. . . . Distinct possible lives of a single person can diverge from a common beginning, but they cannot converge to a common conclusion from diverse beginnings.<sup>7</sup>

In brief, since I could not have been born earlier, I can have no reason to want to have been, but because I can die later, I *can* have a reason to want that.

Derek Parfit's response to this argument seems to have been to reject the premise that things cannot be worse for us even if it is metaphysically impossible that they were otherwise.<sup>8</sup> By way of illustration, Parfit suggests that it was rational for the Pythagoreans to regret the fact that the square root of two is an irrational number, even once they came to accept that it

---

<sup>7</sup> Nagel (1970), p. 79 (emphasis in original).

<sup>8</sup> Parfit (1984), p. 175.

would have been logically (and hence metaphysically) impossible for it to have been anything else.<sup>9</sup>

But this does not show that things can be worse for us even if they could not have been otherwise; all it shows is that regret can be an appropriate attitude to take toward those things. While that more modest conclusion may have been Parfit's aim, the more important version of Lucretius' argument concerns not our emotional attitudes toward being born when we were, but whether being born earlier would have been better for us. If Nagel is right that I could not have begun to exist other than when I did, then any possible world in which I was not born when I was is a world in which I never exist. And while there is some plausibility to the claim that it can be better for a person to exist than not to exist,<sup>10</sup> there is scant plausibility to the reverse claim—namely, that it can be better for a person never to exist than to exist. While the disappointment the Pythagoreans felt upon learning the truth about the square root of two may well have been rational, the only lesson we can draw from this is that we can rationally regret something that did not make us worse off.

---

<sup>9</sup> Ibid., p. 175.

<sup>10</sup> See, for example Arrhenius and Rabinowicz (2010).

Let us therefore grant to Nagel that *if* I could not have been born substantially earlier, then being born when in fact I was could not be bad for me. Several authors have taken issue with the other premise in his argument—namely, that one could not have come into existence earlier than when she in fact did. Since Nagel seems to allow that a slightly earlier birth would not have prevented one from coming into existence, and since it is anyway not plausible that an individual comes into existence precisely at birth, presumably he intended birth to function as a proxy for whatever event does mark the start of one’s existence—for example, conception. Nagel appears to have thought that whatever that event is, had it occurred earlier, I would never have existed.

Unfortunately, Nagel does not say why he thinks it would have been impossible for us to come into existence earlier than we did. A natural interpretation of his argument, however, is that it appeals to the standard interpretation of Saul Kripke’s doctrine of the “necessity of origin”, whereby one could only have developed from the substance from which one in fact developed, for instance a particular sperm and egg.<sup>11</sup> Nagel’s discussants have assumed that this is the thought he was invoking in claiming that an earlier

---

<sup>11</sup> See Kripke (1980); Forbes (1985; 1986); Noonan (1983).

birth is impossible.<sup>12</sup> But as these discussants have also pointed out, it seems false at least of some people that they could not have existed earlier on this condition: we might imagine that the sperm and egg that fused to bring about a person's conception had been frozen many years earlier and then thawed.<sup>13</sup> McMahan suggests that maybe even those individuals who were conceived in more traditional ways could in fact have begun to exist earlier than they did.<sup>14</sup>

But some of Kripke's comments also indicate a different, less widely discussed rationale for the necessity of origin, which I think may have been what Nagel really had in mind. Kripke's other argument for the necessity of origin concerns the conditions for transworld identity. According to a standard approach, to find the possible world relevant for analysing a counterfactual statement we hold fixed the history of the actual world up until the point at which the counterfactual event would have occurred, and then insert that event and ask how things would have played out. On the "no-branching" criterion for transworld identity, some object or individual in the actual world is identical to an object or individual in the relevant possible world only if it

---

<sup>12</sup> Kamm (1993), p.26; McMahan (2005a); Kagan (2012) pp. 228–9; Johansson (2013); Brueckner and Fischer (1986), p. 220n6.

<sup>13</sup> Kamm (1993); McMahan (2005); Kagan (2012) pp. 228–9; Johansson (2013).

<sup>14</sup> McMahan (2005a) pp. 215–16.

exists in both worlds prior to the time at which they diverge—namely, prior to the time at which the relevant counterfactual event would have happened.<sup>15</sup>

This deeper metaphysical thesis might have been what Nagel had in mind when he said that different possible lives of a single person cannot converge from different beginnings. If the no-branching criterion for transworld identity is correct, then an individual could not have begun to exist earlier than when she did in the actual world, because this would require her to exist in both the actual and possible world *prior* to the specified event. This premise implies that it is metaphysically impossible for an individual to have begun to exist earlier than she did, even if, for instance, the gametes from which she was eventually formed already existed then.<sup>16</sup>

For all that I have said, though, Nagel’s “impossibility” justification for the Asymmetry is unsatisfactory. Even if it is true that I could not have come into existence earlier than when I did, there are simple analogues of the Asymmetry that cannot be defended on that ground. Suppose you learn three surprising things about your life. First, immediately after you came into

---

<sup>15</sup> Kripke (1980), pp. 113–114. More explicit supporters of the no-branching criterion include J. L. Mackie (1974); Penelope Mackie (2006); and George Sher (1979; 2005).

<sup>16</sup> Kripke (1980), pp. 112–114.

existence you were placed into suspended animation and revived ten years later. Second, at some point in the future, you will be placed into suspended animation for ten years, after which you will die without being revived, but had you not been placed into it, you would have lived those ten years normally and happily. And third, your date of death would have been the same whether or not you had passed the first ten years of your existence in suspended animation. If you had not been placed into it, then you would have had an extra ten conscious years in the past, and your future thereafter would have been no different.

If you are anything like me, you would feel that the badness of the deprivation of conscious life that results from the suspended animation you underwent during the first decade of your life is dwarfed by comparison with the deprivation that will result from passing your final ten years in it. What is more, in the case we have been considering I would have the selfsame asymmetrical attitudes towards the deprivation of future and past conscious life that I have in the ordinary case, in which the deprivation, if there is one, is due to nonexistence.

For convenience, I shall continue to compare deprivations of good life due to early death with those due to coming into existence later than one

could have, on the assumption that the latter is indeed metaphysically possible. But even if it is not possible, the analogue sketched above raises all the same puzzles. For that reason, the Asymmetry cannot be defended along the lines Nagel proposes.

## 1.2. The “biased toward the future” answer

While Nagel maintains that the impossibility of being born earlier *justifies* having asymmetrical attitudes towards birth and death, he is careful to deny that this provides the *explanation* for why we have the relevant attitudes. That explanation, he surmises, requires “a general treatment of the difference between past and future in our attitudes toward our own lives.”<sup>17</sup> Nagel is hardly alone in thinking that the explanation for Lucretius’ observation is just a particular case of a more general asymmetry between our attitudes toward past and future goods. The claim has been repeated by many authors, among them Parfit, Kagan, McMahan, Meghan Sullivan, and Samuel Scheffler.<sup>18</sup>

---

<sup>17</sup> Nagel (1970), p. 80n3.

<sup>18</sup> Parfit (1984); Kagan (2012); McMahan (2005a); Sullivan (2018) chapter 7; Scheffler (2018) chapter 4.

Call a person “biased toward the future” or “future-biased” if for a given good she prefers it to be in the future rather than the past or for a given bad she prefers it to be in the past rather than the future, and does so merely on the basis of its temporal location relative to the present.<sup>19</sup> (I use “bias” in a non-normative sense.) In a widely discussed article, Anthony Brueckner and John Martin Fischer claim that the proposition that future-bias is the source of the relevant attitudes is alone sufficient to justify the Asymmetry. As they write:

If death occurs in the future, then it is a deprivation of something to which we look forward and about which we care—future experienced goods. But prenatal nonexistence is a deprivation of past experienced goods, goods to which we are indifferent. Death deprives us of something we care about, whereas prenatal nonexistence deprives us of something to which we are indifferent.<sup>20</sup>

In short, they claim that because it is bad for us to be deprived of something we care about, but not of something about which we are indifferent, the Asymmetry is rationally defensible.

---

<sup>19</sup> Parfit (1984), p. 165.

<sup>20</sup> Brueckner and Fischer (1986), p. 219.



Brueckner and Fischer's argument has generated a good deal of discussion, much of which has centred on the question of whether desiring something is necessary for it to be in one's interest to have it. Thus, Jens Johansson replies that things can be good for us as long as we *would* have wanted them had we had them,<sup>21</sup> and Fred Feldman argues that pleasures can be good for even those individuals who lack the capacity to desire them.<sup>22</sup>

But we might instead wonder whether desiring something is sufficient for it to be in one's interest to have it. Brueckner and Fischer argue *from* the thought that we are future-biased *to* the conclusion that our attitudes toward birth and death are justified. On their account, our more general pattern of desires regarding the past and the future simply comes pre-justified, on the grounds that we have a reason to want something whenever we actually want it. But that is a controversial assumption.

So-called "Humeans about reasons" believe that I have a reason to want something when and only when it would fulfil, or help to fulfil, some desire that I have.<sup>23</sup> And among those who are not Humeans about reasons, many are preference-satisfactionists about wellbeing, some of whom hold that

---

<sup>21</sup> Johansson (2013; 2014).

<sup>22</sup> Feldman (2013).

<sup>23</sup> The most famous contemporary defence of this position, which is also called "reasons internalism", is Williams (1979).

something is in one's self-interest to have if and only if it would satisfy, or help to satisfy, at least one of her current, actual preferences.<sup>24</sup> But these philosophers would not think that there is an independent puzzle about whether the attitudes in question are rational. They may be curious *why* we have these attitudes, and one plausible hypothesis for that psychological question is that we have them because in general we are biased to prefer good experiences be in the future rather than the past. But those who believe that something is in our interest whenever we desire it should not have doubted the Asymmetry in the first place.

For those who, like me, are neither Humeans about reasons nor actual-preference-satisfactionists about wellbeing, it is an open question whether we have a stronger reason to want to extend our lives by postponing death than we have to extend them by having begun to exist earlier. And it is not clear how Brueckner and Fischer have provided any answer to that question.

As I say, those who have discussed the Asymmetry generally all agree that our preference for a later death over an earlier birth is *explained* by the

---

<sup>24</sup> Preference-satisfactionists of other stripes, such as those who think that a satisfied preference contributes to my wellbeing only if it is informed or ideally rational, should still find Brueckner and Fischer's argument unsatisfying. For it remains an open question whether we would prefer a later death to an earlier birth with full information and after due deliberation.

fact that we are biased toward the future. It is instructive to ask why, in turn, we are thought to be biased toward the future in the first place.

Some claim that the explanation is our belief that the future is open or undetermined in a way that the past is not, or, relatedly, that we can affect the future but not the past. But these explanations are not persuasive. As Parfit pointed out, we would not normally care either more or less about some experience in the future—some pain, for example—merely when and because we could be sure that we could do nothing to prevent it occurring.<sup>25</sup>

The much more plausible and widely endorsed explanation is that we exhibit the bias toward the future because of a deep-rooted conviction in *the passage of time*. We all feel, intuitively, as though time moves forward—what is present will become past, what is future will become present, and what is already past will retreat further away from the present as time flows onward. This conviction seems to provide a natural explanation of the bias toward the future. In short: we care about experiences precisely because of what it's like to undergo them in the present, and future experiences, but not past ones, will one day be present. As Kamm puts it,

---

<sup>25</sup> Parfit (1984), pp. 168–70. Kamm (1993, p. 28) also objects on this basis to the explanations in question.

Presumably the preference for future over past experienced goods is related to the fact that a person in the present, though not now enjoying either past or future goods, is yet to experience the future because of time's passage.<sup>26</sup>

This provides a plausible explanation for why I would care much less about having an earlier birth than about postponing death. Because I instinctively believe that time passes, I feel that by delaying my death I will secure additional good life that I will one day get to experience in the present. But by the same token, I also feel that whatever good life I might have realised by having begun to exist earlier than I did could never again be experienced.

The alleged connection between future-bias and our belief in the passage of time has led some philosophers to infer that whether future-bias is justified will ultimately depend on truths about the metaphysics of time. Thus Parfit concluded that “[i]f time's passage is an illusion, temporal neutrality cannot be irrational”—by which he meant that if it is an illusion then we cannot have more reason to care, for its own sake, whether an experience is in the past or the future.<sup>27</sup> Other authors have more explicitly argued that

---

<sup>26</sup> Kamm (1993), p. 28. See also Parfit (1984), p. 178.

<sup>27</sup> Parfit (1984), p. 181.

if time's passage is real then the Asymmetry is true.<sup>28</sup> Parfit himself expressed some scepticism whether even the reality of time's passage would be sufficient to justify the bias toward the future, but ultimately he left that possibility open. He *was* clear, though, in his belief that if time's passage is not real then future-bias in general, and the Asymmetry in particular, are indefensible.<sup>29</sup>

Unfortunately, as far as I can tell Parfit never spelled out in detail his argument to that effect. But Meghan Sullivan has suggested an argument which corresponds well to some of Parfit's own remarks and may have been what he had in mind.<sup>30</sup> Sullivan's argument against the bias toward the future exploits an analogy between time and space. If we believe that there are objective, mind-independent facts about what times are past, present, and future, we accept what is known as the "A-Theory" of time. If we deny this, we accept what is called the "B-Theory". According to the B-Theory, the temporal location of an event in time is analogous to its location in space. Just as there is no non-indexical fact of the matter about what is "here",

---

<sup>28</sup> Deng (2015); Robson (2014).

<sup>29</sup> Parfit (1984), pp. 175, 181.

<sup>30</sup> Parfit (1984), p. 178; Sullivan (2018), pp. 109–110. Note that Sullivan stops short of endorsing the argument, for she does not want her arguments against future-bias to depend on the metaphysics of time and does not herself affirm the B-Theory.

there is, according to the B-Theorist, no non-indexical fact about what is “now”. Just as “London is here” is true if and only if uttered in London, “2019 is present” is true if and only if uttered in 2019. Only if the A-Theory is true can time’s passage be mind-independent and real. The argument against future-bias from the B-Theory of time seems to be this:<sup>31</sup>

(1) If the B-Theory is true, the temporal location of an event relative to the present is like the spatial location of an event relative to what location is here.

(2) The spatial location of an event is arbitrary from the rational point of view.

(3) Hence, the temporal location of an event is also arbitrary from the rational point of view.

Premise (1) follows by definition. In support of (2), let us grant for the sake of argument that it is in itself bad for me when my partner is unfaithful, even if I never learn about this fact. Surely, I should not care, for its own sake, whether my partner was unfaithful to me here or elsewhere. According to the argument at hand, then, if the B-Theory is true it would also be no

---

<sup>31</sup> Sullivan (2018), pp. 109–110.

worse for me if my partner is unfaithful in the past rather than the future, all else equal.

We have already seen, though, that the temporal location of non-experiential goods does not matter much to us. Yet when it comes to *experiential* goods and bads, the argument at hand is not as straightforward as it has been made out to be. The reason is that experiences are essentially spatially indexed in a way that other events are not. It is not true that an experience I am having could be anywhere but here. And if, counterpossibly, an experience I am now having *could* be elsewhere, then it is unclear that its location *would* be arbitrary from a rational point of view. So it is not clear why we should accept (2) when it comes to the very sort of goods to which the bias toward the future is meant to apply.

Someone who wants to reject the bias toward the future on the basis of the metaphysics of time should not, I think, appeal to the metaphysical similarity of time and space to defend that premise. But it seems to me that a better argument is available. If the B-Theory is true, there is a sense in which we should view all times as equally present, or, alternatively none of them as present. Although only one set of my experiences is taking place now (to wit, taking place at the time this sentence is being read), all of my experiences

are still *happening* in the tenseless sense of that verb. Perhaps in that tenseless sense of happening, we should not care, for its own sake, whether these experiences are happening last year or happening next week.

If successful, this argument would provide a powerful case against the Asymmetry. Many physicists and philosophers accept the B-Theory of time, and that theory suggests that one does not have an independent reason to care, for its own sake, whether some experience is in the past or the future. On the assumption that our attitudes toward earlier birth and later death really are, as many authors have claimed, the products of the more general bias toward the future, then this is itself a *prima facie* case against the Asymmetry. For if we come to believe that a normative attitude we hold is based on a metaphysical claim that is false, we should take this to constitute a presumption against the attitude, since any independent support for it would obtain only by coincidence.

### **1.3. We are not biased toward the future**

But the bias toward the future is not the source of these attitudes, for we do not, in fact, care more about deprivations of life on the basis of their location



in time. Furthermore, once we get clear on what our attitudes toward various deprivations of goods actually involve—a bias toward what I'll call the “subjective future”—we shall see both that their source is unrelated to any independent conviction that time really passes and that the metaphysics of time can do nothing whatsoever to undermine them.

To begin with, suppose that tomorrow I will either die or else travel back in time to the year 1350, in which case I will live out the remainder of my days happily in the fourteenth century. I would greatly hope to travel back in time and, by comparison with that prospect, would regard dying instead to be a grave misfortune. More importantly, however, I would regard living out the rest of my life in the fourteenth century to be *no worse* a prospect than living the same length and quality of life as usual, here in the twenty-first century. Of course, I have contingent reasons to want to stay in my own epoch. Perhaps I am eagerly anticipating the 2024 Olympics; perhaps I would be sad to part with my friends and family (though we might, without issue, just suppose I can take them with me). But when I factor out these extrinsic considerations, living out the rest of my days in an age long before my birth seems about as good as living them here in the present day.

One thing this case seems to show is that, contrary to what Lucretius and his discussants have taken for granted, far from being indifferent toward periods of nonexistence that precede our birth, we sometimes regard them with the same concern we have toward the nonexistence that will follow our death. Much more importantly, however, it seems to be a case in which, when it comes to securing additional good life, the strength of my preference for it is unaffected by its temporal location relative to the present.

Before going any further let me pause to forestall two objections. First, it may be said that our intuitions about far-fetched cases involving backward time travel can tell us nothing about how we ought to apportion our prudential concern in the real world. But at this stage I am not concerned to answer that normative question. My intention is just to bring out our actual attitudes toward deprivations of life in the past and the future. When we imagine this case, and the others to come, we have strong reactions about them, so even if physical or philosophical investigation revealed backward time travel to be impossible, this would do nothing to undermine this observation about the structure of our actual attitudes.<sup>32</sup>

---

<sup>32</sup> Parfit (1984, p. 199) makes a similar point about the conceivability of his famous teletransportation cases. See also Unger (1990), pp. 84–85.

Second, it may be said that when we imagine travelling back in time in this case, and respond to it as we do, we are implicitly conceiving the fourteenth century to be in the metaphysical future. If that were true, then the case would do nothing to show that we are not future-biased, since the strength of the desire to go on living in the fourteenth century would be based on our impression that we would thereby go on living as usual in the future.<sup>33</sup> But I think that when we imagine this case we *do* fully grasp that to live on in the fourteenth century is to live on in what is now the past. To see this, we need only pick some event that is scheduled to occur later than the date of my time travel; say, the return in 2061 of Halley's Comet. If, as this objection holds, I were really implicitly imagining the fourteenth century to be in the future when considering the case at hand, then, upon believing that I was soon going to travel back in time, I would conceive of the 2061 return of Halley's comet as though it would happen 700 years from now. But I do not think of it in that way. Rather, I understand both that forty-odd years in the future Halley's Comet will breach the inner solar system and that by that time I will have long since travelled back to the fourteenth century. That

---

<sup>33</sup> Of course, I do not deny that there is a way in which it would seem to me *as though* the fourteenth century is in the future. One day I would be in the twenty-first century and the next I would be in the fourteenth century; absent external markers, my *perception* of time's passage is the same whether or not I travel back in time.

seems to answer the objection that in this case I am implicitly conceiving the fourteenth century to be in the future. In any event, whatever one thinks of this reply, the next case I offer is not vulnerable to the objection.

Now suppose that I have successfully travelled back in time. I have settled into my new life here in the fourteenth century. One day I awaken in the sick-house, having just recovered from a brief bout of the Plague. I find myself unsure whether I am forty or fifty years old, because I cannot remember whether the year of my birth is 1986 or 1976. I would, I submit, have no more concern for whether my date of birth is in 1976 or 1986 than I ordinarily have. And yet, as I ponder this question from my sickbed, I would understand that my birth is in the future. I would gaze forward to a time, long after my death, and understand that after all the intervening historical events, I will be born. But the fact that I conceive my birth to be in the future does nothing whatsoever to make me care more about securing additional good life by being born earlier. My desire for an earlier birth is unchanged with respect to the ordinary case.

Let me take stock. We have, first, a case in which I regard delaying my death no differently whether the additional life I would thereby secure would be in the future or the past; and, second, a case in which I regard having an

earlier birth no differently whether the additional life it would secure is in the past or the future. And in neither case do the relative intensities of my attitudes differ from those in the cases in which my birth and death are in the customary temporal directions from the present.

These cases show that our asymmetrical attitudes toward birth and death do not betray a bias toward the future. Moreover, they show those attitudes to be independent of any implicit conviction in the objectivity of time's passage. Consider my first time travel case. Relative to the present, the good life I stand to gain by travelling back to the fourteenth century is in the past, whereas the good life which I stand to gain by living on in the usual way would be in the future. But I regard the deprivation of one to be no harsher than the deprivation of the other. If my attitudes about the temporal locations of experiential goods were really based on the belief that only goods in the future will one day be present, then I would find living on by traveling back in time markedly less desirable than living on in the usual fashion. But I do not. My preference for later death over earlier birth has nothing to do with whether the life I stand to gain is earlier or later than the present moment. The assumption that the source of these attitudes is a bias toward the future is a red herring. Far from the metaphysics of time determining wheth-

er these attitudes are rational or not, neither our belief in the passage of time nor the truth of that belief has anything to do with them.

Let me conclude this section by noting that there are straightforward parallel arguments that the bias toward the future is not the explanation for our attitudes toward good and bad experiences which do not affect our lifespans. Here is an example. Suppose that at midnight I might or might not travel back in time. I would care no less, for its own sake, about a pain or a pleasure that occurs a week after I arrive in the past than I would care about one that would occur in a week's time here in the twenty-first century, should it turn out that I will not travel back.

#### **1.4. Our actual attitudes**

I shall take myself to have established that our attitudes toward deprivations of life at different times are not attitudes about the temporal location of events relative to the present. We can now ask what our attitudes toward those deprivations actually are. As a natural first step, borrowing some terminology from David Lewis, we can distinguish external time from

personal time.<sup>34</sup> When tomorrow I will travel back to 1350, the year 1375 is in my future in personal time: between now and then my beard will continue to grow, my wristwatch will count the passing hours as usual, and so forth. But in external time, 1375 is in the past: the fourteenth century is earlier than the present day.

A natural first try, then, is to say that what matters to me is not the procurement of additional good life in the external future but rather additional good life in my personal future. But a moment's reflection shows this will not quite do. For in the second case I considered—in which I have already settled in the fourteenth century—my birth *is* then in my personal future—it's just that it occurs at a time long after my death.<sup>35</sup> And yet, as we saw, in the fourteenth century I do not care much at all about how early or late my date of birth is. When it comes to extending our lives, it seems that our preference is for that good life which is not only in our personal futures but also before our deaths (or after our counterfactual deaths, depending on the question we are asking). Let us call that portion of a person's life her *subjective future*. When it comes to extending my life, I am, it seems, biased toward the subjective future.

---

<sup>34</sup> Lewis (1976a), pp. 147–48.

<sup>35</sup> It is in what Lewis called my *extended* personal future. Lewis (1976a), p. 147.

We can next distinguish a person from that person's life, where a life is not identical to a person, as these terms are sometimes used, but is rather what she lives—it is the sequence of events that happen to her.<sup>36</sup> What defines that part of a person's life that is in her subjective future? It is, I submit, that part of her life which will be lived by those temporal stages of herself which are causally downstream from her present self, connected to it by unbroken chains of causal dependence. Certain stages of oneself are causally dependent on oneself now, in the sense that they arise out of or continue the current stage. What properties are involved in that dependence will simply be those specified by the correct theory of personal identity over time; perhaps the properties are mental or physical or both.<sup>37</sup> And what counts as the right kind of dependence is just whatever counts in other cases of personal identity (I shall return to this question in more detail later).

We can now clarify how, in the fourteenth century, I do not care much whether my date of birth is in 1976 rather than 1986. Although by being born in 1976 I would secure an extra decade of good life, and although that

---

<sup>36</sup> As far as I know Kagan (1993) was the first to explicitly identify this difference.

<sup>37</sup> For more developed discussions of the nature of this causal dependence see Lewis (1976b); Nozick (1981); Shoemaker (1984); Parfit (1984); Sidelle (2011).



life would be in the future, that segment of my life will not be lived by a temporal stage of myself that is causally downstream from me now.

Philosophers who have discussed the bias toward the future have, understandably, been tempted to describe it as the special concern one has for one's "future self". But, as I hope is now clear, characterising the pattern of concern using temporal concepts is problematic. Let us therefore call those stages of me that are in my subjective future my *successive selves*, and say that all of them together constitute my *successive self*.<sup>38</sup> My special concern with what happens *in* the part of my life that constitutes my subjective future is really a special concern for what happens *to* my successive self. (Of course, in ordinary cases, that concern coincides with my concern for what happens in the external future. It is in cases involving backward time travel that my concern for my successive self and my concern about what happens in the future diverge.)

Having established that our attitudes toward birth and death track not a bias toward the future but rather a special concern for our successive selves, we can now ask whether *that* attitude is justified. To that end, let us leave time travel behind, and confine our discussion to cases in which one's

---

<sup>38</sup> I borrow these phrases from Parfit (1984, p. 301), who uses them interchangeably with "future selves" and "future self".

subjective future coincides with the external future. In what follows I want to suggest that the desire for a given good experience to be in our subjective future is justified on the same basis as the desire to want good experiences at all.

### **1.5. Personal identity and prudential concern:**

#### **another asymmetry**

The last half-century has witnessed a profusion of interest in the question of the grounds of rational prudential concern for what happens at other times—what is sometimes styled “what matters in survival”.<sup>39</sup> This question concerns when and why I should now regard something that will happen to an individual at another time with that same special sort of egoistic solicitude with which I would ordinarily regard it were it happening to me, as opposed to how I would ordinarily regard it were it happening to somebody else. For instance, Parfit famously argued that personal identity itself is not among the bases of rational prudential concern, and accordingly that a person should

---

<sup>39</sup> Parfit (1984), p. 298 and elsewhere.

not be rationally concerned with the end of her existence so long as certain other relations will continue to hold between her and someone else.<sup>40</sup>

It is a striking fact about all discussions of the bases of rational prudential concern of which I am aware that they consider only cases that involve one's relationship to future individuals. Thus Bernard Williams titled his early and influential essay on the topic "The Self and the Future", and in it considers cases in which one will undergo certain drastic, potentially identity-altering changes, and asks the reader to consider whether he or she would regard a subsequent episode of torture with same the dread she would have toward it were the changes not to take place.<sup>41</sup> Peter Unger generalises and defends Williams's methodology, exploiting at length what he calls the "sacrifice for future wellbeing test"; in these cases, one can sacrifice some amount of present wellbeing for the sake of the wellbeing of some future individual who may or may not be oneself.<sup>42</sup> Many others have employed the sacrifice for future well-being test either explicitly or implicitly, and most have also made use of Parfit's well-known examples of brain division and teletransport-

---

<sup>40</sup> Parfit (2007) later admitted to having conflated the so-called "desirability" and "prudential" questions of what matters in survival. I am concerned here with the (more general) prudential question, which is the one I describe. For a useful analysis of the demarcation see also Unger (1990), pp. 93–97.

<sup>41</sup> Williams (1970).

<sup>42</sup> Unger (1990).

ation, all in an attempt to uncover the true bases of rational prudential concern.<sup>43</sup>

But Williams never considers whether, if the torture had already occurred and was followed by the drastic changes to oneself, one would regard it to be as bad as a past episode of torture after which no such changes had occurred. Nowhere in his otherwise impressively comprehensive book does Unger consider a sacrifice for *past* wellbeing test. Parfit does not ask whether having half the brain of a previous individual, or being recreated according to the scan of the brain and body of someone else, is as good as *having survived*.<sup>44</sup> To the best of my knowledge, the other authors who make use of these and other cases never consider the conditions under which rational prudential concern is preserved between oneself and an individual in the past.

And yet, in reading all these authors, one might be forgiven for supposing that they have in mind that the conditions they entertain or defend would apply between oneself and individuals that exist at *any* time, not just times in the future. By fixating on cases that involve one's relationship to future individuals, those who aim to discover the grounds of rational prudential

---

<sup>43</sup> For instance, Holtug (2010); Noonan (2019); McMahan (2002); Otsuka (2017).

<sup>44</sup> Interestingly, Parfit does consider these cases in an argument that reductionism about personal identity implies that no one can deserve to be punished for past wrongs (1986a, pp. 838–840), but never extends the discussion to prudential concern for past individuals.

concern have thus neglected to consider whether the answers they entertain can be generalised to give us the grounds of rational prudential concern for individuals at any time.

Now, when it comes to the question of personal *identity* across time, the assumption that the conditions for identity to a future individual will generalise to the conditions for identity to an individual at any time is warranted. That is because the identity relation is symmetric. A relation is symmetric when its converse is the same relation as itself. (The converse of a binary relation is the unique relation that obtains when its elements are reversed—the converse of “child of”, for instance, is “parent of”.) “Sibling of” is an example of a symmetric relation. And because identity is symmetric (if  $A=B$  then, necessarily,  $B=A$ ), the converse of any relation that is necessary or sufficient for identity to hold between oneself and an individual in the future will also be necessary or sufficient, respectively, for identity to hold between oneself and an individual in the past.<sup>45</sup>

Note that this is true even if the relation that *underwrites* identity is not itself symmetric. For, while some relations sometimes thought to play a role in personal identity are themselves symmetric—spatio-temporal continu-

---

<sup>45</sup> I here take for granted that we are reductionists about personal identity, but this assumption is not necessary for my eventual main argument, which concerns not the grounds of identity but those of rational prudential concern.

ity, for example—others are asymmetric, like those involved in anticipation, intention, and memory. The converse of “remembers having the experiences of”, for instance, is “is currently having experiences that will be remembered by”. Still, because identity is symmetric, if it turns out that Y is the same person as X whenever Y remembers having the experiences of X, then X must be the same person as Y whenever X is currently having experiences that will be remembered by Y.

Since personal identity is symmetric, it is natural to assume that the prudential concern relation is, too. After all, it is generally unquestioned that in conventional cases in which identity across time is preserved, what matters in survival is also preserved.<sup>46</sup> But I will assume, for well-known reasons, that personal identity itself is not necessary for rational prudential concern.<sup>47</sup> And if that is right, then it is an open question whether the *prudential concern* relation is also symmetric. Suppose, for the sake of illustration, that it is true that at  $t_1$  (which is now) I should be prudentially concerned with what happens to me at  $t_2$  because, at  $t_2$ , I will remember having the experiences I am

---

<sup>46</sup> The exception being cases involving infancy and early childhood, of which it is not uncommon to deny that the grounds of rational prudential concern track identity. See, for instance, McMahan (2002).

<sup>47</sup> In particular, I have in mind cases involving the fission of one individual into multiple individuals.

now having. It remains an open question whether at  $t_2$  I should be concerned with what happened to me at  $t_1$ .

Why has the possibility that the prudential concern relation is asymmetric been largely overlooked? I have already suggested one possible explanation: it is typically taken for granted that throughout most parts of most lives rational prudential concern and personal identity coincide, and the identity relation is symmetric. But the primary explanation, I conjecture, is just that discussions of the grounds of rational prudential concern have confined themselves to what matters in *survival*—to cases involving one's relationship to individuals, or to temporal stages of individuals, that exist at times later than the present moment. Those who restrict their investigations in this way will never confront the question of whether or not the prudential concern relation is symmetric.

### **1.6. In defence of the Asymmetry**

It is my contention that the grounds of rational prudential concern for what happens at other times are indeed asymmetric. In particular, I claim, they hold only between oneself and individuals, or the temporal stages of

individuals, that are causally downstream from oneself now. We can put this claim in terms of one's time-relative interest in having some good. Relative to a given time, I will argue, a person has an interest in having some good only if she is having it now or it will be had by her successive self. To prevent misunderstanding, I do accept that there are some prudential reasons to be concerned with what happened to oneself at times that are not in one's subjective future, and I shall have more to say about what those reasons are later. I maintain, however, that even those reasons are merely derivative: they obtain only insofar as they improve the condition of oneself now or of one's successive self.

We can begin our investigation with the modest observation that almost every contemporary account of the conditions for rational prudential concern across time includes among those conditions some kind of causal continuity. Accounts vary widely with respect to both the relevant states of individuals that are to *be* continuous if the prudential concern relation is to hold, as well as with respect to the scope of causal dependence that is sufficient for that continuity. Parfit's position, for example, seems to have been that the relevant states are psychological ones, and that almost any kind of causal dependence will do. Other philosophers place physical states among the relevant



ones, and yet others agree with Parfit that the only relevant states are psychological but think the causal dependence must be narrower for the grounds of prudential concern to be preserved, perhaps requiring the continuity of parts of the brain.

What all these accounts have in common, however, is that they seem to implicitly assume what we might call the “Two-Directional View”:

*Two-Directional View:* The causal condition on rational prudential concern is symmetric. Provided the right states obtain in X and Y, and provided Y’s states depend causally on X’s in the right way such that X should be prudentially concerned for Y, then Y should be prudentially concerned for X.

In contrast, I defend the “One-Directional View”:

*One-Directional View:* The causal continuity condition on rational prudential concern is asymmetric. X should be prudentially concerned for Y if and only if the right states obtain in X and Y and Y’s states depend causally on X’s.<sup>48</sup>

---

<sup>48</sup> It is worth noting that in his discussion of temporal biases, Parfit rejects the claim that the “direction of causation” justifies the bias toward the future (1984, p. 168). But by the direction of causation he means only the proposal that the bias could be justified on the grounds that it matters whether or not I can *affect* the future. His arguments there do not concern, and go no way toward undermining, my own thesis that the direction of causation is among the grounds of rational concern themselves.

At the risk of stating the obvious, let me be explicit about how this all relates to the Asymmetry. On the Two-Directional View, the prudential concern relation runs in either causal direction. If that view is true, then the Asymmetry is *prima facie* unjustified. If it is to be justified, it must be so on the basis of some independent premise that supplements the rational prudential concern relation. On the One-Directional View, the Asymmetry is *prima facie* justified. What once appeared to be a bias toward the future, and which we now know is actually a bias toward one's successive self, is built into the prudential concern relation itself.

In weighing up these two views, it is helpful to begin by asking why causal continuity has been thought indispensable in the first place. The reason is that its inclusion is necessary to avoid the counterintuitive implication that rational prudential concern is preserved in cases of “random duplication”.<sup>49</sup> These are cases in which, at some time after your death, a molecule-to-molecule duplicate of you just happens to pop into existence. This duplicate will not be created by scanning your body and reproducing it from

---

<sup>49</sup> See for example Nozick (1970), pp. 41–2; Shoemaker (1984), pp. 115–120; Unger (1990), p. 67. Kolak and Martin (1987) and Elliot (1991) maintain that these cases preserve identity; they are also the only authors I know of who have attempted to reject the causal condition wholesale.

new matter according to the scan—there is no counterfactual dependence at all. It is simply a random quirk of chance that happens to occur.

Many of us would regard the prospect of random duplication to be a very bad one. Suppose you know that you will be vaporised tonight in your sleep and a random duplicate of you will later pop into existence, replacing you in your bed. If you are anything like me, this prospect fills you with horror. True, because presumably your duplicate will carry out the life you otherwise expected to live, this may not be as bad as vaporisation *without* replacement. For instance, some desires you now have will go on to be fulfilled, desires which would not be fulfilled had you not been replaced by a duplicate. But, intuitively, vaporisation is nonetheless a profound misfortune. Even if it would not be as bad as ordinary death, it seems far closer to it than to ordinary survival.

Because this instance of vaporisation-replacement is in your subjective future, the One- and Two-Directional Views treat it the same. Both correctly imply that the prospect would be very bad for you, since all the good life you would otherwise have had in the future will now accrue to somebody who is causally discontinuous with you.

Suppose now that you learn you were mistaken: in fact, the vaporisation took place yesterday, and *you* are the random duplicate. If you are like me, this news does not disturb you in the same way. It simply does not feel to me like I have lost something of profound prudential value. There are grounds to regret being a duplicate and I shall shortly discuss some of them. For now, however, it suffices to note that learning that I am soon to be vaporised and randomly replaced is very different to learning that someone else was vaporised and randomly replaced by me.

Now, if the explanation of why in ordinary cases we care about having good experiences in the future rather than the past was that we believe in time's passage, a defender of the Two-Directional View could account for why we react differently to these two cases as follows. Upon learning that we are the duplicates, we are less concerned about the past good life we have lost because it seems to us that the experiences it contains are forever in the past. When we will be vaporised tonight, on the other hand, the prudential concern relation is severed between us and someone whose experiences will one day be *present*.

But as we have seen, *pace* Parfit and many others, our preference about when good experiences happen is independent of our belief that time passes

and cannot be explained by the conviction that, unlike future experiences, those in the past will never again take place.

The One-Directional View, on the other hand, implies that our reactions to the two duplication-replacement cases are what they should be. When I will be vaporised tonight, I stand to lose out on all the good life that would otherwise have accrued to stages of myself that are downstream from me now. In contrast, when I learn that I am the random duplicate of somebody who was vaporised yesterday, the good experiences I have lost are not something I had a prudential reason to want in the first place, because they would have belonged to stages of myself that are causally upstream.

At this point it may be objected that restricting the causal condition to a single direction is a move that is *ad hoc* or arbitrary. But its endorsement could be no *less* arbitrary than the endorsement of the already widely-assumed Two-Directional View, since the chief motivation for the inclusion of causal condition in the first place is to exclude future random duplication cases, and the One-Directional View handles those equally well.

Moreover, the fact that the One-Directional View can easily account for our intuitions about both duplication cases is already an additional reason to accept it. Lastly, for those of us who think coherence between a principle and

case-based judgments provides additional justification for each, the fact that the One-Directional View implies the bias toward the successive self already provides some support for both. Since at this point the Two-Directional View has no advantages of its own, that is already reason, I believe, to accept the One-Directional View—and with it the Asymmetry.

For all that, however, I accept that some theory better according with our intuitions than its alternatives is not by itself enough to warrant high confidence in its truth. For that, we need some intuitively compelling underlying rationale for why the view would actually matter. While the Two-Directional View faces the same challenge—showing *why* causal dependence in either direction matters in the first place—ideally we can find some independent, normatively satisfactory reason to think that not just casual dependence, but causal dependence in one direction only, is rationally significant. Is there any such rationale available?

I believe that there is. When what is now my future self is also my successive self, it is someone I will become; into whom I will grow and evolve.<sup>50</sup> Many of my present thoughts, desires, and intentions will develop into new thoughts and desires or give rise to actions. My successive self is someone

---

<sup>50</sup> Indeed, Michael Otsuka (2017) contends that prudential concern is rational only for those individuals whom one will “be or become”. Taken literally, the verb “become” applies only to the relationship between oneself now and one’s successive selves.

who will continue me. That these considerations matter has a ring of truth to it that its converse—that I have an egoistic interest in what happens to the individual whose life *I* have continued—lacks.

I will conclude my discussion by considering an objection. There may seem to be one respect in which our attitudes toward later death and early birth do not accord with the picture I have been defending. At the beginning of this chapter, I restricted the Asymmetry to experiential goods only, on the grounds that arguably it is in a person's interest to have *non*-experiential goods like accomplishments and successes at any moment in her life, not just in her subjective future. Does the One-Directional View fail on this count?

I suggest that it does not. The reason is that having had more non-experiential goods in one's past makes one better off in the present. Other things being equal, it is better for a person now to exist within a life that contains more and greater accomplishments. Think back to the case we considered at the beginning of this chapter, of which we asked whether it would be better to now be forty or fifty years old, given that one's date of death would be the same in either case. When I reasoned that if I am fifty rather than forty I am likely to have achieved more worthwhile things, it does not seem to me a stretch to say that I was reasoning that if I turn out to be fifty,

then I am better off *now*. Existing within a grander life is, on this picture, a higher-order good that one has an interest in having at any given time within the life, which in turn explains why we have a reason to want non-experiential goods that are not in our subjective futures. Of course, this way of looking at things assumes that a person can be made better off at a given time in virtue of things which have no intrinsic effect upon her at that time, but those who believe in non-experiential goods must already accept that possibility.

We can note, in closing, that this point also explains why I should not be wholly unconcerned, even on purely egoistic grounds, to learn that I am the random duplicate of someone vaporised yesterday. If I am a duplicate, then all the things I believe I have achieved are not, in fact, part of my life. For that reason, if I am a duplicate, I am worse off, now, than I would otherwise be.

## 1.7. Conclusion

Most of us wish we could live on at least for a while longer than we in fact will, yet care little that we were not born earlier. This pair of asymmetrical



attitudes is, I have argued, defensible. We are perfectly right to want to die later but not to care much about having been born earlier. I have tried to show that this pair of attitudes is not, as has been supposed, the product of a more general preference for experiential goods to be in the future rather than in the past. Instead, it tracks the preference that good things happen to a part of oneself that is downstream from the present.

Philosophers have assumed that if I have a non-instrumental reason to want some good to be located at one time rather than another, this must be because the event's temporal location is itself rationally significant. In other words, the established order of investigation has begun by taking for granted that I should be equally prudentially concerned with goods for their own sake in the absence of any specification of their temporal properties, and then asking whether there is something about a good state or event occurring in the past that makes it less worth wanting than it would be were it in the future.

I have argued that this picture gets things the wrong way around. There is nothing about the temporal location of a good experience relative to the present that makes it less valuable to have—rather, its causal relationship to the present already determines whether it is in one's interest to have it. Our typical interest in having good experiences in the future rather than the past

is justified internally, on the basis of the same considerations that ground our interest in having good experiences at all. Unlike the good experiences I will lose out upon by dying earlier than I otherwise would have, the good experiences I would have had by being born earlier than I was are not in my interest to have had.

## Chapter 2: Is Conscious Life Itself a Good?

In the previous chapter, I argued against one of the major challenges to the common view that death can be bad for its subject. I concluded that a person can have a prudential reason to go on living in virtue of the good things that would thereby accrue to someone who is causally dependent on her in the right sort of way.<sup>1</sup> Of course, the flip-side of that thought is that a person has a competing prudential reason *not* to go on living in virtue of the bad things that would happen to her successive self. This raises the obvious question of when a person has an all-things-considered prudential reason to go on living.

Here is a natural answer to that question: “You have a prudential reason to go on living just in case your future will contain more good than bad on balance.” One way this answer might be thought incomplete is the following. Some would say that even if my future will not be on-balance good when

---

<sup>1</sup> To avoid unnecessary distractions, I will take for granted that if a good is in the future it is also in one’s present subjective future.

taken on its own, if it would positively contribute to the value of my life as a whole then that would give me a prudential reason to live it. I shall not consider that reply further here. A more interesting way in which the natural answer might be incomplete is if the prudential value of existence has positive prudential value in its own right. Perhaps, as Samuel Johnson is said to have declared, by contrast with death “mere existence is so much better than nothing, that one would rather exist even in pain, than not exist.”<sup>2</sup>

It is doubtful, though, that Johnson would have thought it better to live on in a persistent coma than to cease to exist. On the plausible assumption that living on in a coma is no better for its subject than death, one way to rescue the claim that existence is worth wanting for its own sake is to maintain that being in a persistent coma is bad for a person to an extent that outweighs whatever positive value is afforded by existence itself. But no aspect of existence seems positive when it is spent in a total coma. A more charitable construal of Johnson’s assertion is that there is positive value not in existence *tout court* but in conscious existence.

This was roughly the picture that Thomas Nagel sketched, unfortunately without further argument, in his well-known study of death. He wrote:

---

<sup>2</sup> Boswell (1953/1799), p. 950.

There are elements which, if added to one's experience, make life better; there are other elements which, if added to one's experience, make life worse. But what remains when these are set aside is not merely neutral: it is emphatically positive. . . . The additional positive weight is supplied by experience itself, rather than by any of its consequences. . . . That is what is meant, I think, by the allegation that it is good simply to be alive, even if one is undergoing terrible experiences.<sup>3</sup>

One way to understand Nagel's idea is that every individual constituent of one's overall conscious state (for example, the sound of the clock tolling one) is in one way valuable for its subject, just in virtue of the fact that it is an experience. But there does not seem to be anything to be said for the pain one feels in one's foot, for example. A more plausible understanding of what he meant would distinguish the constituent elements of experience from the whole unified experience one has over a period of time. Although there is no way in which I benefit from experiencing a given pain, I am in one way made better off, relative to not being conscious, by the fact that I am experiencing anything at all. On this picture, if going on living with pain was the only way to go on living at all, one might have some reason to do so. But it is not the

---

<sup>3</sup> Nagel (1970) p. 74. (In the text, the last sentence of the quoted passage precedes the ones before it.)

pain that would generate that reason but the fact that one is experiencing at all, despite the pain, that does. I shall call the extended state of having experiences “conscious life”. The plausible claim that was probably intended by Johnson is that a period of conscious life has some positive value for its subject over and above the particular good and bad things that together constitute it.

The thought that conscious life is itself good for its subject has not received too much attention, but Joseph Raz has taken up the task of criticising it. Raz asks us to imagine a person whose waking life involves nothing but agony, terror, and racing paranoid thoughts.<sup>4</sup> He thinks we will agree that there would be *nothing* of prudential value in this person’s life. From this he infers that experience (by which I will assume he means a period of conscious life) is not worth wanting for its own sake. “It all depends”, he writes, “on its content”.<sup>5</sup>

That this person’s life seems to have nothing of value in it is, I think, questionable. It is not so implausible that this hapless person might, in her more lucid moments, find some solace in the fact that she is alive to experience anything at all. Even if Raz is right about the intuition his case evokes,

---

<sup>4</sup> Raz (2001), p. 117.

<sup>5</sup> Ibid.

there are two further problems with his argument. First, he will have shown at most that a period of conscious life is not in any way valuable when its only contents are very bad. It does not follow that a period of conscious life is not in any way valuable when at least some of its contents are either neutral or good, or even when the bad contents it has are only somewhat bad. The more pressing problem with his example, however, is that it may only appear that nothing speaks in favour of the person's wretched life because its many horrors eclipse the positive value derivable from conscious life alone.

To avoid the swamping problem that a case like Raz's faces, we should ask whether there seems to be something independently valuable about a period of conscious life that is not as bad as all that. Ideally, we could ask whether a person's conscious life has value for her when its contents are overall neutral (either because the contents of experience are all themselves neutral, or because the good and bad ones cancel each other out in the final account).

Let us try to do so. Imagine the Grim Reaper appears and informs you it's your time to die. Because he is in a good mood, however, he will offer you the following deal. Instead of immediate death, you may have another five years of conscious life. But the particular contents of your experience will be

neutral on balance, and the same will be true of any non-experiential goods. He clarifies: for each positive element of wellbeing, with the (possible) exception of the element of wellbeing that is having a conscious life itself, there will be a corresponding negative one, such that one would be indifferent between having both elements and having neither.

I believe that most people would readily accept the Grim Reaper's offer. This provides a *prima facie* case that conscious life itself is among the elements of wellbeing. In what follows, I will argue that this position is the correct one. Of course, if I am right, it follows that the future the Grim Reaper is offering would not be *overall* neutral. We might simply think of the position I will defend in the following way. There is some list of the things we ought to fill our lives with to make them go well. Prominent candidates for inclusion on that list are happiness, accomplishment, knowledge, personal relationships, goodwill towards others, and the fulfilment of desires. We should also place, on that list, the much more general good that is the having of experience itself, independent of its contents.

I will begin by giving some reasons for doubting a competing justification for accepting the Grim Reaper's offer: that we have a reason to go on living because by doing so we temporarily avoid the non-comparative bad



that is the permanent loss of conscious existence. I will then respond to several challenges to the claim that the Grim Reaper case shows that conscious life is in itself good for its subject. These challenges aim to show that the case does not reveal a genuine preference for prolonging our conscious life, either because we do not grasp what such a life would really be like or because we are subject to a bias that colours our judgment about it. After replying to these challenges, I will return to the thesis that conscious life is valuable in its own right. I will try to spell out that thought in more detail and offer a rationale for why it might be so.

### **2.1. Is the goodness of life the badness of death?**

Some events or states are bad for us in virtue of the alternatives they preclude. For example, it may be bad for a person to lose her car keys, not because losing them is itself bad for her but because it means she cannot attend the party she wanted to attend. Of course, preclusion cannot be the only way in which something can be good or bad for a person; the regress must end somewhere. We would then hold that misplacing her keys is comparatively bad for her in virtue of depriving her of something that is non-

comparatively good. Obviously enough, there are also comparative goods and non-comparative bads. Taking painkillers may be comparatively good for a person in virtue of preventing the non-comparative bad that is the pain she would have otherwise suffered.

Many philosophers believe that death—or the non-fatal permanent end to consciousness—can be comparatively bad for a person; Chapter 1 was an attempt to refute one prominent argument against that view. But some philosophers have suggested that death can also be bad in the non-comparative sense.<sup>6</sup> Some of those philosophers also think that the non-comparative badness of death gives us a reason to postpone it. If they are right, this might explain why it would be in one's interest to accept the Grim Reaper's offer.

One of the main reasons given by proponents of the view that death is non-comparatively bad cannot be used to show that one has a reason to accept the Grim Reaper's offer. It has been claimed that one way in which death is non-comparatively bad is that it involves the loss of the continuation of the good things in life we already have, and the loss of something good is in itself bad, irrespective of the comparative badness that comes from its sub-

---

<sup>6</sup> Kamm (1993; 2017); Kagan (2012); Benatar (2017).

sequent absence.<sup>7</sup> While this idea has some plausibility, it cannot by itself justify accepting the Grim Reaper's offer. Unless there is already something valuable in a period of conscious life over and above its particular contents, having a future that is on-balance neutral in those contents would involve no smaller a loss than death would.

There are, however, other properties of death that might make it non-comparatively bad and which may seem to justify accepting the offer. Frances Kamm—perhaps the most forceful contemporary defender of death's non-comparative badness—suggests that the permanent end of consciousness is non-comparatively bad in part because it means that “everything is all over for one”.<sup>8</sup> If everything being all over is in itself bad for me, then it might make sense to prolong my conscious life even if to do so would not be non-comparatively good for me.

Accordingly, it could be in a person's interest, Kamm thinks, for him to postpone his death by entering suspended animation for millions of years after which he will regain consciousness, even though doing so will not change the length or quality of his conscious life (Kamm dubs this person the

---

<sup>7</sup> Kamm calls this the “Insult Factor” (1993, p. 40; 2017, pp. 735–6); see also Kagan (2012), pp. 278–9 and Benatar (2017), pp. 102–110.

<sup>8</sup> Kamm (1993), p. 19.

“Limbo Man”).<sup>9</sup> By doing so, this person could ensure for as long as possible that his conscious life is not all over. She believes that the Limbo Man’s decision is rational and that this in turn provides evidence that the permanent end of consciousness can be non-comparatively bad.<sup>10</sup>

Not all philosophers share Kamm’s intuition about this case.<sup>11</sup> In any event, since I maintain both that reason we have for extending our lives is not the non-comparative badness of death and that there is no positive prudential value in unconscious life, I must deny that the Limbo Man has a reason to enter suspended animation. To that end, consider

*Local Time Freeze:* The human race can together enter a collective limbo by triggering a temporary “time freeze” that is localised only to our planet. If we do so, it will seem to us as though no time has elapsed from one moment to the next, but in external time thousands of years will have gone by.

I believe that there is no relevant sense in which triggering a local time freeze would be in anyone’s interest. If the decision were mine alone, I believe I would have no reason at all to trigger it. This seems to show that it is not

---

<sup>9</sup> Ibid., pp. 19, 49–50.

<sup>10</sup> Ibid.

<sup>11</sup> Thomas Scanlon does not, for one. (2015), p. 490.

the mere postponement of death that is in one's interest. An alternative explanation for the force of the original Limbo Man case, insofar as it has force, is that by entering suspended animation the Limbo Man's will get to experience a distant and exotic future.

Someone might disagree that a collective time freeze would be in no one's interest, or point to some feature of the limbo man case I have overlooked to show that it does not have this implication. Or else one might accept that one lacks a reason to enter a limbo state, but does have a reason to go on living provided the life one would have is conscious on the basis of some other feature of death that makes it non-comparatively bad. I shall, therefore, offer a more direct argument against the proposition that death can be non-comparatively bad, and afterward give a reason to think that, even if it was non-comparatively bad, this still would not, on its own, provide a reason to go on living.

If death really is a non-comparatively bad thing to happen to someone, we should in principle be able to ask how bad it is for her. In the first place, however, it is utterly unclear how to go about assigning to a given death a non-arbitrary value such that it is equivalent, for example, to some degree and duration of pain. But let us grant that we can assign it a value, if only

an imprecise one. It is still implausible that a person's death constitutes a non-comparative bad for her, for it is implausible that there is any significant amount of non-comparative good that it could outweigh. Suppose, for the sake of exposition, that death's non-comparative badness is such that to offset its presence in a person's life she would need to live for at least three happy years. Then a life of two happy years would not be worth living, a life of one year worse still, and so on. But I cannot believe that the life of a child who dies after two happy years was overall bad for her. If anything, most people would be glad for the child's own sake that she got to live at all.

Of course, perhaps death is just not as bad as all that. Let us suppose it is only bad enough to outweigh three happy months. Would anyone conclude, upon the discovery of an extra-terrestrial species whose members die painlessly after only one merry summer, that the total prudential goodness in the universe is lower than previously thought? Defenders of the position that death is non-comparatively bad for its subject might reply that the position applies to persons only, but things seem no different if we clarify that these extra-terrestrials come into existence as fully-formed persons. For my part, it is hard to believe that death could be non-comparatively bad enough to outweigh even one happy day.

Suppose that this argument does not succeed; death really is non-comparatively bad. For all that, it is still not obvious that this could provide a prudential reason to postpone death. Were immortality possible, then choosing to live forever might be in one way good for me, since only then could I escape the non-comparative badness of death. But I must die eventually, and when I do my death will be no less bad for having been rescheduled. For death's non-comparative badness to justify postponing death in the Grim Reaper scenario, or for that matter in the Limbo Man case, it seems that it must also be true that it is in a person's interest to undergo a bad event later rather than sooner. That is a questionable premise which to my knowledge has not been adequately defended.

It may be replied that one needn't appeal to the mere temporal distance of death to explain how its non-comparative badness provides a reason to postpone it. Instead, the following explanation might be proposed. Suppose we accept that identity is not what grounds rational prudential concern across time, and suppose we also accept that what *does* ground it comes in degrees (degrees of psychological connectedness and continuity between person-stages, for instance). It may then seem that I now have a present time-relative interest in postponing my death when and because this would mean

that the non-comparative badness it will bring is suffered by a stage of myself to whom I am more weakly connected in the relevant way, and hence to a stage for which I should now have less prudential concern. But this cannot be right. If identity is not what grounds rational prudential concern, then at a given time it seems that it would be equally bad to die as to come to lack the prudential concern relation to any successive individual.<sup>12</sup> And if the rational prudential concern relation comes in degrees, then to the extent that I will be less-than-fully connected in the relevant way to myself at the moment of my death, I will have already suffered the non-comparative badness of death gradually, by way of the weakening of the connections that determine how much prudential concern I should have about the bad things that happen to me at that time. On a reasonable interpretation of time-relative interest account, then, there is a sense in which a person dies only gradually, until at last she dies completely by way of her literal death.

---

<sup>12</sup> For good measure I repeat that I am assuming for simplicity that one's subjective future coincides with the external future.



## 2.2. Do we really want a life that is neutral in its contents?

Because I believe that the reason one has for continuing conscious life is not that by doing so she can avoid the non-comparative badness of death, I maintain that the better explanation for the Grim Reaper case is that conscious life is non-comparatively good. At this point, an opponent might object that the Grim Reaper case at best only *seems* to show that. It might be said that despite our intuitive reaction to the case it does not establish what I say it establishes, either because when we react as we do we are not correctly imagining the future we are getting, or else because even if we *are* correctly imagining it, we think we have a reason to want it that we do not really have.

To begin to answer the first of these worries, let us try to see what a period of conscious life that is on-balance neutral in its contents would be like. In filling in the details, two variables stand out. First, we need to specify the elements of wellbeing or ill-being in the period at each given moment, as well as their respective intensities. The second variable is how the value of the part of a life supervenes on the elements of wellbeing occurring at each

moment. It is at least not obviously true that this value is just a sum of momentary wellbeing, just as it is not obvious in the case of a whole life.<sup>13</sup>

We can attempt to skirt these issues by concentrating on a life whose only particular contents are experiences that are themselves roughly neutral (call this the “flat-line future”). At least the first issue does not arise, and it seems unlikely that there would be either an additional higher-order good or a higher-order bad generated by the shape of a totally flat-line period of life. It is unclear what the day-to-day details of this flat-line future would look like. Parfit suggested facetiously that a life just barely worth living would contain only muzak and potatoes.<sup>14</sup> That may be already taking into account the value of conscious life itself, on the assumption that it has value. Would the contents of a conscious life neutral in its contents be only potatoes? (No one who has been on hold with a call centre could deny that a life of only muzak is worse than death.)

Still, it is not as though we have no idea how to imagine such a life. Picture a future existence consisting of humdrum wanderings about, thinking, but about nothing much in particular, and the occasional spell of daytime TV. Perhaps this is as close as we can get to a period of neutral con-

---

<sup>13</sup> See, for example, Velleman (1991); Kamm (1993), pp. 67–71; Dorsey (2015).

<sup>14</sup> Parfit (1986b).

scious life that is recognisably our own. (I shall have something to say later about lives that are not recognisably our own). These seem close to neutral experiential contents. When I try to picture a future like this one, admittedly I cannot help but imagine some mild contentment. But I also imagine boredom, loneliness, and a dim awareness of the relative meaninglessness of my existence. The latter elements of experience are not neutral but decidedly bad, so it is not clear that the overall value of the contents of the future I am imagining is overall better than neutral. Yet I still think this future is far preferable to death. And if others agree, that would seem to indicate that we take the continuation of our conscious life itself to be good for its own sake.

It might be said that when I am told I am getting neutral future, the future I *think* I am getting is better than the one I am *actually* getting. Many philosophers have argued that we are systematically overoptimistic about how good a period of life will be, and psychological research supports that claim.<sup>15</sup> At this stage it is important to distinguish what we might call *descriptive optimists* and *normative optimists*. We can best illustrate the difference by an example, in which a person is imagining what it will be like to eat a certain chocolate bar. If the person is a descriptive optimist, then she over-

---

<sup>15</sup> Schopenhauer (1851/1981); Mackie (1985); Tännsjö (1992); Huemer (2008); Benatar (2006; 2017). For a survey of the psychological research to that effect, see chapter 3 of Benatar (2006).

estimates how much enjoyment she will get out of eating it. When it comes down to it, it will be less enjoyable than she thought. But it is conceivable that this person might correctly imagine just how much enjoyment she will get out of the chocolate bar while mistakenly thinking that the reason she has to want that enjoyment is stronger than it really is. She would then be a normative optimist.

Much of the literature which argues that we are optimists focuses on descriptive optimism. But if we are systematically overoptimistic in that sense, this is no obvious challenge to the Grim Reaper argument. For the optimism that I have has already been added, in the manner of a background assumption, to my evaluation of the period of life I am being offered. If I believe the experiential contents of the life I am imagining are neutral overall, and I would still prefer living on to dying, then that must be because I am placing value on conscious life itself. I may, for example, be imagining more happiness in the future there will actually be. But because I (mistakenly) judge the life to be neutral, if I prefer it, then it seems that would be for something besides the happiness itself.

Normative optimism, on the other hand, is a problem for the Grim Reaper argument. If I am a normative optimist, then even if I could correctly

imagine what a conscious future neutral in its contents would look like, I will mistakenly believe I have more reason to want it than I actually have. And there may be good reason to think we are normative optimists about continuing our conscious lives.

### **2.3. Am I mistaking the reasons I have?**

One reason to think we are normative optimists is that it seems that our evolutionary ancestors would have evolved the predisposition to prolong their lives when possible. Sharon Street has pointed out “how beneficial (in terms of reproductive success) it would be to judge that the fact that something would promote one’s survival is a reason in favour of it”.<sup>16</sup> Street’s target is realism in metaethics, but it is easy to see how the same observation threatens the first-order judgment elicited by the Grim Reaper case. Believing one had a reason to prolong one’s life would have been advantageous to our ancestors. So it seems that the disposition to judge that one had a reason to continue living would have been naturally selected for whether or not one really *did* have that reason. So we should expect that we

---

<sup>16</sup> Street (2006), p. 114.

would find the Grim Reaper's offer attractive whether or not it was really in our interest to accept it. If we would accept the offer whether or not we ought to accept it, then our acceptance is no evidence for the claim that it is in our interest to accept it.

We should bear in mind that evolutionary debunking arguments like this threaten to undercut many other evaluative beliefs we would be reluctant to let go—that it is in our interest to avoid pain, for example.<sup>17</sup> Indeed, an evolutionary story can probably be told for most of the mainstream constituents of wellbeing, so it would be prejudicial to treat the evolutionary objection as a special challenge to the claim that having a conscious life itself has positive value for its subject.

Even still, it is not obvious that Street's argument can wholly debunk the intuition evoked by the Grim Reaper case. The belief that one has a reason to stay alive for its own sake would have been advantageous only in those cases in which it enabled one to reproduce or to make a significant contribution to the survival and reproduction of one's descendants. But people who are beyond the age at which their continued existence is useful for those ends do not lose their will to live. Indeed, since by staying alive these indi-

---

<sup>17</sup> Ibid. For an argument to the effect that any evolutionary debunking argument with implications for normative ethics also implies that pain is not bad, see Kahane (2011).

viduals would consume resources that could instead be used to promote the survival of younger generations, if anything we should expect evolutionary pressures to push in the other direction after a certain age. Of course, I do not claim that no evolutionary story can be told to account for this and other related complications. But to my knowledge none currently exists. And until we have a plausible and non-*ad hoc* explanation, rejecting the Grim Reaper intuition on the basis that having it would have been evolutionarily advantageous is too hasty.

I shall consider two final objections to the Grim Reaper argument. First, it might be said that the only reason I would accept his offer is that I am afraid to die. (Of course, I might fear death precisely because it deprives me of more conscious life, which I recognise to be non-comparatively valuable, but clearly that would not challenge what the case is supposed to show.) But consider a case in which I can secretly accept the Reaper's offer on behalf of someone else whose wellbeing I was concerned to promote, but whose death I would not be alive to experience, such as my grandchild. It seems that to do so would be to confer upon that person a benefit, but fear of my own death cannot plausibly explain that intuition.

Lastly, it may be allowed that we desire more conscious life for its own sake, but said that it is a “mere” desire, not something we have an independent *reason* to desire. If that is so, then while there may be something of interest in my proposal as a psychological observation, it has little obvious philosophical significance. One way to respond to this worry is the one I have just given: I would take myself to have a reason extend the conscious life of someone else who would otherwise die, even if I could be assured that the particular contents of the additional period of life would be on-balance neutral; and I would feel that to do so would be to confer a benefit on her. It might be said that this is just an inference to the fact that this person would want to extend her life in that way, just as I myself would, and I have a reason to help her realise what she wants. At this point, I can only retreat to the phenomenological argument that it feels that I want to extend my conscious life because I have an independent reason to do so, not that what reason I have to extend it is parasitic on the desire. In the end, I know of no other way to distinguish a mere desire from what we have a reason to desire.



#### 2.4. When conscious life might be good

I conclude that there is good reason to think that conscious life is prudentially valuable for its subject in its own right. We should ask, finally, when exactly it is valuable and why. On the most extreme view, any period of conscious life, irrespective of its content, is in one way prudentially valuable for its subject. But suppose that in place of dying one could enter a minimally conscious state in which one will experience only muted colours, white noise, and the faint sensation of bedclothes on skin. Is there anything to be said for this painless but pleasureless future, which nonetheless contains experience itself? For my part, I see no reason to want this life, which seems, if anything, worse than death.

It might be replied that one would not have a reason to want such a life either because it would be too psychologically discontinuous with oneself now, or alternatively because the moment-to-moment psychological connections would be so weak as to have little prudential value for the one whose life it is at any given moment. I shall have a bit more to say about this in the next chapter, but we can note for now that this is probably too restrictive a view to take of a psychological view about either identity or the grounds of prudential concern. At any given time, much of a person's psychology is non-

conscious. It seems that someone in a minimally conscious state might be related to me in the way necessary to ground rational prudential concern even if he was too benumbed to be aware of who he was.

The better suggestion is, I think, the somewhat Millian one that if conscious life is to be seriously worth wanting it must be sufficiently sophisticated. In his conditional defence of preference-hedonism against its rivals, Parfit cited James Griffin's observation that toward the end of his life Freud refused sedating painkillers—preferring instead to suffer with a clear mind—and suggested that, given his preference, Freud's life went better for it.<sup>18</sup> An alternative explanation for why this was better for Freud is that he was able to continue to enjoy the richness of experience that is necessary for additional conscious life to be good for the one whose life it is. (This suggestion is compatible with the possibility that Freud's tacit recognition of that value is what explained his preference.)

Of course, not least because the complexity of consciousness itself exists on a continuum, it would be implausible to think that there is any sharp cut-off point here. We should say that independent of its particular contents, a given period of conscious life is more valuable for the one whose life it is the more complex it is. Hence, a year of the conscious life of a slug has at most

---

<sup>18</sup> Parfit (1984), pp. 493–4.

minimal value for it, a year of the conscious life of a pig more valuable, and a year of the conscious life of a normal adult human being better still. The rough explanation for the difference, I suggest, is that while a person's normal conscious life continues on, she remains "around"; aware of and engaged with herself and her surroundings. Obviously that is not so when one is in a coma, but it is also not true of a being that is only barely conscious. As opposed to being minimally conscious or a corpse, the continuation of the robust and complex life of a person means that she remains a subject in the world.

## Chapter 3: Not All Killings Are Equally Wrong

Ronald Dworkin once observed that “most people think . . . that the murder of a depressive handicapped octogenarian misanthrope is as heinous, and must be punished as seriously, as the murder of anyone younger or healthier or more valuable to others.”<sup>1</sup> Jeff McMahan, a prominent defender, calls this view the *Equal Wrongness Thesis*.<sup>2</sup> As he puts it, the moral objection to killing persons “does not vary with such factors as the degree of harm caused to the victim, the age, intelligence, temperament, or social circumstances of the victim, whether the victim is well liked or generally despised, and so on.”<sup>3</sup>

Whether the Equal Wrongness Thesis is true is of obvious theoretical interest. It also has important practical implications, since any factor (age, for example) that is irrelevant to the strength of the moral objection to killing a person is *ipso facto* irrelevant to its overall permissibility. Suppose

---

<sup>1</sup> Dworkin, (1993), p. 85.

<sup>2</sup> McMahan (2002). See also McMahan (1995).

<sup>3</sup> McMahan (2002), p. 235.

we must decide, as the German Federal Constitutional Court once did, whether to pass legislation permitting the gunning down of hijacked passenger planes when doing so is necessary to prevent their use as a terrorist weapon to kill many people.<sup>4</sup> If how wrong it is to kill a person is unaffected by how bad it is for her to die, then it makes no difference to the permissibility of shooting down these aircraft that the passengers on board them would almost certainly have died soon anyway.

Kasper Lippert-Rasmussen has argued that we should reject the Equal Wrongness Thesis on the basis of the moral equivalence of killing someone and knocking her unconscious. Several authors have raised objections to his argument for that equivalence, however, and, as I shall show, there are other issues with his approach that have not yet been noted. Nevertheless, in what follows I shall argue that killing someone and knocking her unconscious are indeed morally equivalent in the way that is required to show that not all killings are equally wrong.

The structure of this chapter is as follows. In section 3.1, I clarify the Equal Wrongness Thesis. In section 3.2, I introduce Lippert-Rasmussen's ar-

---

<sup>4</sup> Bundesverfassungsgericht, *Judgment of the First Senate of 15 February 2006*. Of particular relevance is paragraph 132, in which the court argues that “[h]uman life and human dignity enjoy the same constitutional protection regardless of the duration of the physical existence of the individual human being.”

gument against that thesis and discuss some of the ways I believe it falls short. In sections 3.3 and 3.4, I defend the claim that, other things being equal, it is no less wrong to deprive someone of a certain amount of conscious life by knocking her unconscious than it is to do so by killing her, regardless of when in her lifetime it takes place. I conclude, in section 3.5, by explaining why that claim is sufficient to defeat the Equal Wrongness Thesis. In brief, along with the Equal Wrongness Thesis, it implies that knocking someone temporarily unconscious for a short period is just as wrong as killing someone with many good years left to live. Because that result is absurd, we should reject the Equal Wrongness Thesis.

### **3.1. Preliminary remarks**

Several initial clarifications of the Equal Wrongness Thesis are necessary. First, as McMahan explains, the sense of “wrongness” that is relevant for the Equal Wrongness Thesis is one that admits of degrees.<sup>5</sup> An act’s wrongness in this sense is the degree to which it is morally objectionable: one act is more wrong than another insofar as it is subject to a moral objection that is harder

---

<sup>5</sup> McMahan (2002), p. 190. See also Lippert-Rasmussen (2007), p. 717. Cf. Soto (2013), pp. 551–4.

to outweigh by competing morally-relevant considerations. It is important to note, though, that the Equal Wrongness Thesis applies only to acts of killing considered on their own. Suppose that on Monday I kill a young person and on Tuesday I kill an old person. The Equal Wrongness Thesis tells us that what I do on each day is equally wrong and would have been equally difficult to justify. But two acts being equally hard to justify does not imply that in a choice between them we ought to be indifferent. Suppose we must choose *between* killing a young person and killing an old person, for example as a necessary side effect of averting the deaths of many other innocent people. It is consistent with the Equal Wrongness Thesis that in this case we ought to kill the old person.<sup>6</sup>

Second, following McMahan and others, I understand the Equal Wrongness Thesis to be restricted in scope to a certain class of killings.<sup>7</sup> It does not apply to killings that differ in morally relevant but extrinsic ways, such as their effects (positive or negative) on third-parties, or the relationship between the perpetrator and the victim. Nor does it apply to killings that differ with respect to the presence or absence of potential justifying consider-

---

<sup>6</sup> McMahan (2002), p. 237.

<sup>7</sup> McMahan (2002), pp. 236–7. Other discussants of the Equal Wrongness Thesis also apply these restrictions to it. See Lippert-Rasmussen (2007), pp. 718–19; Soto (2013), p. 544; Hanser (2013), pp. 392–409.

ations, such as the victim's moral responsibility or her free consent to being killed. It also does not apply to killings that differ in their modes of agency—for instance, whether the victim is killed as a means to some end or as a side effect of the pursuit of that end. Lastly, the Equal Wrongness Thesis concerns only the killing of individuals who satisfy the conditions of personhood, whatever those conditions are. It is compatible with any view about the wrongness of killing non-persons. I will usually omit the preceding qualifications and take them to be understood when I discuss an instance of killing.

### 3.2. Lippert-Rasmussen's trilemma

In “Why Killing Some People Is More Seriously Wrong than Killing Others”, Kasper Lippert-Rasmussen challenges the Equal Wrongness Thesis by way of a trilemma consisting of the following three claims<sup>8</sup>:

- (1) *The Equal Wrongness Thesis*: The degree to which different killings of persons are wrong does not vary: all such killings are equally wrong.

---

<sup>8</sup> Lippert-Rasmussen (2007).



(2) *The Unequal Wrongness of Renderings Unconscious Thesis:*

For any period  $\omega$ , it is more wrong to render a person unconscious for a period longer than  $\omega$  than it is to render a person unconscious for  $\omega$ , other things being the same.

(3) *The Equivalence Thesis:* It is neither more, nor less, wrong to deprive a person of a certain amount of conscious experience by killing her than it is to deprive her of the same amount of conscious experience by rendering her unconscious, other things being the same.

To see that these three claims cannot all be true, consider the following two pairs of acts. In the first pair, I kill someone with fifty years left to live and then kill someone with just one year left to live. According to (1), these two acts are equally wrong. In the second pair, everything is the same but instead of killing them I administer to each a powerful drug that induces a total coma that will last until her death. According to (2), now what I do to the person who will live another fifty years is more seriously wrong than what I do to the person who will live for just another year. Yet if the two preceding judgments are correct, then (3) cannot be true. For if depriving someone of a certain amount of conscious experience by killing her is no more or less wrong than doing so by knocking her unconscious until her death, the relative

wrongness must be constant across both pairs. Lippert-Rasmussen argues that of these three claims we should give up the Equal Wrongness Thesis.

I take the crucial premise in Lippert-Rasmussen's argument to be (3). There are a number of ways to challenge this premise. For one thing, unlike when she is knocked unconscious, when a person is killed she will never regain consciousness. Since it may matter for the wrongness of an act whether it deprives its victim of her last period of conscious life, we might doubt that depriving someone of a given amount of conscious experience by knocking her unconscious is morally equivalent to killing her.

Lippert-Rasmussen suggests two individually sufficient ways of responding to this worry. The first is to restrict (3) so that it covers only instances of unconsciousness that persist until the victim's death.<sup>9</sup> (Whether we take this restriction to be covered by a suitable interpretation of the premise's *ceteris paribus* clause or else to constitute a revised version of the premise itself is not a substantive matter.) If we take this route, it cannot be argued that the two types of act are morally different on the grounds that it makes a difference whether or not the victim will regain consciousness.

The other response is to keep the more expansive, unrestricted version of (3), and argue independently that it is not in itself morally significant

---

<sup>9</sup> Ibid., p. 727.

whether an act deprives its victim of what would have been her last period of consciousness. To that end, Lippert-Rasmussen proposes that it only seems more wrongful to deprive someone of what would have been her last conscious period because we mistakenly assume that it would be an especially great loss for her, or that it is especially disrespectful to do so.<sup>10</sup>

With respect to the deprivation of what would have been a victim's last conscious period constituting an especially great loss for her, Lippert-Rasmussen notes that there is no reason, in principle, as to why the final period of a person's life is more valuable or important than any other.<sup>11</sup> It is worth noting that he does not consider one reason it might be more valuable—namely, that the permanent end of consciousness is non-comparatively bad for a person—but I have already argued against that view in the previous chapter. In any event, however, it is not far-fetched to think that ensuring that someone will never again experience anything is particularly egregious in its own right. In a similar vein, Daniel Cohen and Morgan Luck have argued that the wrongness of a given deprivation of consciousness depends on how much of the victim's remaining conscious life it eliminates.<sup>12</sup> Before we can

---

<sup>10</sup> *Ibid.*, p. 726.

<sup>11</sup> *Ibid.*

<sup>12</sup> Cohen and Luck (2009), pp. 396–401.

accept the more expansive version of (3), we need to show that the wrongness of knocking someone unconscious does not depend on when in the victim's life the period of unconsciousness occurs.

Because of these issues with the more expansive construal of (3), perhaps an opponent of the Equal Wrongness Thesis should restrict (3) in such a way that it covers only those instances of knocking someone unconscious that last for the rest of her life. Lippert-Rasmussen suggests that restricting (3) in this way is already sufficient to answer the worry at hand, since the version of the trilemma that emerges is just as forceful as the unrestricted one.<sup>13</sup> In the first place, however, even if that is right, it may still be doubted that depriving someone of consciousness by killing her is morally equivalent to doing so by knocking her unconscious until her death. Lippert-Rasmussen offers several arguments for their equivalence, but, as I will argue in section 3.3, none is wholly convincing. There are also disanalogies between killing someone and knocking her unconscious that he does not consider. For example, Matthew Hanser has since objected that the fact that killing someone terminates all her basic biological operations is a factor in making killing

---

<sup>13</sup> Lippert-Rasmussen (2007), p. 727.

wrong.<sup>14</sup> Alternatively, it might matter simply that by killing someone, but not by knocking her unconscious, one causes the victim to cease to exist.

There is a deeper problem with restricting (3) to cover only deprivations of consciousness that continue to the end of the victim's life. If we take this route, then in order for the trilemma to go through, (2) must also be understood to apply only between instances of knocking someone unconscious that last for the rest of the victim's life. But the original, unrestricted version of (2) may have been compelling in the first place because of an implicit assumption that the instances of making people unconscious that it compares do not involve ending the victim's conscious life forever. Lippert-Rasmussen states that he will not pursue an independent argument for (2).<sup>15</sup> But for those who, like me, believe the Equal Wrongness Thesis has considerable *prima facie* plausibility, it may not be obvious that the wrongness of rendering someone permanently unconscious does depend on the amount of conscious life she loses. Since the force of the trilemma lies precisely in the independent plausibility of the Equal Wrongness Thesis, if we restrict (2) and (3) to cover only those acts that permanently end their victims' conscious life,

---

<sup>14</sup> Hanser (2013).

<sup>15</sup> Lippert-Rasmussen (2007), p. 722.

there is probably insufficient reason to resolve the trilemma by rejecting the Equal Wrongness Thesis, rather than by rejecting (2).

To defeat the Equal Wrongness Thesis, therefore, we need to show both that the wrongness of knocking someone unconscious does not depend on when in the victim's life the unconsciousness occurs *and* that any other properties of killing someone that distinguish it from knocking someone unconscious—such as the fact that killing someone ends her life—are not themselves morally relevant.

The rest of this chapter is an attempt to do just that. In particular, I shall argue for the following claim:

(C) It is as wrong to deprive someone of a certain amount of conscious life by knocking her temporarily unconscious in the middle of her life as it is to deprive her of the same amount of conscious life by killing her, other things being the same.

My strategy is to first defend, in a novel way, the *restricted* version of (3)—namely, that killing someone and knocking her unconscious until her death are morally equivalent, other things being the same. With that conclusion established, I will then offer a number of arguments to show the wrongness of knocking someone unconscious for a given period is unaffected by when in

the victim's lifetime it takes place. Together, these two premises entail (C), which is sufficient to refute the Equal Wrongness Thesis. That thesis implies that killing someone a month before she would otherwise have died is as wrong as killing someone with fifty good years left to live. Hence, if we accept (C), we can maintain the Equal Wrongness Thesis only by accepting that it is just as wrong to knock someone temporarily unconscious for a month as it is to kill someone with fifty good years left to live. Since that result is absurd, we should abandon the Equal Wrongness Thesis.

### **3.3. The moral equivalence of killing someone and knocking her unconscious until her death**

In this section, I will defend my argument's first premise, which is the restricted version of Lippert-Rasmussen's Equivalence Thesis mentioned earlier:

(A) It is as wrong to deprive someone of a certain amount of conscious life by knocking her unconscious until her death as it is to deprive someone of the same amount of conscious life by killing her, other things being the same.

In his own defence of that thesis, Lippert-Rasmussen employs a pair of cases in which

(1) I kill a person twenty years before he would otherwise have naturally died;

(2) I knock a relevantly similar person unconscious for twenty years, at the end of which he dies a natural death, and I do so knowing that he will not regain consciousness before dying.<sup>16</sup>

Lippert-Rasmussen proposes, first, that these two acts are equally bad for their victims.<sup>17</sup> But the fact that two acts are equally bad for their victims does not imply that they are equally wrong. (Indeed, if an act's wrongness always correlated with its harmfulness we could reject the Equal Wrongness Thesis out of hand.) Lippert-Rasmussen acknowledges this, but argues that in the absence of an explanation for why two acts differ morally the fact that they are equally harmful is a reason to believe them equally wrong.<sup>18</sup> But if

---

<sup>16</sup> Ibid., p. 722.

<sup>17</sup> Ibid. He adds that any cost it would be permissible to impose upon me to prevent my doing (1) could also be permissibly imposed upon him to prevent him doing (2). I agree with Matthew Hanser (2013, p. 379n17) that this argument would sway only someone to whom these acts already seem morally equivalent.

<sup>18</sup> Lippert-Rasmussen (2007), p. 722.



two acts seem to differ morally, then even lacking an explanation for *why* they differ, we are probably unwarranted in concluding them to be equally wrongful on the sole basis that they are equally harmful—and it is questionable that most people would intuitively believe that (1) and (2) are equally wrong. Moreover, there *are* a number of reasonable candidate explanations for that difference: for example, that killing someone ends that person's life, whereas knocking her unconscious does not.

The other argument Lippert-Rasmussen gives for the moral equivalence of killing someone and knocking her unconscious is that their equivalence follows from existing theoretical accounts of *why* killing is wrong.<sup>19</sup> If the factors that explain why killing someone is wrong are present in equal measure for knocking someone unconscious, it seems reasonable to assume that the two types of act are morally equivalent. For instance, if the intrinsic wrongness of killing is wholly explained by its harmfulness, then, on the assumption that being knocked unconscious is no better or worse for a person than being killed, it is reasonable to conclude that the strength of the moral objection is the same for both types of act. Alternatively, suppose the wrongness of killing is fully explained by its disrespectfulness to the victim. Then, on the assumption that the factors that make killing someone disrespectful apply

---

<sup>19</sup> Ibid., pp. 723–4.

equally to knocking her unconscious, quite probably the two types of act are morally equivalent.<sup>20</sup>

The general problem with this argument is that whether or not all killings are equally wrong is itself a major consideration that any account of why killing is wrong must accommodate. Indeed, it is primarily on the basis of his conviction that all killings are equally wrong that McMahan rejects theoretical accounts of the wrongness of killing that imply otherwise, and it is also on the basis of that conviction that he proposes his own Intrinsic Worth Account, whose chief selling point is that it entails the Equal Wrongness Thesis.<sup>21</sup> If knowing whether the Equal Wrongness Thesis is true is necessary to adjudicate among competing accounts of the wrongness of killing, we cannot justifiably use accounts of the wrongness of killing themselves to decide whether the Equal Wrongness Thesis is true.

---

<sup>20</sup> Ibid., p. 724. In defence of his claim that knocking someone unconscious is *ceteris paribus* no less disrespectful than killing her, Lippert-Rasmussen points to the fact that persons deserve respect not simply in virtue of being alive, but rather in virtue of "certain cognitive and emotional capacities". But an unconscious person might retain her capacities for cognition and emotion. Furthermore, this argument seems to conflate the *bases* of respect for persons and the *manner* in which we should respect them. That we should respect people in virtue of certain capacities they have does not imply that respect for them consists in respect for those capacities.

<sup>21</sup> McMahan (2002), pp. 241–3.

In the rest of this section, I will attempt to establish the moral equivalence of killing someone and knocking her unconscious until her death in a different way. My argument runs as follows. If, other things being equal, killing someone really is more wrong than knocking her unconscious until her death, then killing someone must have some wrong-making property not shared by knocking someone unconscious until her death. But there is no such property. Therefore, the two acts are morally equivalent.

I take it that the conceptual difference between killing someone and knocking her unconscious until her death is that the former causes the victim's death, whereas the latter merely ends her conscious mental activity until she dies in a causally unrelated way. Thus, how we should conceptually distinguish the two kinds of act will depend on how we define death. Though many definitions of death (and by extension of killing) have been offered, all of them fall into one or the other of two categories. The first category, which we can call the *Existence Conception* of death, identifies a person's death with the cessation of her existence. If death is the end of existence, then metaphysical difference between killing someone and knocking her unconscious until her death is, roughly, that the former ends the person's existence, while the latter merely causes her to have no conscious mental activity until

she ceases to exist for a causally unrelated reason. As we shall see shortly, as a conception of killing, this notion needs fine-tuning, but this rough characterisation of it will do for the moment.

The second category of definitions of death, which we can call the *Biological Conception*, identifies a person's death with the cessation of functioning of one or more of her basic organismic processes. On the Biological Conception, the difference between killing a person and knocking her unconscious until her death is that the former causes the permanent cessation of the vital biological processes while the latter merely causes her to have no conscious mental activity up to the time that those processes cease due to independent causes.

These two conceptions—Existence and Biological—cover all plausible definitions of death. Hence, they cover all plausible views about what distinguishes killing someone from knocking someone unconscious until her death. Therefore, if it can be shown that on neither conception of death does the fact that one causes it constitute a wrong-making property of an act, we will have shown that the two types of act are morally equivalent.

I begin with the Existence Conception of death. On this conception, killing ends someone's existence while knocking someone unconscious until

her death merely causes her conscious mental life to cease up to the time that her existence ends for other reasons. Hanser sometimes adopts this conception of killing, for instance when he writes that killing a person is more seriously wrong than knocking her unconscious because “[s]omeone who has been . . . knocked unconscious continues on in an impaired state; someone who has been killed does not continue on at all.”<sup>22</sup>

This point recalls one made by Frances Kamm. She writes:

Suppose we put someone into a coma, knowing that he will never recover from it . . . In such a case, we would not, I believe, be treated as harshly as if we had killed someone . . . The fact that one person determines the nonexistence of another against his will . . . is a factor in making killing wrong.<sup>23</sup>

Conceiving killing simply as causing non-existence for the purposes of moral analysis is a mistake, however, because of well-known branching cases in the theory of personal identity, whereby an individual is caused to cease to exist in ways that seem morally quite unlike killing her.

Consider first a stock case of what looks like a wrongful killing:

---

<sup>22</sup> Hanser (2013), p. 399.

<sup>23</sup> Kamm (1993), p. 21.

*Riding Roughshod*: I am driving to the hospital to receive an urgent life-saving treatment when I notice you crossing the road ahead of me. Pressed for time, I run you over, killing you.

Now contrast that case with the following one:

*Your Involuntary Division*: I am driving to the hospital to receive an urgent life-saving treatment when I notice you crossing the road ahead of me. I press a button on my dashboard that causes you to divide amoeba-like into two individuals and then deposits each of your two “fission offspring” on either side of the road, where my car passes safely between them.

Like *Riding Roughshod*, pressing the button in *Your Involuntary Division* arguably causes you to cease to exist. It does so because, for well-known reasons I shall not rehearse here, you would not be numerically identical with any individual who exists afterward.<sup>24</sup> Yet to press this button is plainly not as objectionable as fatally running you over. My own intuitions are that causing your nonexistence en route to saving my own life is impermissible in *Riding Roughshod* but permissible in *Your Involuntary Division*. But whatever one

---

<sup>24</sup> See Parfit (1984), pp. 255–260.

thinks about the overall permissibility of these two cases, it is clear enough that the moral objection to my action is weaker in the second case than in the first.

At this point, someone who would still like to conceive killing as causing nonexistence might reply in the following way. Unlike causing your nonexistence by running you over with my car, causing your nonexistence by splitting you into two individuals preserves, or mostly preserves, “what matters in survival”. That is why, it may be said, it is less objectionable to kill you in that case. Since this reply presupposes that the moral objection to causing a person’s nonexistence depends on the badness for the victim of ceasing to exist, though, to affirm it is already to reject the Equal Wrongness Thesis.

I do not deny that there is some moral objection to causing a person to divide in two against her will. But it seems that whatever objection there is does not arise from the bare fact it brings about her nonexistence. It arises, rather, from the fact that it involves the violation of the victim’s bodily integrity or some cognate consideration concerning her moral rights against interference. For that reason, *Your Involuntary Division* should not be described as an instance of killing—at least not on the conception of that term that is useful for analysing the moral wrongness of killing people.

These remarks show that analysing killing as causing nonexistence is inadequate. But they should not, I think, lead us to abandon the Existence Conception wholesale. Instead, we should accept that there are two morally salient ways of causing an individual's nonexistence, which we might label "killing by annihilation" and "killing by duplication". Killing by annihilation is a matter of causing an individual to cease to exist by eliminating its essential *intrinsic* properties (i.e., those intrinsic properties of it which it could not possibly lack). Killing by duplication is a matter of causing an individual to cease to exist by eliminating its essential *extrinsic* properties—for example, by creating one or more "continuers" of it, each of which is at least as good a candidate for numerical identity to the original individual as another candidate, and none of which is numerically identical to the original.

The version of the Existence Conception of killing that is useful for moral theorising about killing is killing by annihilation, and I shall henceforth understand the Existence Conception of killing to refer only to that way of causing someone to cease to exist. Recall now that (A) claims that depriving someone of a given amount of conscious life by knocking her unconscious until her death is as wrong as depriving her of the same amount of conscious life by killing her. In light of our more sophisticated understanding of the Ex-



istence Conception, to assess (A), we need to consider the various candidates for what intrinsic properties of a person are essential. I shall now argue that on none of the mainstream accounts of these properties is there a moral difference between killing someone and knocking her unconscious until her death.

Consider, first, the psychological continuity view, according to which a person will exist in the future if and only if there will be one person to whom she is appropriately connected via overlapping beliefs, memories, desires that lead to actions, and so on.<sup>25</sup> Because most of our psychology can exist without being consciously active, on a plausible understanding of the psychological continuity condition it will be possible to knock someone unconscious until her death without causing her to cease to exist. On this condition for continued existence, the difference between killing someone and knocking her unconscious until her death is that the latter does not cause the nonexistence of the psychological states that ground her persistence.

Next, consider what McMahan calls the “embodied mind account”.<sup>26</sup> According to this view, a person continues to exist if and only if there is the unique physical and functional (or potentially functional) continuity of enough of those parts of the brain to retain the capacity for consciousness.

---

<sup>25</sup> Shoemaker (1970; 1984); Perry (1972); Lewis (1976b); Parfit (1984).

<sup>26</sup> McMahan (2002). Parfit (2012) later defended a similar view.

On the embodied mind account, the difference between killing someone and merely knocking her unconscious until her death is that only the killing involves the destruction of enough of the relevant parts of the brain.

I believe that on neither the psychological continuity condition, nor on the embodied mind account, is it morally relevant whether one's action causes someone to cease to exist. To see why, it will be instructive to consider two real ways a human being can be biologically alive even though she will never again be conscious.<sup>27</sup> The first is a permanent vegetative state, in which the victim has permanently lost consciousness because of irreversible damage to her cortex and limbic system. In a permanent vegetative state, the only parts of the brain that retain any functional capacity are the brainstem's vegetative centres: those responsible for heart rate, blood pressure control, temperature control, and respiration. The other way involves damage to the ascending reticular activating system, a network of cells in the brainstem that is necessary for any consciousness to occur.<sup>28</sup> A patient with localised damage to the reticular activating system retains the *capacity* for consciousness, as well as the bulk of her distinctive psychology, which is stored in (or consti-

---

<sup>27</sup> For illuminating surveys of these phenomena, see McMahan (2002), especially pp. 423–455, and Horne (2009).

<sup>28</sup> Zeman (2001); McMahan (2002); Horne (2009).

tuted by) parts of the cerebrum. But because of the damage, as a matter of fact she will never again have any conscious experience.

Suppose that the conditions for an individual to continue to exist are either the continued existence of her psychology or the physical and functional continuity of those parts of the brain that have the capacity for consciousness. If it was in itself morally objectionable to terminate a person's existence, then on either of these accounts it would be objectionable to end a patient's life by terminating life support if she was in a total coma from which she would never recover due to damage to the reticular activating system, but not to remove life support from a patient in a permanent vegetative state. That is because the former ends the continuity of the patient's psychology as well as the capacity for consciousness, whereas in a permanent vegetative state they have already been destroyed. It is very hard to believe that there is this difference.

The preceding argument shows, I believe, that on neither of these two accounts of what it takes for someone to continue to exist is it in itself morally objectionable to cause a person's nonexistence. But it shows this by appealing to atypical cases in which the conscious life of the victim has already ended forever. It might be replied that terminating a person's existence *is*

especially objectionable, but only so long as the conscious life of the victim is not already over. On this view, the wrongness of ending a person's existence is a "combination effect": when ending a person's existence *also* ends her conscious life, the former constitutes an additional moral objection over and above the wrongness of the latter. It would not, however, be more wrongful to cause a conscious person to be in a permanent vegetative state than it would be to cause her to be permanently unconscious by destroying her reticular activating system. That is so even though only the former would involve the termination of psychological continuity as well as the destruction of those parts of the brain that have the capacity for consciousness. It is not plausible that there is a stronger moral objection to ending someone's conscious life forever based only on which part of the brain one damages. Although the psychology of a patient with permanent damage to her reticular activating system survives, it is only "stored". The fact that a person's *non-conscious* psychology is not destroyed would not, by itself, mitigate the wrongness of terminating her *conscious* life.

Lastly, consider a biological continuity condition of existence, according to which a person will exist in the future if and only if her organism continues to exist and function biologically.<sup>29</sup> If continued existence is conditional

---

<sup>29</sup> Olson (1997); Snowdon (2014).

upon biological continuity, however, then being killed on the Existence Conception of death coincides with being killed on the Biological Conception. Because I will also argue that killing someone and knocking her unconscious until her death are morally equivalent when killing is understood on the Biological Conception, I can argue that they are morally equivalent on the biological version of the Existence Conception incidentally, since my arguments will also apply to it.

Let us turn, then, to the Biological Conception of death, according to which the difference between killing someone and knocking her unconscious until her death is that the former but not the latter involves the termination of the victim's vital biological processes. In some passages, Hanser seems to be adopting this conception of killing, instead of the Existence Conception. He writes:

An agent arguably interferes with someone in a much more fundamental way by killing him than he does by knocking him unconscious . . . The former causes the complete and permanent cessation of the victim's basic life-sustaining (or life-constituting) bodily operations; the latter leaves the victim alive but, for a time,

unable to exercise a variety of his practical and perceptual capacities.<sup>30</sup>

Hanser here cites two distinct properties of killing someone which allegedly make it more wrong than knocking someone unconscious. One is that it causes the cessation of the victim's vital bodily operations; the other is the permanence of that cessation. These components are separable. Medical advances might one day permit the biological resuscitation of a body that has ceased to function completely.

We should ask which of the two properties (or both) is morally significant. Consider first the appeal to the *permanence* of the state that killing brings about. Of course, assuming that one's victim is mortal, knocking her unconscious does not change the fact that there will be *some* time at which she will be permanently dead. So, if killing a person is more objectionable than knocking her unconscious because of the permanence of death, that must be because by killing someone one is the *cause* of her permanent biological death. To see whether this is a wrong-making property of killing, contrast a pair of cases in which, absent any interference, the victim would have lived forever:

---

<sup>30</sup> Hanser (2013), p. 399.

(1) I cause my victim's vital biological operations to permanently cease at  $t_1$ , thereby pre-empting their permanent cessation at  $t_2$ .

(2) I cause my victim's vital biological operations to cease for the interval  $t_1-t_2$ . At  $t_2$ , they will be made to cease *permanently* by some independent process (one I fully foresee but do not cause). That other process would have occurred when it did whether or not I had interfered with my victim first.

It seems to me that what I do in (1) is no more wrong than what I do in (2), even though in only the former case do I *cause* the victim's permanent biological death. Causing the temporary cessation of someone's biological functions which will thereafter cease permanently without ever regaining functioning is no less wrong than causing their permanent cessation. In support of this claim, consider an analogous pair of cases in which the harm for which I am responsible is non-fatal:

(3) Albert will be infected by the incurable debilitating Virus X next week if and only if I don't pre-emptively infect him with it today.

(4) Bernard will be infected by Virus Y next week whether or not I infect him with Virus X today. Virus Y has identical symptoms to Virus X, but it will also completely inhibit Virus X if that virus is present in the host when Virus Y is contracted.

What I do in (3) is not, I believe, more seriously objectionable than what I do in (4). Some support for this claim comes from the observation that if I had to infect just one of these men, I would have no more reason to choose Bernard. That is so even though only in Albert's case would I be responsible for the event that is someone becoming debilitatingly ill for the rest of his life. Suppose, furthermore, that we re-imagine the cases so that the viruses are beneficial, rather than harmful, to their hosts. I would have no more reason to give Albert the beneficial Virus X\*, on the assumption that he would anyway contract it next week, if I don't, than I would have to give X\* to Bernard, who will contract the superseding but equally beneficial Y\* next week no matter what I do. Assuming that benefits and harms are analogous in the relevant way, this supports the claim that, in instances of non-lethal harming, the wrongness of one's act is determined by the difference one makes to what happens to the victim, not by the effect on her for which one is responsible. Without a good reason not to, it seems to me that we should



generalise this result to lethal harms. The mere fact that killing someone causes the permanent cessation of her vital biological operations is not, I conclude, a factor in making it wrong.

Instead of appealing to the permanence of death, one might try to show that killing is especially wrong by appealing to its *irreversibility*. It might be said that depriving someone of conscious life by killing her is more objectionable than doing so by knocking her unconscious because the effect of the former on the victim cannot be undone. We can understand this claim to be about metaphysical, physical, or practical irreversibility. Consider first the proposal that killing is especially wrongful because it is either metaphysically or physically irreversible. Suppose I wilfully terminate all of a person's mental processes along with her biological life. Her condition is not metaphysically or physically irreversible, let us suppose, for I know that the drug that would revitalise her has been successfully manufactured in the past. However, I also know that all samples of the drug have been destroyed and the knowledge of how to manufacture it has been lost forever. It is both metaphysically and nomologically possible to make the drug again, so my victim's condition is reversible in both those senses—it is just that I know that it will not be reversed. I believe that it would be no less wrong to end my victim's life

in this case than it would be if the drug to revitalise her had never existed and indeed never could have existed.

Consider next the claim that killing is especially wrong because death is irreversible in a *practical* sense. Killing a person, it might be said, is objectionable in part because by doing so one makes it the case that there is no feasible way to bring that person back to life. Against this claim, suppose the revitalising drug exists, but the last remaining sample is owned by an independently acting agent who I know will never use it to restore the life of my would-be victim. Assume that I know, too, that there is nothing I can do to change this fact. When I end my victim's life in this scenario, my effect is reversible in the practical sense, because there is a real way to bring her back to life. Yet it seems to me that it would be no less wrongful to end her life in that scenario than it would be if the drug had never existed. Because this argument appeals to what I and another agent bring about, it may seem misleading to focus on the wrongness of my action alone. But the mere fact that there is someone *else* who could revive my victim does not mitigate the wrongness of my own action in killing her, given that I know, when I act, that this other person will never in fact revive her and cannot be made to do

so. Ending her life would be no less wrongful if the reason the drug could not be used was some impassable natural obstacle instead of an agential one.

I have argued that neither the permanence of death nor its irreversibility is a factor in making killing a person more wrongful than knocking someone unconscious until her death. Let us turn, then, to the more fundamental wrong-making feature of the Biological Conception: simply that it brings about the cessation of a person's life-sustaining (or life-constituting) bodily operations. I believe that the fact that one act causes the cessation of a person's bodily operations does not, on its own, make its performance more objectionable than another. Suppose that someone in a temporary and medically non-serious coma will soon be connected to a life support system. While she is connected to the machine, if the parts of her brain that support her mental life are destroyed, her bodily operations will continue to function: her heart will pump blood, her lungs will inhale and exhale, and so forth. If those parts of her brain are destroyed while she is *not* connected to the machine the injury will also cause her biological death. If bringing about the cessation of a person's vital biological operations were in itself wrongful, then it would be less objectionable to obliterate this person's brain while she was on the life support system than to do so before she was connected to it, for in

the former case her biological operations would continue to function even after the totality of her mental life was gone. But it is very difficult to believe that these two acts differ in their degrees of wrongness.

Earlier in this section, I postponed arguing that it is not in itself wrongful to terminate a person's existence if personal persistence is a matter of biological continuity. I take the preceding arguments to have now established that conclusion, too. Hence, on none of the mainstream accounts of the conditions for our continuity is terminating someone's existence a factor in making killing wrong.

To summarise, on neither the Existence Conception nor the Biological Conception of killing is there an intrinsic moral difference between killing a person and knocking her unconscious until her death. Because these are the only two plausible conceptions of killing, and because one type of act is more wrong than another only if there is some morally significant property of the former that is not shared by the latter, I conclude that it is just as wrong to knock someone unconscious until her death as it is to deprive her of the same amount of conscious life by killing her.

### 3.4. The temporal irrelevance of knocking someone unconscious

I turn now to the second premise in my argument against the Equal Wrongness Thesis:

(B) It is as wrong to deprive someone of a certain amount of conscious life by knocking her temporarily unconscious in the middle of her life as it is to deprive someone of the same amount of conscious life by knocking her unconscious until her death, other things being the same.

Before proceeding, it is important to clarify two things about (B)'s *ceteris paribus* clause. First, it should be understood to exclude any differences in the prudential value for the victim of the conscious period of which she is deprived. It may be said that the last month of a person's life would have been especially important because it would have allowed her to say her goodbyes, or to realise some end toward which she had been working. But these are contingent facts about a life, and anyway I suspect that for many people who will live until old age a month in the middle of their life is worth at least as much to them as the one at its end. Second, (B) should be understood to exclude differences in the length of the victims' counterfactual conscious lifespans. It is plausible that it is more wrong to deprive someone of a

given amount of consciousness the shorter her total conscious life would have been, but the premise applies only when that difference is factored out.

To get a better sense of (B), consider the following illustration. Calvin and Dylan will both live for exactly 1,000 months (about 83 years). Calvin has lived for 500 months so far while Dylan has lived for 999 months. Apart from their ages, the two men differ in no morally relevant respect. Suppose that the prudential value of what would be Dylan's 1,000th conscious month is the same as that of what would be Calvin's 501st month. (B) says that the strength of the moral objection to knocking out either man for one month is the same.

In a response to Lippert-Rasmussen's article, Daniel Cohen and Morgan Luck have argued that the wrongness of depriving a victim of a given amount of consciousness is greater the nearer the victim is to the end of her conscious life.<sup>31</sup> More precisely, the authors defend the following principle:

The wrongness of rendering someone unconscious corresponds with the proportion of the victim's *remaining* conscious life that is thereby removed, other things being the same.<sup>32</sup>

---

<sup>31</sup> Cohen and Luck (2009), pp. 396–401.

<sup>32</sup> *Ibid.*, p. 399 (emphasis added).

Their argument for that principle involves imagining two alien species. Members of the first species (call them the “short-livers”) always live for just two years, while members of the second species (call them the “long-livers”) live for 1,000 years. The authors imagine coming across a newborn member of each species.<sup>33</sup> As they point out, it seems more seriously wrongful to render the newborn short-liver unconscious for one year—half its remaining life—than to knock the newborn long-liver unconscious for a year, leaving it to experience another 999 years when it awakens.

Cohen and Luck’s principle gets the right answer in this case, since the short-liver is deprived of a greater proportion of its remaining conscious life (one year of two) than the long-liver is of his (one year of a thousand). But there is a simple variation on their case that shows that their principle has an implausible implication. The principle they defend would imply that it would be no more wrong to knock unconscious a short-liver who has been alive for one year than a long-liver who has already lived for 999 years, for in either case one removes the entirety of the victim’s remaining conscious life. But that is hard to believe. When we knock the short-liver unconscious we deprive it of half its total conscious life, whereas when we do so to a long-liver we reduce its conscious life by just one one-thousandth.

---

<sup>33</sup> Assume that members of both species are born full persons.

It might be thought that the *ceteris paribus* clause of Cohen and Luck's principle is meant to factor out differences in how much conscious life the victim has so far enjoyed. But their conclusion is that a person's age has no bearing on the wrongness of killing her, so their argument cannot go through if cases in which the victims differ in the amount of conscious life they have lived are not covered by their principle.

Cohen and Luck's newborn aliens case relies on the intuition that it matters, when knocking a person unconscious, how much future conscious life she loses out on. But as the variation I have just offered shows, it also seems to matter for its wrongness how much conscious life a person has so far enjoyed. To account for both factors requires a middle ground. One plausible candidate is that the wrongness of knocking a person unconscious for some period corresponds not with that period's proportion to the victim's remaining conscious life, but rather with its proportion to the total lifetime conscious experience she would otherwise have had.

That new principle would be:

The wrongness of knocking someone unconscious corresponds with the proportion she loses of the *total* amount of conscious life she would otherwise have enjoyed, other things being the same.



This principle gets the right answer in both of the two short- and long-lived alien cases. It also entails (B), which compares equally long deprivations of consciousness at different times within equally long conscious lives.

I now turn to a positive argument for (B). In it, Calvin and Dylan again make an appearance. Suppose that I possess two drugs: a *fast-acting drug* that immediately knocks its taker unconscious for one month and a *delayed-acting drug* that has no effect until one month before its taker's death, when it will knock him unconscious for that final month.<sup>34</sup> Consider, first, the following case:

*Case One:* I secretly slip the fast-acting drug into the coffee of the 999-month-old Dylan and the delayed-acting drug into the coffee of the 500-month-old Calvin.

It is not plausible that what I do to either man in Case One is more wrong than what I do to the other. I deprive each man of the very same amount of experience, at the very same location in his life, and in the very same manner. It is true that the effect of my action takes place immediately for Dylan

---

<sup>34</sup> McMahan considers an objection to the Equal Wrongness Thesis that involves fast- and slow-acting poisons—causing either an immediate or a delayed death—which is the inspiration for these cases (2002), p. 248.

but is delayed by nearly forty years for Calvin. But it seems to me that mere temporal distance between an act and its effect could not be morally significant in itself.

Next, consider:

*Case Two:* Calvin has an identical twin, Caleb. I slip the delayed-acting drug into Calvin's coffee, and the fast-acting drug into Caleb's.

I do not think it is plausible that the acts in Case Two are wrong to different degrees. To dispute that claim, one cannot appeal to the proportion of remaining conscious life of which each man is deprived, for it is the same. The salient difference seems to be that only Calvin loses the final month of his conscious life. But since it is not the last month he has left, and since being deprived of that month of consciousness does not, as I have argued in the last chapter, constitute a greater non-comparative loss to him, it is hard to see why the mere fact that the month is at the end of his life should make its deprivation more wrongful.

Taken together, these two equivalences—that in Case One and that in Case Two—show that it is morally irrelevant whether, when I deprive

someone of a month of conscious experience, I do so at the end of his life or its middle. If giving the fast-acting drug to the 999-month-old Dylan is morally equivalent to giving the delayed-acting drug to the 500-month-old Calvin (Case One), and the latter act is morally equivalent to giving the fast-acting drug to the 500-month-old Caleb (Case Two), then giving the fast-acting drug to the 999-month-old Dylan is morally equivalent to giving the 500-month-old Caleb the fast-acting drug. That result, of course, is the same as the case with which we began this section: knocking out Calvin now, or knocking out Dylan now.<sup>35</sup>

### 3.5. Conclusion

It is now possible to bring together the pieces of my argument. The first two premises were

(A) It is as wrong to deprive someone of a certain amount of conscious life by knocking her unconscious until her death as it is to deprive someone of the same amount of conscious life by killing her, other things being the same.

---

<sup>35</sup> The same argument can be given for longer deprivations of consciousness than one month. But even this specific equivalence is sufficient to undermine the Equal Wrongness Thesis.

And

(B) It is as wrong to deprive someone of a certain amount of conscious life by knocking her temporarily unconscious in the middle of her life as it is to deprive someone of the same amount of conscious life by knocking her unconscious until her death, other things being the same.

These two premises entail

(C) It is as wrong to deprive someone of a certain amount of conscious life by knocking her temporarily unconscious in the middle of her life as it is to deprive her of the same amount of conscious life by killing her, other things being the same.

(C) is sufficient to defeat the Equal Wrongness Thesis. Suppose we take the period of unconsciousness to be one month. We would then hold that knocking someone temporarily unconscious for one month in the middle of her life is as wrong as killing someone who has one month left to live. The Equal Wrongness Thesis tells us that killing a person who has one month left to live is as wrong as killing a person with fifty good years to live. Hence, the conjunction of the Equal Wrongness Thesis and (C) entails that, other things

being equal, knocking a person unconscious for a month in the middle of her life is just as wrong as killing a person with fifty good years ahead of her. This result is absurd. Whatever considerations might justify knocking someone out for a month in the middle of her life, those considerations cannot be as strong as the ones required to justify killing a person with fifty good years left to live. On pain of absurdity, therefore, we should reject the Equal Wrongness Thesis.

## Chapter 4: Non-Compensable Harms

It is more or less uncontroversial that when we harm someone through wrongful conduct we incur an obligation to compensate her. But sometimes compensation seems impossible: when the victim has been killed, for example. Other times, only partial compensation is possible. There may be little one can do for the person one has made permanently quadriplegic; or that a state can do for a person who has spent most of his life in prison on a wrongful conviction. Call any harm whose victim cannot be fully compensated by its perpetrator a “non-compensable harm”.

Many authors have discussed the duties offenders have to compensate their victims. But to my knowledge few have asked how these duties change when full compensation is impossible. This chapter is an attempt to explore this largely unexamined issue. I will try to show the many interesting questions it raises, whose answers have important and far-reaching consequences for both theoretical and applied ethics.

It will be useful to sort non-compensable harms into two categories. The first category comprises harms whose perpetrators cannot, at least beyond a certain point, compensate their victims at all. In these cases, there is either nothing whatsoever that the perpetrator can do to improve the victim's condition, or else nothing she can do once some threshold below full compensation is reached. On the controversial assumption that the living cannot improve the lives of the dead, the clearest example of this sort of non-compensable harm is one in which the victim has been killed.

The other category of non-compensable harms comprises those harms whose victims, while unable to be fully compensated, can nonetheless always receive *more* compensation. The victim's capacity for compensation is "asymptotic": the perpetrator can get closer to full compensation, but can never reach it. In many more typical cases, it is natural to assume that the marginal cost to a perpetrator of compensating his victim is roughly equal to the marginal benefit to the victim of being compensated. Non-compensable harms can involve cases where the cost of compensating and the portion of the harm compensation offsets differ dramatically.

The bulk of this chapter is divided into two long sections, each corresponding to one of these two categories of non-compensable harm. In section

4.1, I will defend a position I call the “Non-Compensable Harms Thesis”, which holds that to the extent that it is impossible for the perpetrator of a non-compensable harm to compensate his victim further, he acquires a new undirected duty to promote the impartial good.

In section 4.2, I will explore the nature and extent of the duties of perpetrators who cannot compensate their victims fully but can, eventually at a very high cost, still get closer to full compensation. These cases raise the issue of the limits of proportionality in corrective justice. Because full compensation is impossible, we have to ask how much compensation—if any—is enough.

#### **4.1. How to make up for what we cannot make up for**

It may understandably be wondered whether there really are perpetrators who cannot further compensate their victims. After all, isn't there always *something* that can be done to improve the life a quadriplegic, however little the improvement may be? Couldn't she always have better therapeutic treatment, more advanced medical care, a more comfortable bed? Indeed, on some popular theories of death and wellbeing, victims can be compensated



posthumously, for example by satisfying interests they once had in what happens after their deaths. If so, then it seems there is some room to compensate even the person one has killed.

Even if this complaint were on target, there would still remain the important question of how far a perpetrator must go in these cases. That is a question I will consider in the chapter's second main part, section 4.2. For even assuming posthumous compensation is possible, few of the perpetrators of lethal harms would be able to *fully* compensate their victims. The vast majority of most people's interests require for their satisfaction that they remain alive, and, moreover, it is hard to believe that there is anything one could do for a victim posthumously to fully make up for the loss of, say, decades of happy life.

Yet I am convinced that the category of non-compensable harms whose perpetrators cannot compensate their victims *at all* is both philosophically significant and widely applicable. For one thing, there is the obvious fact that on other popular theories of wellbeing, like hedonism, posthumous compensation *is* impossible. Conditional on those theories' truth, it is impossible to compensate a person one has killed. Since lethal harming makes up an important portion of wrongful harming in general, it is important to examine

the duties perpetrators of lethal harms have if these theories of wellbeing are true.

Secondly, my concern in this section is with is the perpetrator's duty in cases of non-compensable harm. While it is probably true that a victim can always in principle be compensated, it is not always true that her perpetrator will be able to provide that full compensation. The perpetrator might be totally indigent, or simply irredeemably ignorant of where her victim is located, or of how to even begin to benefit her. Even if posthumous compensation is possible, most of the interests of the victims of large-scale and indiscriminate killings are probably unknowable in practice.

Finally, there are cases in which the receipt of compensation from a perpetrator is strongly contrary to the wishes of the victim. Some victims of violence may simply want nothing whatever to do with their attackers, including the receipt compensation from them. One thing this might mean is that it is actually impossible for the perpetrator to compensate the victim, since any attempt at improving the victim's condition would in fact leave her worse off by virtue of being contrary to her interests. But even if compensation were not impossible in such cases, perpetrators ought to respect the wishes of their victims even at the cost of not fulfilling their *prima facie* du-

ties to compensate them. I have defined a non-compensable harm as one whose victim cannot be fully compensated, but we might then understand “cannot” in a looser sense that means “cannot permissibly”. It is interesting and important to learn what duties a perpetrator has whose victim rejects compensation, and, as I hope will be clear shortly, these cases raise the same puzzles, and have the same answers, as those in which full compensation is actually impossible.

I take myself to have defended the significance of a general treatment the duties of perpetrators who cannot do anything more to compensate their victims. I will therefore proceed to my main topic of inquiry, which is what those duties are.

Let me begin by stating explicitly the thesis I will defend in this section:

*The Non-Compensable Harms Thesis:* When one person wrongfully harms another, she acquires a duty to promote the impartial good in proportion to the amount of the harm that cannot be made up for by compensating the victim.

Before pressing on, several elements of this thesis require some brief explanation. For the purposes of this chapter, I will assume a counterfactual-comparative account of harming, according to which an act harms a person just in

case, and to the extent that, it leaves her worse off than she would have been had the act not been performed. Correspondingly, I will say that an act compensates someone for a harm to the extent that it leaves in a state that is as good as that in which she would have been, had that harm not been done. The counterfactual-comparative account is not, of course, the only conception of harm on the market. My arguments are compatible with alternative conceptions of harm, however, because they will still require a corresponding account of compensation, and any plausible account of harm will allow for harms that are not fully compensable.

By “the promotion of the impartial good” I mean the promotion of that which has intrinsic value, including the elimination of intrinsic disvalue. For the most part, I will remain neutral on what sorts of thing are intrinsically valuable. It is important to stress, however, that by “impartial” I do not mean non-person-affecting. Although it might be that the world is improved when it contains more knowledge or beauty, on any plausible axiology it is improved when its inhabitants enjoy more wellbeing or less suffering. Central to my argument is the assumption that possible worlds or states of affairs can be ordered in terms of better and worse overall, where this notion is conceptually independent of their being better or worse for a particular individual,

and where one state of affairs is better than another just in case it contains more intrinsic value on balance. Although widely accepted, this view does have its detractors<sup>1</sup>, and those who deny that we can make sense of one state of affairs being better or worse *simpliciter* than another will not find my arguments persuasive.

#### 4.1.1 THE NON-COMPENSABLE HARMS THESIS

I turn now to a defence of the Non-Compensable Harms Thesis. Suppose Jones wrongfully breaks Smith's finger. Jones has an obligation to fully compensate Smith for this by, for example, paying his medical bills and offsetting any pain and suffering or lost wages. But suppose Jones had instead chosen to kill Smith. Then there would be no way for Jones to compensate Smith.<sup>2</sup> So by "ought" implies "can", Jones would have no duty to compensate Smith. So, it seems that by committing a far more egregious act, Jones can ensure that he has no obligation to compensate his victim, instead of the costly one of compensating him for a broken finger.

---

<sup>1</sup> For example Taurek (1977); Foot (1985).

<sup>2</sup> I here assume that compensation is impossible because Smith is dead, but the same point applies if compensating Smith is impossible or impermissible for any of the other reasons listed earlier.

Of course, by killing Smith, Jones may owe compensation to Smith's family or loved ones because his action also harmed them. And he may be required to submit to legal punishment on, for instance, grounds of fairness or political obligation. But these duties obtain in virtue of facts that are connected only contingently to the original harm. There is no reason to think they will always obtain, or that when they do their strength will be greater than that of the compensatory duty that would have arisen had the less harmful act been performed instead, so we cannot appeal to them to avoid the result. For those who prefer tidiness, we can add that Smith has no family or friends (or even that Jones killed them, too) and that the offence takes place in the state of nature.

The same phenomenon arises when it is *full* compensation that is impossible. Suppose Wright wrongfully causes Edwards to be moderately disabled for a year. Assume that, while doing so would be onerous and expensive, it is possible for Wright to fully compensate Edwards. But suppose Wright had instead chosen to assault Edwards in a way that made him quadriplegic for life. There is still something we can do to improve the lives of the permanently quadriplegic. But it is unlikely that we can raise their wellbeing to the level it would have been at had they not been injured, and let me

stipulate for good measure that in the case of Edwards there is a point (relatively easily reached) at which there is *nothing* more that can be done for him. Here, again, it seems we must accept that by perpetrating a much more serious harm a person can ensure that he incurs a far weaker obligation to his victim.

In “A Difficulty Concerning Compensation”, Saul Smilansky considers several cases like these and concludes that they reveal an intractable “paradox of morality”.<sup>3</sup> His reason for thinking these cases paradoxical seems to be that they are incompatible with the following principle:

(A) Other things being equal, the strength of the obligation one incurs in virtue of wrongfully harming another correlates with the degree of harm done.

As he puts it,

if [a perpetrator’s] burden can be lightened (or even dismissed altogether) because of the increased harm to the victims, this situation is manifestly morally unsatisfactory, and indeed absurd.<sup>4</sup>

---

<sup>3</sup> Smilansky (2013).

<sup>4</sup> *Ibid.*, p. 333.

Seeing no alternative, Smilansky reluctantly concludes that we must abandon the intuitively attractive view represented by (A).<sup>5</sup> Instead of rejecting (A), however, we could deny that

(B) The obligations one incurs in virtue of wrongfully harming another are exhausted by the obligations one has *to her*.

We could, instead, say that a perpetrator incurs an obligation that tracks the extent to which he has harmed the victim, rather than the extent to which it is possible for him to compensate her. Then the situation would no longer be absurd. In doing so, we might appeal to

*The Non-Compensable Harms Thesis:* When one person wrongfully harms another, she acquires a duty to promote the impartial good in proportion to the amount of the harm that cannot be made up for by compensating the victim.

Killing a person is, we can suppose, much more harmful than breaking his finger. So, the Non-Compensable Harms Thesis implies that Jones's obligation in virtue of killing Smith *is* more demanding than his obligation in the world where he merely breaks Smith's finger—it is just that the non-com-

---

<sup>5</sup> Ibid., p. 336.



pensable portion must be made up for elsewhere. Similar remarks apply when it is full compensation that is impossible. When Wright causes Edwards permanent full paralysis, the thesis implies that his overall obligation is more demanding than it would have been, had he instead caused him temporary moderate disability.

Below I will consider a number of objections to the Non-Compensable Harms Thesis in general and to the argument just made in favour of it in particular. But it is worth registering that the thesis already enjoys at least some intuitive appeal. Many people, I suspect, would find it hard to accept that agent-neutral morality requires us to engage in highly demanding or risky activities to assist the less fortunate by, say, devoting several years of our lives to volunteering in the Peace Corps. But it is much easier to believe that Jones, who has killed an innocent person, could be morally required to do something of the kind.

To be sure, one can uphold the general claim that the perpetrator of a wrongful non-compensable harm can incur an obligation that is not directed *to* her victim, while denying that the content of that obligation is the promotion of the good. Two alternative ways of filling out that content are that it involves the duty to suffer for one's wrongdoing, and that it involves the duty

to repent for it. The view that people should suffer for their misdeeds has some pedigree by way of the religious doctrine that evildoers should atone for their sins by engaging in penance, but it is not a serious rival to the Non-Compensable Harms Thesis for several reasons. First, imagine Jones ends up enjoying his experience in the Peace Corps and becomes a happier and better person for it. We would not, I think, judge that he has failed to discharge the duty he incurred by killing Smith. Second, although many people believe that wrongdoers deserve to suffer, it is harder to believe that they could have a moral duty to suffer, such that it would be impermissible for them not to inflict suffering upon themselves.

Finally, even if we do maintain that the perpetrator of a non-compensable harm is required to suffer for it, in fact it is not obvious that this view supports the claim that the content of the obligation is suffering rather than the promotion of the good. For, many people also consider the suffering of evildoers to be an intrinsically good state of affairs. Thus, even if perpetrators of non-compensable harms have a duty to suffer, this can be understood as the view that they have a special duty to promote the good. Because those who believe that it is good when wrongdoers suffer believe that the suffering of any wrongdoer makes things better impartially, it may seem that together

with the Non-Compensable Harms Thesis this would imply, unpalatably, that the perpetrator of a non-compensable harm can discharge the duty she incurs by inflicting suffering on deserving others rather than on herself. But that is not so, for it is independently plausible that if a person has a duty to bring about some outcome then she ought to bear the costs of bringing it about, rather than impose them on others.

A second possibility for filling in the duty acquired by the perpetrator of a non-compensable harm is that wrongdoers have a duty to repent. There are at least two problems with this view. The first is that it is not clear that the remorse one should feel for a wrongful action tracks its harmfulness rather than other factors like how malicious or callous it was. Second, whether one feels remorse is generally not under one's control. The racist perpetrator of a lethal hate crime may feel no remorse; indeed, he may regret not having done more harm. Because one can have a duty to feel or express remorse only if one is able to do so, the present view would imply that these wrongdoers do not incur any obligation, while those who are capable of sincere regret do. I think we should conclude that while we may justifiably think less of people who do not feel or express remorse for their misconduct, wrongdoers themselves do not have a duty to do so.

The Non-Compensable Harms Thesis avoids the morally absurd result that by doing more harm a wrongdoer can ensure that she acquires a less demanding obligation, and it does so in a way that is more plausible than rival ways of filling out the content of that duty. It also enjoys some independent intuitive appeal. These considerations provide some reason to accept it. But they are not sufficient to justify the thesis for three reasons:

i) Even if we can explain why the promotion of the impartial good can make up for non-compensable harms, we need an explanation for why *compensable* harms cannot be made up for in that way. For I accept that when it *is* possible to compensate a victim for a wrongful harm done to her, no amount of good done elsewhere can count as discharging that duty.

ii) We lack an account of the internal connection between the object of the moral violation (a particular individual) and the object of the duty it supposedly generates (the world in general). When it comes to the duty to benefit a victim, in contrast, that connection is clear. My victim has a claim *against* me, and I have a corresponding duty *to* her. But no similar link seems forthcoming for the move from harming a person to an obligation to do something that will likely have no effect on her. By implying that one person's loss in wellbeing can be made up for by increasing value elsewhere, the

Non-Compensable Harms Thesis may be thought to disregard the separateness of persons.

iii) We lack an account of why the content of the duty is the promotion of the good in particular. Again, in the case of compensation the analogous explanation is more or less plain. Compensating a victim is the way to rectify the harm one has done to her. But it is not yet clear how making the world better in ways that have no effect on one's victim would go any way toward rectifying a wrongful harm done to her.

The answer to all three issues, I shall argue, proceeds from the following premiss:

(C) When a person wrongfully makes the world a worse place, one has an obligation to add value to the world to the extent that one has wrongfully removed it.

The claim is not, I hope, too controversial. The reasoning is straightforward: when one does wrong, one has a duty to reset things as far as possible to the way they would have been if not for one's wrongdoing. Indeed, this seems to be one rationale for the more familiar obligation to compensate victims in the first place. Joel Feinberg writes that the reason a perpetrator has a duty to

compensate his victim for the wrongful harm he has done to her is “not only to repair the damage but also to ‘restore the moral equilibrium’”.<sup>6</sup> According to Feinberg, compensation “sets things straight” and calls for “the redressing of the moral balance or the restoring of the *status quo ante culpum*”.<sup>7</sup> It is clear what these metaphors are meant to capture. The point of compensation is to make things as close as possible, normatively speaking, to the way they were (or would have been) in the absence of the wrongful harm. Both reasons Feinberg cites—repairing the damage and restoring the moral equilibrium—apply *mutatis mutandis* to the redress of wrongdoing that makes the world worse impartially. So if, as seems plausible, Feinberg is right about one rationale behind the duty of offenders to compensate their victims, we should also accept that one has an obligation to restore to the world what impartial value one’s wrongdoing has eliminated from it.

Let me now relate in more detail how (C) rationalizes the Non-Compensable Harms Thesis, and how it resolves problems (i)–(iii) raised earlier. Some ways of making the world worse seem to be non-person-affecting, in that the reason why they make it worse is not any harmful effects on particu-

---

<sup>6</sup> Feinberg (1970), p. 74. Feinberg himself uses the term “rectification”, reserving “compensation” to refer to the redressing of only those harms whose infliction was not wrongful.

<sup>7</sup> *Ibid.*, p. 76.

lar individuals. At least on their orthodox interpretation, non-identity cases are paradigmatic examples. But one can also make the world worse by behaving in a way that is person-affecting. A world in which one more wrongful killing occurs is worse for someone, but on any plausible axiology it is also worse impartially. Therefore, (C) implies that when one person wrongfully harms another she incurs not one but two distinct duties, deriving, respectively, from the personal and impartial aspects of her transgression. When and to the extent that a wrongful harm is compensable, the victim has a claim against the perpetrator for the full amount of that compensation. The perpetrator's duty to satisfy this claim is independent of her duty to restore impartial value to the world. However, by fully compensating her victim, a perpetrator also makes the world as good, or nearly as good, as it would have been had the harm never been done. Accordingly, to the extent that one fulfils an obligation to compensate a victim, the duty to restore impartial value to the world is discharged *incidentally*. To the extent that a harm is non-compensable, the duty to restore impartial value remains, and it is this remainder that gives rise to the perpetrator's obligation to promote the good.

This explains why, as observed in (i), one cannot rectify a *compensable* harm by promoting the general good. That worry arose, I think, from the as-

sumption that the Non-Compensable Harms Thesis implies that when a victim cannot be compensated, the duty to compensate transforms into a duty to promote the impartial good. But, rather, the imperative to restore to the world the value one has wrongfully eliminated from it is independent of whatever obligations one has *to* one's victims. So it makes sense that if one has an undischarged personal duty, promoting the impartial good does not count toward its fulfilment.

Because it is the impartial aspect that generates the obligation to promote the good, this rationale for the Non-Compensable Harms Thesis avoids the charge posed in (ii)—namely, that by allowing that one person's loss in wellbeing can be made up for elsewhere, the thesis fails to respect the separateness of persons.

Because the obligation is given by the duty to restore the impartial value one has wrongfully eliminated, the account explains why the content of the obligation is the promotion of the good, thereby answering (iii).

Furthermore, because the good required is proportional to the impartial disvalue of the non-compensable amount of the harm, it accords with (A), the principle that the obligations a perpetrator incurs in virtue of wrongfully harming another correlate with the amount of harm she has done.



(C) provides a normatively significant underlying rationale for the Non-Compensable Harms Thesis. Before concluding, however, I should address an issues it raises. It may be objected that it lets some wrongdoers off the hook too easily. By some estimates, the current cost of saving a life in the more miserable parts of our planet is just several thousand pounds. The argument so far may therefore seem to imply that an affluent murderer can make amends by donating a relatively negligible sum to an effective charity.

There are two things to notice about this objection. The first is that it depends on the facts that the perpetrator is well off and that there are many people alive who are very badly off and whose lives can be greatly improved at little cost to the wealthy. These facts are both contingent and almost certainly reflect pre-existing unfairness. The second thing to note is that the objection arises even for ordinary compensatory duties. How demanding it is for a perpetrator to compensate her victim almost always depends on how well off she is. It seems plausible, therefore, that what is problematic is not the Non-Compensable Harms Thesis, but rather the current distribution of those factors which determine how easy it is to add value to the world.

#### 4.1.2. TWO PRACTICAL IMPLICATIONS

I have argued that the perpetrator of a wrongful, non-compensable harm has an obligation to promote the impartial good that is proportional to the amount of the harm that is not compensable. If we have committed wrongs in the past, then we may learn from this that we have a duty we did not know we had. I now want to briefly discuss two more specific implications of the Non-Compensable Harms Thesis.

##### *A. The Justification of punishment*

One of the more divisive issues in legal philosophy is the justification of criminal punishment. One family of justifications for punishment is retributivist, holding, roughly, that punishing wrongdoers is justified because they deserve to suffer for what they have done. Those who reject retributivist justifications typically adopt forward-looking rationales for state punishment, perhaps most prominent among them that punishment can serve to deter others from misconduct. But, at least for non-consequentialists, the deterrence justification has at least one implication that is difficult to swallow: it can recommend punishing the innocent.

In an influential recent book, Victor Tadros has developed an alternative justification for punishment, which he calls the “Duty View”.<sup>8</sup> The Duty View aims to show how punishment can be warranted because of its deterrent effects without implying that punishing the innocent could be justified on those same grounds. In brief, Tadros argues that punishment can be justified when it would deter others from crime, but the reason it is justified is not consequentialist in nature but rather that wrongdoers owe their victims compensation. According to Tadros, that compensation is often best provided by protecting their victims from future harms, which in turn is best achieved by enduring sufficient harm to deter others from crime.

Of course, it is highly improbable that the deterrence provided by a given criminal’s incarceration will benefit his own victim: one more incarceration is unlikely to deter prospective offenders to any appreciable degree and, even if it did, it is unlikely that the minute reduction in crime that resulted from would end up benefitting the particular individuals to whom the offender owes compensation. To answer this objection, Tadros argues that criminals have a duty to “pool” the compensatory obligations they owe to their victims. By all undergoing punishment, they together make it the case that each of their victims is expected to enjoy a safer, better life.

---

<sup>8</sup> Tadros (2011).

A major hurdle faced by the Duty View, which Tadros acknowledges, is that it appears unable to explain why punishing offenders is justifiable when doing so will not provide *any* benefit to their particular victims. A murder victim, for example, seems to have nothing to gain from a reduction in criminal activity. It is hard to see why, in these cases, the offender would be required to submit to punishment, since even by doing so together with many others he will not go any way toward benefitting his own victim.

I shall not attempt to assess Tadros's own answer to this objection, which is complex and turns on the "permissions and obligations that citizens have to transfer their right to protection to other citizens".<sup>9</sup> I want to note, instead, that the Non-Compensable Harms Thesis can provide an easier route to his desired conclusion.

Tadros's account has this difficulty because it appeals only to directed compensatory obligations—those wrongdoers have to *their* victims. It therefore has trouble accounting for cases where the victim of a crime is not in a state to be benefitted. The Non-Compensable Harms Thesis holds that in just these cases, an offender has an additional duty to promote the good. So, if the perpetrator of a non-compensable harm is best able to promote the good by submitting to state punishment to deter others from crime, then it

---

<sup>9</sup> Ibid., p. 281.

can provide the necessary justification for punishing offenders whose own victims have nothing to gain.<sup>10</sup>

*B. Non-compensable harms and the state*

Many if not most modern states have committed historical atrocities. Many of the victims of those atrocities cannot be fully compensated, either because the atrocities involved killing (as in the case of contemporary unjust wars) or else because the victims have since died (as in the case of slavery). The Non-Compensable Harms Thesis may imply that these states have a duty to make up for the wrongful harms they have committed in the past by way of promoting the good. Because my aim here is only to draw out a potential implication of the thesis, I sidestep two difficult further questions: the extent to which the individual citizens of a state can bear responsibility for its wrongdoing, and whether, in the case in which that wrongdoing is in the more distant past, present-day citizens can have the responsibility to make up for the misdeeds of their forebears.

Many people believe that affluent states already have some duty to assist the unfortunate people in this world. They accept that the members of

---

<sup>10</sup> Admittedly, this depends (as does Tadros's justification) on the duty in question being enforceable.

those states ought to give up some of their wealth, collected by taxes, to foreign aid. But this proposition is typically defended, implicitly or explicitly, with one or more of what Robert Nozick called “current time-slice principles”—principles like equality, sufficiency, total utility, or the Difference Principle—as contrasted to historical principles, which appeal to and depend upon what rights have been violated in the past. The Non-Compensable Harms Thesis may imply that these states have additional duties of assistance on the basis of the wrongs they have committed in the past.

To make the point vivid, suppose for the sake of argument that the current time-slice-given duty of assistance that those in affluent states have extends only so far as is required to bring others up to a level of wellbeing that is by some measure *sufficient*.<sup>11</sup> If an affluent state is also guilty of significant non-compensable past wrongdoing, as many are, the Non-Compensable Harms Thesis may suggest that the overall duty of assistance that must be borne by its citizens is higher than that level. And insofar as the historical wrongdoings in question are gargantuan—mass slavery and genocide, for instance—this duty may be *much* stronger than what is required by the correct current time-slice principle.

---

<sup>11</sup> Frankfurt (1987; 2000); Miller (1995); Wiggins (1991).

## 4.2. Compensation and proportionality

The first part of this chapter concerned those cases in which the perpetrator of a non-compensable harm cannot or should not further compensate his victim. Those are the cases to which the Non-Compensable Harms Thesis applies. In this, the second part, I turn to those non-compensable harms whose perpetrators both can and may compensate their victims further. In these cases, the victim's capacity for compensation by the perpetrator is asymptotic: she can always be compensated more, but never fully. Because full compensation is not possible, we need to ask how much compensation, if any, can be enough.

I will attempt to show how this second category of non-compensable harms provides a way into the under-explored issue of the limits of proportionality in the liability to compensate. The question of what limits proportionality places on liability to compensate does not arise that often, because typically the answer to "When has the perpetrator contributed enough compensation?" is simply, "When the victim has been fully compensated." But when it is either strictly impossible or else prohibitively expensive for a perpetrator to fully compensate his victim for a wrongful harm, we need to ask how much compensation he *is* liable to provide. In addition to its more direct

implications for the limits of the duty to compensate in a range of such cases, I shall show how the conditions under which a perpetrator has compensated enough have important consequences for proportionality in the ethics of self-defence.

#### 4.2.1. PROPORTIONATE COMPENSATION

No one but an overzealous pacifist would deny that we may sometimes use force to protect ourselves from those who threaten harm. That said, philosophers typically recognise two broad constraints on the right to harm others in self-defence. The first is that one may use only the minimum amount of force necessary to avert a given threat; there must be no way to save oneself from the same probability of the same amount of harm by inflicting a lesser harm on others. This is called the “necessity” condition. But sometimes one lacks the right to harm someone even when doing so is the only way to avert the threat she poses—it is impermissible for me to kill someone just to stop her from breaking my little finger, for example. Philosophers therefore recognise that even a defensive harm which satisfies the necessity condition can be impermissible if the harmful effect is by some



measure excessive in light of the harm it would prevent. This is known as the “proportionality” condition. Of course, sometimes a given use of defensive force is excessive not because of its effect on the attacker, but because of its effects on third parties. The question of how much harm is proportionate to inflict on the threatener is that of so-called “narrow” proportionality—proportionality as it applies to the effects of an act on perpetrators and victims.<sup>12</sup>

Just as a victim lacks the right to inflict more harm than is necessary to defend herself from a given threat, so a victim lacks the right to extract a greater cost than is necessary for her to be *compensated* to a given degree. Even if my total-loss car is the result of your reckless driving, I cannot justifiably insist that you buy me an identical car from the more expensive of two dealerships. More importantly for our purposes, there is a limit to the amount of compensation it is *proportionate* to extract in compensation for a harm. If for some reason the only way for you to recover damages for your broken fence requires me to sell my kidney, you would probably lack a right that I do so.

---

<sup>12</sup> For a useful recent overview of the literature on necessity and proportionality in defensive harming, see McMahan (2018).

Before moving on I should make a few preliminary remarks. First, I will here conceive proportionality as “internal” to liability, such that by definition a person cannot be liable to pay compensation that is disproportionate.<sup>13</sup> Second, I shall assume that a victim has a right against a perpetrator to be compensated for a given harm if and only if, and to the extent that, the perpetrator is liable to compensate her for it. Third, since I will focus mainly on proportionality, I shall often simply take for granted that the necessity condition has been met. I will suppose, in other words, that for a given compensatory claim, it is not possible for the perpetrator to give the same benefit to the victim at a lesser cost to himself. Finally, because I am concerned primarily with the limits that proportionality places on the compensation for which we are liable—rather than the conditions for being liable in the first place—I will often restrict my discussion to cases of fully culpable harm, for in these cases it is straightforward that the perpetrator can be liable to pay at least some compensation.

It is easy to overlook the fact that compensation has its own proportionality constraint for at least two reasons. First, as I noted earlier, in many cases of corrective justice we take for granted that if a perpetrator is liable to

---

<sup>13</sup> For discussion of the parallel view in the context of defensive harm, see McMahan (2009), pp. 8–10.

compensate his victim at all, then he is liable to compensate her fully if he can do so.

Second, unlike cases of defensive harming, in which the size of the harm necessary to stop an aggressor often differs greatly from the size of the harm he threatens, in cases of compensation we may take for granted that the magnitude of the cost to a perpetrator of providing compensation is roughly equivalent to the benefit to the victim of receiving it. If P unlawfully does £1,000 of damages to V's property, by compensating V by only £700 instead of the full £1,000, P makes it the case that V ends up bearing an additional financial cost of £300. And it seems that, for a given fixed cost, it should be the perpetrator, not the victim, who bears it. But of course the benefit to P of £300 may well exceed the benefit to V of £300, in which case the cost to P of fully compensating V may be considerably greater than the cost to V of being less-than-fully compensated. Cases of non-compensable harms illuminate this difference. In these cases, at some point, the marginal wellbeing-cost to the perpetrator of paying compensation will far exceed the cost to the victim of being deprived of additional compensation. It is primarily when these two variables diverge that full compensation can seem excessive.

Much later, in subsection 4.2.5, I will offer three arguments to the effect that proportionality in the liability to compensate is nothing more or less than the *ex post* flip-side of proportionality in the liability to suffer defensive harm. Hence, I claim, in every case in which a perpetrator is liable to suffer some measure of defensive harm to avert a threat he poses, he is also liable to the same cost if doing so is necessary to compensate his victim should the harm eventuate—and vice versa. I call this thesis the *substantive symmetry* of the liability to defensive harm and the liability to compensate. But whatever one thinks about the substantive symmetry, it is undeniable that there is a certain conceptual similarity between the two notions. And it is worth stopping now to show that, just by acknowledging that there is also a proportionality constraint on compensation, we are already well positioned to explain away cases that have sometimes been thought to be counterexamples to the substantive symmetry.<sup>14</sup> Consider:

*Feinberg's Cabin:* A backpacker is trapped in a blizzard. He breaks into an unoccupied cabin and burns the furniture there to keep himself from freezing to death.<sup>15</sup>

---

<sup>14</sup> McMahan (2008); Steinhoff (2016). Steinhoff focuses on yet a third case in the literature, which involves a diabetic in need of insulin.

<sup>15</sup> Feinberg (1978), p. 102.

*Vincent v. Lake Erie*: Lake Erie Transport Company (defendant) tied its ship to Vincent's (plaintiff) dock in the middle of a storm in order to safeguard the ship and crew. Doing so damaged the dock.<sup>16</sup>

Almost everyone would accept that in both cases the offenders acted permissibly. The transport company was later ruled to have been acting under the privilege of private necessity, as would the backpacker have rightly been. More importantly for our purposes, it is also true that it would have been impermissible for the homeowner to use defensive force against the backpacker to prevent him from entering his cabin, could he have done so, and it would have been impermissible for Vincent to forcibly prevent the transport company from mooring at his dock.

Yet although it would have been impermissible for the owners to prevent the offenders from trespassing, it is also reasonably clear that the offenders incurred an obligation to compensate the injured parties after the fact by acting as they did. Feinberg wrote of his example that "almost everyone would agree that [the backpacker] owes compensation to the homeowner"<sup>17</sup>,

---

<sup>16</sup> *Vincent v. Lake Erie Transportation Co.* 109 Minn. 456, 124 N.W. 221 (1910).

<sup>17</sup> Feinberg (1978). p. 102.

and Vincent, whose dock was damaged, later sued the transport company and was awarded \$500 in damages.<sup>18</sup>

Because in these cases the threatened party lacks a right to self-defence against a given harm *ex ante*, but retains a right to receive compensation for the same harm *ex post*, these examples may seem to challenge the substantive symmetry between the liability to defensive harm and the liability to compensate. Thus McMahan writes that Feinberg's case presents "[a]n obvious example in which the one form of liability does not entail the other".<sup>19</sup> However, the correct defensive analogue of the homeowner collecting compensation is not preventing the endangered backpacker from entering his cabin, but rather forcing the backpacker to pay upfront for the damage that will be done to his property. Likewise, the correct defensive analogue of Vincent collecting compensation would not be preventing the ship from mooring, but rather doing \$500 worth of damage to it to spare the damage to his dock. Provided these pre-emptive acts would impose no additional danger on the offenders—as collecting compensation after the fact would not—neither defensive action is disproportionate, so in fact there is no obvious asymmetry here.

---

<sup>18</sup> *Vincent v. Lake Erie Transportation Co.* 109 Minn. 456, 124 N.W. 221 (1910).

<sup>19</sup> McMahan (2008), p. 233.

We can make this point from the opposite direction, too. The compensatory analogue of preventing the backpacker from taking refuge in his cabin would be the homeowner preventing the backpacker's life from being saved in order to recover the damage to his property. Likewise, the compensatory analogue of Vincent preventing the transport company from mooring is later preventing the ship and crew from being saved from a high risk of harm in order to recover \$500. And everyone would agree that a moral theory which permitted these ways of extracting compensation is ludicrously draconian.

In sum, accepting that there are limits to the liability to compensate, just as there are limits to the liability to defensive harm, such that a victim sometimes lacks the right to claim to full compensation, shows that cases like these are not counterexamples to the substantive symmetry of the two forms of liability.

#### 4.2.2. THE RECEIVED VIEW OF PROPORTIONALITY

Again, I will postpone my full defence of the substantive symmetry of defence and compensation until much later. First, I want to turn to the question of what makes a given case of compensation disproportionate in the first place.

As I said earlier, because we often take for granted that wrongdoers are liable to fully compensate their victims, this question arises mainly when the only way for a victim to fully recover small damages would impose an enormous cost on the person responsible for them. The clearest cases are those of non-compensable harms, in which a perpetrator lacks the means to fully compensate his victim.

When, then, is the payment of some amount of compensation excessive, such that a perpetrator is not liable to pay it? According to the most extreme answer to that question, no amount of compensation, however large, is excessive: if there is something more that the perpetrator can do for the victim, then there is something more that he should do for her. Peter Vallentyne has defended this view. But, as we have seen, this extreme answer is implausible. It is very hard to believe, for instance, that the homeowner would be permitted to recuperate the damage to his furniture by preventing the backpacker's life from being saved, or that Vincent could pay for the repairs to his dock by thwarting the prevention of a serious risk from befalling the company's ship and crew. On the contrary, sometimes, even though a victim has not been fully compensated, she lacks the right that the perpetrator compensate her further.



In light of both the implausibility of the extreme answer and the structural similarity of compensation and defensive harm, it is natural to suppose that the limits of proportionality in compensation are analogous to those of what I will call the *Received View* of proportionality in self-defence.<sup>20</sup> On the Received View, to determine whether some defensive act in a choice set is disproportionate we first ask the extent to which he or she is responsible or culpable for it. Once we know this, we need look only at two features of the available acts: the magnitude of the harm to the victim that each act can avert, and the magnitude of the harm to the perpetrator that averting it will cause. The defensive act is disproportionate if and only if the harm it does to the perpetrator is too great relative to the harm it prevents.<sup>21</sup>

The simplest plausible function relating these two variables is probably a ratio. A proponent of the Received View then holds that for a given degree of responsibility or culpability for some threat there is some multiplier  $M$  such that a self-defensive action to prevent the threat is proportionate if and

---

<sup>20</sup> Two notable dissenters are Jonathan Quong (2015), whose view I discuss later, and McMahan (unpublished). McMahan has recently entertained the view that when a victim stands to suffer an enormous cumulative harm only a tiny portion of which is done to her by a culpable aggressor, then that aggressor may be liable to suffer an enormous harm to prevent his tiny contribution to the total harm. I restrict my own discussion of the Received View only to large harms and those for the total of which the perpetrator is responsible.

<sup>21</sup> The probability of the harm will also need to be taken into consideration, though I will here consider only cases of certain harm.

only if the harm the victim visits upon the perpetrator does not exceed  $M$  times the harm to herself she prevents by doing so.

It should now be clear how to transpose the Received View to the context of compensation. The Received View of proportionality in the liability to compensate says that a perpetrator is liable to provide compensation only if the benefit to the victim of receiving it is not too small relative to the cost to the perpetrator of providing it. In what follows, I will sometimes use “Received View” to refer to the view as it applies to both compensation and self-defence, but context should make it clear which of the two I am discussing.

#### 4.2.3. AGAINST THE RECEIVED VIEW OF COMPENSATION 1:

##### DEGREES OF WRONGDOING

The Received View of proportionality in compensation can easily accommodate cases like *Feinberg’s Cabin*. The reason the backpacker is not liable to lose his life in order to restore the damage to the homeowner’s furniture, on this view, is that that cost is too great relative to the benefit to the homeowner of being compensated. The Received View also correctly implies that the backpacker *is* liable to pay for the damages he has done to the cabin and furniture after the fact. Despite these virtues, ultimately, as a

metric for proportionality in compensation, the Received View is false—or so I shall argue.

I will focus on two ways in which the Received View of compensation is defective. The first holds that the view is at best incomplete, since it fails to account for the degree to which a harm wrongs its victim, which can affect how much compensation is owed. My second criticism charges the view with failing to account for how perpetrators fare *relative* to their victims, and, accordingly, lets perpetrators off far too lightly in cases in which the harm done is either non-compensable or compensable but very costly. Following these two criticisms of the Received View in compensation, I will defend the substantive symmetry of compensation and defensive harm, thereby establishing that both criticisms also apply to the Received View of proportionality in the ethics of defensive harming.

Jonathan Quong has suggested that something akin to the extent to which an individual is wronged by an act is what determines proportionality in the liability to defensive harm.<sup>22</sup> Though I am sceptical that wrongness is the *only* factor that is relevant (for reasons that will become clear later),

---

<sup>22</sup> Quong (2015). Quong appeals to the stringency of the right violated, but we should leave open whether an act that wrongs its victim to a greater extent necessarily violates a more stringent right of hers. One reason to leave this possibility open is that the alternative would seem to imply that for every token act that differs from other token acts, however slightly, in the extent to which it wrongs a victim, the victim has unique right against it.

what follows here can be seen as something of a development of that view as it applies it to compensation instead. Before embarking I should make two brief remarks about what it means for an act to vary in the degrees to which they wrong someone. First, this notion of the degree to which an act wrongs *someone* is not identical to that of the overall degree of wrongness of an act. To see the difference, note that while it is very plausibly more wrong to maliciously break the arm of each of ten people at once than to break just one person's arm, each individual person is *wronged* to the same degree by having her arm maliciously broken regardless of how many others are also wrongfully injured.

Second, if the degree to which an act wrongs someone bears on the liability of the one doing it, then it is important that wrongness really does come in *degrees*. It is not enough that we are able to rank actions ordinally in terms of more and less wrongful. If wrongness is to bear on how much compensation is owed we must also be able to say to what extent an act wrongs a person. Some may doubt that the notion of degrees of wrong done to others makes sense, but in fact we appear to make judgments like this all the time. "Stealing someone's umbrella wrongs her a bit more than cutting in front of her in a queue, but murdering someone wrongs her a lot more than stealing

her umbrella.” This statement makes sense only if we intuitively understand comparisons of wrongdoing to involve more than an ordinal ranking of the relevant acts.

Now, at least when it comes to defensive harm, it seems to me entirely correct that the limits of proportionality correlate with the extent to which a victim stands to be wronged by a given threat, other things being equal.

Compare two cases:

- (1) P could easily stop a loose boulder from rolling down a hill, where it will crush and kill the innocent V, but out of malice for V he chooses to allow it to roll by.
- (2) P intentionally and maliciously pushes a loose boulder down a hill where it will crush and kill V.

Assume that the harm to V of death is the same in both cases. It seems to me that if V could somehow stop the boulder from crushing her, but only by placing P in its path where it will crush him instead, she would be permitted to do so in (2) but not in (1). Granted that in these cases killing V constitutes a greater wrong done to her than letting her die, this is evidence for the fact that a person can be liable to more defensive force the greater wrong his action does, or would do, his victim.

Further, it seems to me that if, once the boulder started rolling, P could stop it, but only at some harm to himself by placing himself in front of it uphill from V, there is some cost such that he would be required to incur it to stop the boulder in (2) but would not be required to do so in (1). That seems to further establish that the permission of V to defend himself by killing P in (2) but not in (1) is grounded at least in part in P's *liability* to suffer more harm in that case—rather than, for instance, considerations of retribution.

If we accept my eventual suggestion that the limits of proportionality *ex post* mirror those of proportionality *ex ante*, the fact that P is liable to more defensive harm in (2) would, by extension, show that the wrongfulness to the victim of an act helps determine the degree of compensation the victim is owed. But I believe it is by itself intuitive that the amount of compensation owed varies according to the wrongness to the individual of the harm.

When it comes to more typical cases of compensation this is easy to overlook. That is because we often take for granted that full compensation is owed. Consider:

- (3) P negligently sets off some fireworks in a residential area and they burn down V's garden shed.

- (4) P intentionally and maliciously burns down V's garden shed thinking he can get away with it.

Most people would agree that what P does in (4) wrongs V to a greater degree than what he does in (3). But it is also reasonably clear that P owes V full compensation in both cases and, further, that he will have fully discharged his compensatory duty once he has replaced V's shed.<sup>23</sup> Because victims do not have a right to claim more than full compensation from those who have harmed them, though, it is quite possible that P's liability to pay full compensation for the more wrongful act he perpetrates in (4) is simply overdetermined, since he is also liable to pay full compensation in (3).

The degree to which an act wrongs someone partially determines how much compensation is owed to her emerges primarily in cases in which it is an open question whether full compensation is owed. Consider:

- (5) P drives carelessly and above the speed limit and hits V with his car, severely disabling her.
- (6) P maliciously and intentionally runs down the innocent V with his car, severely disabling her.

---

<sup>23</sup> Again, I set aside what *other* duties P may incur, like those to apologise, express remorse, or submit to legal punishment.

Suppose that P and V are each initially well off at a level of wellbeing of 100, and that by hitting V with his car P lowers her wellbeing to 75 in either case. Suppose, further, that it is extremely costly for P to compensate V, such that the marginal cost to him of compensating her far outweighs the extent to which the receipt of compensation offsets what harm she has suffered. Imagine, for instance, that fully compensating V would require P to decrease his own wellbeing to a very low level that corresponds to 1.

We can represent the possible distributions as a continuum from the situation immediately after P hits V to the situation in which P fully compensates V, as illustrated by the following table.

**Table 1**

P's wellbeing	V's wellbeing	
100	75	<i>Ex post the harm</i>
⋮	⋮	
70	80	
⋮	⋮	
40	85	
⋮	⋮	
⋮	⋮	
1	100	<i>Full compensation</i>



Let us assume, if only for the sake of argument, that in neither (5) nor (6) is P liable to *fully* compensate V. On that assumption, it seems clear to me that P must go further in the direction of *Full compensation* in (6) than in (5). This shows at minimum that, as applied to the liability to compensate, the picture given by the Received View is too simple. There is at minimum a third factor, apart from the magnitudes of the relevant harms, which helps to determine how much compensation a perpetrator is liable to pay. When it comes to *non-compensable* harms—in which full compensation is possible—we should therefore accept that the strength of the perpetrator's duty to compensate his victim depends in part on the extent to which the harm has wronged her.

#### 4.2.4. AGAINST THE RECEIVED VIEW OF COMPENSATION 2:

##### RESPONSIBILITY FOR UNJUST UNFAIRNESS

I will now turn to a distinct criticism of the Received View as it applies to proportionality in compensation. This criticism appeals not to a factor that the Received View fails to account for, but rather to a deeper issue with how it quantifies proportionality in the first place. It is my contention that the Received View gives the wrong judgment about what is and is not

disproportionate when the marginal cost to a perpetrator of paying compensation exceeds the marginal benefit to the victim of receiving it.

It is worth mentioning that the term “proportionality” can give the misleading impression that something like the Received View is the only possible view. For the term suggests that whether a given act is disproportionate is a matter of some ratio (proportion) between two variables. Yet there is no reason to assume *a priori* that this is the right test for when a certain cost is or isn’t excessive. The excessiveness of inflicting a certain cost on a perpetrator needn’t have anything to do with the relationship of that cost to the benefit secured, the harm prevented, or indeed anything else—all we are after is a test that delivers the right verdict for those acts we intuitively deem excessive or not (a more accurate, if clumsier term might have been “proportionateness”).

One would display no conceptual confusion, for instance, in maintaining that proportionality is just a matter of not imposing large costs on perpetrators, nothing more. Equally, one might appeal only to the effect on the victim: perhaps imposing a cost on a perpetrator is disproportionate just in case doing so would benefit the victim only slightly. (Suitably interpreted, both of these single-variable views could correctly imply that neither the backpacker

nor the transport company could be liable to risk their lives to repair the property they have damaged.) Neither of these alternative conceptions of proportionality is compelling in the end, but we should reject them because they are implausible, not because they are ruled out conceptually.<sup>24</sup>

It is my contention that the Received View should be rejected because it can conflict with the following principle:

*Principle:* When a perpetrator is fully culpable for making his victim worse off than himself, he is liable to compensate her at least until he is no better off than she is, unless the victim can be fully compensated at a lesser cost.

When applied to cases in which the marginal cost to the perpetrator of providing compensation is great relative to the marginal benefit to the victim of receiving it, this principle rules as proportionate costs that would be ruled disproportionate by any specification of the Received View of compensation. For, if a perpetrator remains better off than his victim after he harms her, the principle implies that he can be liable to bear a very great cost, even

---

<sup>24</sup> Appealing only to the cost to the perpetrator incorrectly implies that a very large compensatory claim is disproportionate even when necessary to make up for an even greater harm done to the victim; appealing only to the size of the benefit to the victim incorrectly implies that all small claims are disproportionate.

when doing so will do comparatively little to improve the victim's condition. Suppose, for example, that you were previously able-bodied, and my wrongful act leaves you permanently quadriplegic, and there is no way for me to fully compensate you. According to the principle I am proposing, I can be liable to toil for the rest of my life only to transfer to you whatever goods I can produce beyond those necessary for me to maintain a relatively austere existence. That is so because I would prefer to toil for the rest of my life at little personal benefit than be permanently quadriplegic with whatever benefits I could feasibly receive in that state. We can represent a situation like this—in which a perpetrator visits a severe and non-compensable harm upon a victim—as follows:

**Table 2**

	P	V
<i>Ex ante the harm</i>	100	100
<i>Ex post the harm</i>	100	1

Again, it is my contention that in this situation, the perpetrator is liable to compensate the victim at least until the following state of affairs obtains, provided the necessity condition is satisfied:

Table 2 (continued)

	P	V
<i>Equality</i>	2	2

Because the gain to V of receiving compensation is very small relative to the cost to P of providing it, bringing about the state of affairs represented by *Equality* would be disproportionate on the Received View of compensation. So if I am right that P is indeed liable to bring about *Equality*, we should reject the Received View.

Why would P be liable to sacrifice so much just to improve V's wellbeing only slightly? The reason is that anything else would be manifestly unjust. To show this, it will be helpful to first consider a variation on this case in which the unequal distribution represented by *Ex post the harm* in Table 2 is due to natural causes. Suppose that V's poor condition is the result of an unforeseeable accident. In this case, many distributive egalitarians would already hold that V has at least *some* right against P that he compensate her, on the grounds that it is unjust for one person to be worse off than another through no fault of her own.<sup>25</sup>

---

<sup>25</sup> I here make the simplifying assumption that wellbeing—very broadly construed as whatever a person has a self-interested reason to want—is the metric of distributive justice.

Of course, only a very hard-nosed egalitarian would believe that when the unfairness is due to no fault of his, P has an all-things-considered duty to go so far as *Equality*. Few believe that morality could demand of a person that he make such an enormous sacrifice just for the sake of making the situation as fair as possible. Egalitarians can avoid that counterintuitive conclusion in either of two ways. On the one hand, they can take what we might call an *external* line, according to which the reason why P is not required to go so far to correct the natural unfairness at hand is that there is a counter-vailing consideration that tells against the requirement to do what justice demands.

Two possible considerations come to mind. First, as G. A. Cohen has claimed, P may have an agent-relative prerogative not to fulfil his impartial obligation of justice, perhaps on the ground that he may give special weight to his own interests when they conflict with the demands of impartial morality.<sup>26</sup> Alternatively, there may be a separate aspect of impartial morality that competes with justice, like the reason we have to maximise the good. On the credible assumption that equality is not the only good, perhaps P ought not

---

<sup>26</sup> Cohen (2008), p. 61. Cohen does not entertain the alternative explanation that I have called the “internal” line, possibly because of his stated commitment to value pluralism.

to bring about *Equality* because doing so would be sufficiently worse than not doing so, axiologically speaking, to outweigh what justice itself demands.

On the other hand, those who wish to deny that P must compensate V until *Equality* is reached, even when he is in no way responsible for that inequality, may take an *internal* line. This holds that P would not be required to fully neutralise the unfair inequality not because there is some reason *competing* with justice, but because justice itself simply does not demand so much of a person.<sup>27</sup>

Both lines provide plausible explanations for why a person's all-things-considered duty to correct natural unfair inequality is less than full. But neither is viable when it comes to rejecting the principle I have been defending in the case of wrongful harm.

The external line is unsatisfactory because the existence of morally-relevant reasons not to compensate a person do not show that one is not *liable* to compensate her. The strength of the victim's right to be compensated by the perpetrator is not lessened on the grounds of the perpetrator's agent-rel-

---

<sup>27</sup> The thought that there is an internal limit to what distributive justice demands of us is familiar; see, for example, Fleurbaey (1995); Williams (2006); Olsaretti (2009); Stemplowska (2009). In the present context, this explanation of why P would not be required to bring about full equality recalls the "satisficing consequentialism" developed in reaction to the demandingness objection to standard act consequentialism (by, for example, Slote and Pettit (1984)).

ative permission not to compensate her. And the possibility that there is a competing aspect of morality, like the reason one has to maximise the good, is a matter of “wide” proportionality and does not affect the liability of the perpetrator to compensate.

The internal line is unsatisfactory because the cases with which we are concerned do not involve natural unfairness; they are ones in which one person is *morally culpable* for the fact that the other is much worse off than him. When asked to make such a great sacrifice when P himself bears *no responsibility* for the unfair inequality that exists, P could reasonably say to V, “While I recognise it to be unfair that you’re worse off, you have no right to demand that I sacrifice so much for your small benefit.” But when the unfair inequality is the result of P’s wrongdoing, can he truly say something of the sort? He could not, I believe, justifiably say to V, “I know that I am fully culpable for the bad state you are in, but you have no right to demand that I make a great sacrifice for your own small benefit, even though by doing so I would end up in no worse a condition than the one in which I have wrongfully left you.” Could not V reply, at any point before *Equality* had been reached, “You’re fully responsible for this sad state of affairs. If anything, you should be *worse* off than me. So how could I lack a further right that you



improve my condition, while you still remain better off than I am?" I believe V would be justified in responding in that way.

The central case we have been considering involves a very large harm that carries with it the possibility of only minor compensation. I have claimed that even when doing so would come at great cost to the perpetrator and provide little benefit to the victim, the perpetrator is nevertheless liable to do so at least until he is no worse off than she. Perhaps, however, when the amount of benefit V can derive is truly minuscule this is just too hard to accept. For the principle implies that a perpetrator can be liable for arbitrarily small increases ( $\epsilon$ ) to the victim's wellbeing that can only be compensated by arbitrarily large decreases in his own, as in the following case:

**Table 3**

	P	V
<i>Ex ante the harm</i>	1,000	1,000
<i>Ex post the harm</i>	1,000	1
<i>Equality</i>	$1+\epsilon$	$1+\epsilon$

Though I would maintain that, even in this case, in virtue of his act P becomes liable to compensate V at least until *Equality* is reached, it may be

hard to believe that V retains a right to that compensation. If so, that may be because V herself has an obligation to give P what amounts more or less to a *costless benefit*. We are typically morally required to confer large benefits on others—even those who are already better off than us—when we can do so at negligible cost to ourselves. It is a corollary of that principle that we lack a right that others relinquish the benefits in question even if they are independently liable to do so.

The arguments in this section indicate, I suggest, that we should reject the Received View of proportionality in the liability to compensate. On that view, recall, proportionality is a matter of the relationship between the magnitude of the cost to the perpetrator of compensating and the benefit to the victim of receiving compensation. Accordingly, whether a compensatory claim is disproportionate, and hence whether the perpetrator is liable to pay it, is a question which can be answered without knowledge about how well off the relevant parties are relative to one another. The principle I have defended implies that, as it pertains to proportionality in compensation, the Received View is false. For the principle implies that a sufficiently well-off perpetrator can be liable to compensate his victim to a very small degree, even at a great

cost to himself—a cost which would be deemed disproportionate by any specification of the Received View.

#### 4.2.5. THE SUBSTANTIVE SYMMETRY OF COMPENSATION AND SELF-DEFENCE

On the assumption that the liability to defensive harm and the liability to compensate are substantively symmetrical, the conclusion of the preceding subsection has important implications for proportionality in defensive harm. If the substantive symmetry is correct, a perpetrator who culpably threatens to make someone worse off than himself is liable suffer an enormous harm to prevent even a small portion of the threat he poses. For example, an able-bodied aggressor who culpably threatens to make another able-bodied person permanently paralysed could be liable to suffer permanent paralysis himself even when doing so would serve no greater purpose than rendering the victim's own paralysis a little less severe. That is a result that would be deemed disproportionate on any version of the Received View of proportionality in defensive harming.

Suppose we are reluctant to reject the Received View of proportionality in the liability to defensive harm. We then face a trilemma. On pain of inconsistency, we must reject one of the following theses:

- (I) When a perpetrator is fully culpable for making his victim worse off than himself, he is liable to compensate her at least until he is no better off than she is, unless the victim can be fully compensated at a lesser cost;
- (II) The Received View of proportionality in self-defence;
- (III) Proportionality in the liability to defensive harm mirrors proportionality in the liability to pay compensation.

I have argued that we should accept the principle expressed by (I). Someone who finds my arguments to that end persuasive, but who is reluctant to abandon the Received View, will want to deny (III), the substantive symmetry. That person may insist that, as McMahan has said, “one cannot infer that a person is liable to defensive action *ex ante* from his being liable to pay compensation *ex post*.”<sup>28</sup>

I shall argue that rejecting the substantive symmetry is not a viable option, however, and I shall do so on three counts. First, doing so has implaus-

---

<sup>28</sup> McMahan (2008), p. 233.

ible implications. Second, grounding liability to defensive harm in duties of redistributive justice makes it less plausible that there could be substantive differences between the two types of liability. And third, a more nuanced understanding of the *contingent* differences between the two types of liability reveals that, when comparisons between them are properly equalised, any substantive difference that remained could be based on nothing but timing.

Firstly, then, suppose for the sake of argument that the symmetry between the two types of proportionality did not hold. Then the highest proportionate cost one would be permitted to inflict on an attacker in self-defence against a given threat would sometimes differ from the highest proportionate cost one could extract from him in compensation should that same threat eventuate. If in a given instance these costs did differ, we would then need to ask on which of the two sides it is higher. But if, on the one hand, it is higher on the defensive side, this would unjustifiably favour those who have the ability to defend themselves over those who do not. And if, on the other hand, it was higher on the corrective side, then in some cases (such as non-compensable harms) a victim would have the perverse incentive to allow herself to be harmed, for then she could permissibly extract more compensation

and thereby end up better off overall. Neither result is easy to accept, which provides some support for the substantive symmetry.

My second argument draws on the theoretical underpinnings of proportionality in defensive harming generally. Obviously enough, sometimes when we harm others we also wrong them. When harming a person is both necessary for the sake of some end and (narrowly) proportionate, however, then it does not *wrong* the person to inflict that harm for the sake of that end, by which I mean that doing so neither violates, nor merely infringes, his rights.<sup>29</sup> It is difficult to see how one could inflict harm upon another person without even infringing his rights unless he also had a moral obligation to suffer that harm. So it follows that imposing a cost on someone is (narrowly) proportionate only if they have a duty to bear that cost.

What could explain why a responsible threatener could have a duty to bear some or all of the cost a potential victim can impose upon him? A plausible rationale that has been suggested is that the duty is one of redistributive justice.<sup>30</sup> In cases of self-defence, if the necessity condition is met then some cost will be borne by someone. A natural way to explain why the

---

<sup>29</sup> See Thomson (1986), p. 51; (1990), p. 122.

<sup>30</sup> For the suggestion that right to defensively harm is grounded in considerations of the just redistribution of costs, see McMahan (2005b), p. 395; Montague (2010).

perpetrator often has a duty to bear the bulk of the cost in question is that he is responsible for the fact that *someone* must bear a cost. Because people have a duty not to excessively *externalise* the costs of their choices onto others without those others' consent, responsible threateners have a duty to *internalise* a greater portion of the total cost that must in the end be suffered by someone.

If this is right, and the duty not to *externalise* the costs of one's choice is what explains the duty *to* internalise them, then it is hard to see how the limits of what proportionality permits could differ between the liability to self-defensive harm *ex ante* and the liability to pay compensation *ex post*, when other things are equal. The choice of the perpetrator and the potential cost someone stands to bear are the same, and it seems unlikely that, when other things are equal, the cost he is forbidden to externalise onto others could depend on whether he did so directly (through an un-averted threat) or indirectly (through unpaid compensation).

My final argument for the substantive symmetry involves identifying and factoring out a contingent difference between many cases of self-defence and those of collecting compensation. To that end, it is necessary to distinguish two different modes of correcting for harm, or for the prospect of harm:

*negating* and *offsetting*.<sup>31</sup> An act aims at negating a harm when the outcome it would bring about aims to approach the possible world in which the harm was not done in the first place. In the context of corrective justice, the *restoration* to a victim of lost property (rather than compensation for it) aims at negating (rather than offsetting) a harm. On the other hand, an act aims at *offsetting* a harm when it aims to bring about an outcome that is merely *as good as* the possible world in which the harm was not done. Again in the context of corrective justice, monetary compensation paid to a physically injured party for damages aims at offsetting (rather than negating) harms.

Now, one manifest difference between self-defence and compensation that has not, to my awareness, been hitherto acknowledged is this: while paradigmatic acts of defence aim to negate harm, those of compensation aim to offset it. And attending to that difference helps reveal the deeper structural relationships between the two types of act. The *ex post* analogue of defensive harm is restoration, not compensation. And the *ex ante* analogue of compensation is offsetting a harm before it occurs, not defending oneself against it.

---

<sup>31</sup> For accounts of this distinction in the context of corrective justice only, see Goodin (1989); Gardner (2011), pp. 28–37; Slavny (2014); Gardner (2018), pp. 102–124.



We should begin, therefore, by properly equalising cases, such that we are comparing a case of defensive harm with one of negating the harm *ex post*, or a case of compensating a harm with one of offsetting the harm *ex ante*. It is easiest to do so by way of some unrealistic examples. In the following cases, I enter the scene having already lost my left leg for unrelated reasons, and you act so as to cut off my right leg. Assume that you and I are equally well off *ex ante*, that the prudential value for me of keeping my right leg is equal to that of having a left leg and that I am indifferent between having either, and that there is only one corrective action available to me in each of the cases (that is, if I am able to defend myself, I am not able to claim compensation, and vice versa).

Compare, first, two cases of negating:

- (1) *Self-defence*: Before you are able to act, I can prevent you from cutting off my right leg, though doing so will cause you to suffer some harm *H*.
- (2) *Restoration (ex post negating)*: After you cut off my right leg I can, with minimal pain and effort, reattach that same leg in full working order, though doing so will cause you to suffer *H*.

And next, two cases of offsetting:

- (3) *Ex ante offsetting*: I cannot stop you from cutting off my right leg, but before you are able to attack me I have the opportunity to proactively regrow my left leg, though doing so will cause you to suffer *H*.
  
- (4) *Compensation*: I cannot reattach my right leg, which you have just cut off, but I can regrow my left leg, though doing so will cause you to suffer *H*.

Since the effect on you of my acting (i.e., *H*) is the same in all four cases, we can distinguish in the following way the four outcomes in which I act:

**Table 4**

	Negating the harm	Offsetting the harm
<i>Ex ante</i>	(1) <i>Self-defence</i> Keep my right leg.	(3) " <i>Ex ante offsetting</i> " Regrow my left leg.
<i>Ex post</i>	(2) <i>Restoration</i> Lose and regain my right leg.	(4) <i>Compensation</i> Regrow my left leg.

When we compare any one of these four acts with its analogue in the same column, it seems to me that the difference between them could not make a difference to proportionality. This is clearest when we compare cases of offsetting: *ex ante* offsetting a harm is just collecting compensation for it before it eventuates. If that is the only way for the victim to have restitution, it is

hard to believe that whether  $H$  is proportionate or not could depend on whether the compensation is collected before or after a threat eventuates.

Provided the threat is already engaged and certain to occur, the difference is nothing but a matter of timing. Consider an analogue of *Feinberg's Cabin*. Suppose the cabin is on a mountain frequented only by wealthy and over-prepared hikers who always take with them a credit card. If the homeowner is entitled to compensation for damages, then surely he could permissibly install a credit card machine on his cabin door such that, in the event of a blizzard, endangered hikers would need to pay upfront the estimated cost of the damages they will do by entering, assuming the necessity of paying would impose no additional risk on the hikers (which compensation *ex post* likewise does not).

When it comes to *negating* harms there is, however, one disanalogy between doing so *ex ante* and doing so *ex post*. Restoration almost always will involve an unavoidable temporary loss of the good in question, while successful defensive harm can mean that not even a temporary loss occurs. Could that difference itself imply that the limits of liability to self-defence differ from those to provide restoration? I believe it could not. For we ought to

view a claim to restoration *ex post* as an instance of defence against the *permanent* loss of the relevant good.

Thus, consider a variation on (1) (*Self-defence*) in which a temporary loss is unavoidable:

- (1\*) Whatever I do, you will soon cut off my right leg, but if and only if I strike first, the severed leg will be preserved so that it can be easily reattached thereafter, though striking first will cause you to suffer *H*.

The difference between (1\*) and (2) (*Restoration*) really is just one of timing, and it is accordingly highly implausible that there is any value of *H* such that inflicting *H* on you is proportionate in the one case but not in the other.<sup>32</sup>

Of course, when a victim can, by engaging in necessary self-defence, ensure that she incurs not even a temporary loss, then the amount of harm the attacker is liable to bear may well differ from the amount he would be liable to bear in order to negate the injury *ex post*. For, in the latter case, the victim will also have suffered a greater setback. But that is not an argument against the equivalence of *ex ante* and *ex post* negating; rather, it is a matter

---

<sup>32</sup> Though I shall not pursue the matter, depending on when the relevant events occur, their temporal order may not be preserved across inertial reference frames. Certain views of the metaphysics of time therefore imply that there is not even a stance-independent fact of the matter about whether my act constitutes defence or restoration. The same goes for the comparison between compensation and *ex ante* offsetting.

of the threat of harm being smaller in a case in which it can be averted than in a case in which it eventuates and can only be negated *ex post*. And that difference would typically call for additional restitution. So the limits of liability can sometimes differ between negating by defence and negating by restoration, but the difference here would not be based on when the negating occurred.

These observations cast serious doubt on the notion that proportionality *ex ante* could differ from proportionality *ex post* when we keep constant whether the comparison is between instances of negating or instances of offsetting. The other step in this argument for the substantive symmetry is to show that what proportionality demands also does not differ on the basis of whether an act involves offsetting or negating.

The equivalence of negating and offsetting harm would be easy to establish if I could show that the harm some act does to a victim is *fungible*, in the sense that its moral relevance *qua harm* is just the extent to which it impacts the victim's wellbeing. Then negating and offsetting would be interchangeable in the requisite way, for what cost the perpetrator is liable to bear would be independent of the manner in which he is required to bear it (whether by offsetting the harm or by negating it). But I do not believe that harms are fun-

gible in that way. If harm were in this way fungible, then it would be up to a perpetrator whether to restore the goods he has wrongfully taken from his victim or merely compensate her for their loss. But it seems clear that the perpetrator is, in the first place, liable to negate a harm, and may compensate his victim for a harm instead only when negating the harm is impossible or she consents to him doing so. Nor, moreover, may a *victim* unilaterally claim compensation for a harm when the perpetrator can negate it and prefers instead to do so. In short, if negation is possible, then the perpetrator is liable to do so and only to do so.

Be that as it may, it is very hard to believe that the amount of cost a perpetrator is liable to bear to offset a given amount of his victim's lost well-being could differ from the amount he is liable to bear to negate the same loss, other things being equal. To see this, it is helpful to compare cases of harm that cannot be fully rectified and in which the perpetrator can go some way toward making up for it only by partially *negating* it, with those in which the perpetrator can do so only by partially *compensating* for it. It seems that the highest cost a perpetrator can be liable to bear to offset his victim's loss to a given extent could not be *higher* than the cost he would have had to bear to negate it to that same extent, for what would circum-

scribe the perpetrator's liability to offset a harm, if not the extent to which would have been liable to negate it, were doing so possible? But it also seems that the highest cost a perpetrator is liable to bear to offset a victim's loss to a given extent could not be *lower* than that which he would have had to bear to negate it, if doing so had been possible. Surely wrongdoers cannot get away with less, merely because the best way to make up for the harm they have done—negating it—is unavailable.

I have offered three separate arguments for the substantive symmetry of the liability to defensive harm and the liability to compensate. I conclude that to the extent, and only to the extent, that a perpetrator is liable to defensive harm *ex ante*, he is also liable to pay compensation *ex post*.

#### 4.2.6. A SUMMARY OF SECTION 4.2

We have covered a lot of ground in this section. Before concluding this chapter, it may be helpful to briefly summarise the points I have made and the arguments I have given. The springboard for this second section was that class of non-compensable harms whose victims are nonetheless compensable to some extent—those for which there is no point at which the perpetrator

has more to give but the victim nothing more to receive. I began by illustrating that compensation has its own proportionality constraint and showing how this helps to answer some putative counterexamples to the substantive symmetry of the liability to defensive harm and the liability to compensate.

I then discussed the Received View of proportionality as it applies to both defensive harm and compensation, according to which given a certain degree of responsibility or culpability for some harm, the variables relevant to determining the limits of proportionality are the cost imposed on the perpetrator and the cost to the victim that is prevented. Following this, I offered two arguments against the Received View as it applies to compensation. The first was that it does not account for the relevance to proportionality of the extent to which a given harm has wronged its victim; the second was that it fails to account for the relative conditions of the perpetrator and victim which, as a matter of distributive justice, ought to count—and therefore lets perpetrators off far too lightly in certain cases, especially those of non-compensable harm.



I then focused on the latter thesis (i.e., that for claim (I) in the tri-lemma below), and noted that we must reject at least one of the following three claims:

- (I) When a perpetrator is fully culpable for making his victim worse off than himself, he is liable to compensate her at least until he is no better off than she is, unless the victim can be fully compensated at a lesser cost;
- (II) The Received View of proportionality in self-defence;
- (III) Proportionality in the liability to defensive harm mirrors proportionality in the liability to pay compensation.

Finally, I argued that rejecting (III) is not a viable option: other things being equal, the cost a perpetrator is liable to suffer to avert the threat he poses can be neither more nor less than the cost he is liable to bear to compensate his victim *ex post*. We must therefore choose between (I) and (II). My suggestion is that we reject (II), departing even further from the Received View of proportionality in the ethics of self-defence.

### 4.3. Conclusion

Often the steps required to rectify wrongdoing are straightforward, at least in principle. When wrongful harms to others are fully compensable and the costs to their perpetrators of compensating for them is not prohibitively expensive, what is owed is usually just full compensation. But things are not always so straightforward, even in principle. One sort of case in which they are not is the one in which a perpetrator cannot or may not compensate his victim, at least beyond a certain threshold that is below full compensation. I have defended the thesis that the remainder of what would have been his compensatory obligation had compensation been possible must be discharged in the service of the impartial good.

The second less straightforward case occurs when further compensation is both possible and permissible, but the marginal cost to the perpetrator of compensating his victim is significantly greater than the marginal benefit to the victim of being compensated. Among other claims, I have defended the thesis that when a perpetrator is culpable for the fact that his victim is worse off than him, he can be liable to compensate her at least until they are equally well off. This means that perpetrators must sometimes bear great costs even to provide small benefits to their victims.

I would like to close by asking how the two main theses I have defended might interact. Early in this chapter's second section, I rejected the most extreme view of what proportionality in compensation demands, according to which there is in principle no upper limit to how much compensation a culpable perpetrator can be liable to provide. Provided we reject that view, then although in some cases the upper limit is very high, there is a point at which a perpetrator can be said to have done enough and at which a victim lacks the right to further compensation notwithstanding the fact that she has not been fully compensated. In these cases, the duty described by the Non-Compensable Harms Thesis surfaces, for this perpetrator would not have wholly discharged his *pro tanto* duty to restore the impartial goodness he has wrongfully eliminated from the world.

There are cases in which a given amount of cost or effort on the part of a perpetrator would go further toward bettering the world impartially if deployed elsewhere than it would go toward compensating his victim if he used it instead to benefit her. About these cases, my own intuition is that the duty to further compensate one's victim has some priority, but not lexical priority, over one's duty to restore impartial value to the world.

It is worth keeping in mind, however, that this need not mean that whenever a perpetrator has the opportunity to compensate his victim, and is liable to do so, that this is what he all-things-considered ought morally to do. For one thing, an offender may well have pre-existing duties of beneficence or justice which are stronger than any compensatory obligations he has toward his victims. That he should fulfil these other duties instead of compensating his victim is most likely when his own victim is unjustly well off relative to other prospective beneficiaries, for then he already has an independent reason of justice to redirect potential benefits away from his victim and toward those others. Furthermore, it may be that one can sometimes permissibly fail to discharge a *pro tanto* duty one has in order to promote the greater good instead, even when doing the latter would be supererogatory taken on its own. How we should balance these various competing considerations when they conflict is a difficult question that a complete theory of the rectification of wrongdoing, whether compensable or not, must eventually address.

## Bibliography

- Arrhenius, Gustaf and Rabinowicz, Wlodek (2010). “Better to Be than Not to Be?” In H. Joas, and B. Klein (eds.) *The Benefit of Broad Horizons*. Brill: pp. 399–414.
- Bailey, Cyril (trans.) (1963). *Lucretius: De Rerum Natura*. Oxford University Press.
- Brueckner, Anthony and Fischer, John Martin (1986). “Why is Death Bad?” *Philosophical Studies* 50 (2): 213–221.
- Benatar, David (2006). *Better Never to Have Been: The Harm of Coming into Existence*. Oxford University Press.
- Benatar, David (2017). *The Human Predicament: A Candid Guide to Life’s Biggest Questions*. Oxford University Press.
- Bradley, Ben (2004). “When Is Death Bad for the One Who Dies?” *Nôus* 38 (1): 1–28.
- Bradley, Ben (2009). *Well-Being and Death*. Oxford University Press.
- Broome, John (1993). “Goodness is Reducible to Betterness: The Evil of Death is the Value of Life.” In P. Koslowski and Y. Shionoya (eds.) *The Good and the Economical: Ethical Choices in Economics and Management*. Springer.
- Broome, John (2004). *Weighing Lives*. Oxford University Press.

- Bundesverfassungsgericht, *Judgment of the First Senate of 15 February 2006*. BvR 357/05, paras. 1–156. Available online: [http://www.bverfg.de/e/rs20060215\\_1bvr035705en.html](http://www.bverfg.de/e/rs20060215_1bvr035705en.html)
- Campbell, Tim and McMahan, Jeff (2010). “Animalism and the Varieties of Conjoined Twinning.” *Theoretical Medicine and Bioethics* 31 (4): 285–301.
- Cohen, Daniel and Luck, Morgan (2009). “Why a Victim’s Age is Irrelevant When Assessing the Wrongness of Killing.” *Journal of Applied Philosophy* 26: 396–401.
- Cohen, G. A. (2008). *Rescuing Justice and Equality*. Harvard University Press.
- Deng, Natalja (2015). “How A-Theoretic Deprivationists Should Respond to Lucretius.” *Journal of the American Philosophical Association* 1 (3): 417–432.
- Dorsey, Dale (2015). “The Significance of a Life’s Shape.” *Ethics* 125 (2): 303–330.
- Dworkin, Ronald (1993). *Life’s Dominion*. Alfred A. Knopf.
- Elliot, Robert (1991). “Personal Identity and the Causal Continuity Requirement.” *Philosophical Quarterly* 41: 55–75.
- Feinberg, Joel (1970). “Justice and Personal Desert.” In his *Doing and Deserving: Essays in the Theory of Responsibility*. Princeton University Press: 55–94.
- Feinberg, Joel (1978). “Voluntary Euthanasia and the Inalienable Right to Life.” *Philosophy and Public Affairs* 7 (2): 93–123.
- Feldman, Fred (1991). “Some Puzzles About the Evil of Death.” *The Philosophical Review* 100: 205–27.

- Feldman, Fred (1992). *Confrontations with the Reaper: A Philosophical Study of the Nature and Value of Death*. Oxford University Press.
- Feldman, Fred (2013). "Brueckner and Fischer on the Evil of Death." *Philosophical Studies* 162: 309–317.
- Fleurbaey, Marc (1995). "Equal Opportunity or Equal Social Outcome?" *Economics and Philosophy* 11: 25–55.
- Foot, Philippa (1985). Utilitarianism and the Virtues. *Mind* 94: 196–209.
- Frankfurt, Harry (1987). "Equality as a Moral Ideal." *Ethics* 98: 21–42.
- Frankfurt, Harry (2000). "The Moral Irrelevance of Equality." *Public Affairs Quarterly* 14: 87–103.
- Gardner, John (2011). "What is Tort Law For? Part 1. The Place of Corrective Justice." *Law and Philosophy* 30 (1): 1–50.
- Gardner, John (2018). *From Personal Life to Private Law*. Oxford University Press.
- Goodin, Robert E. (1989). "Theories of Compensation." *Oxford Journal of Legal Studies* 9 (1): 56–75.
- Hanser, Matthew (2013). "The Wrongness of Killing and the Badness of Death." In B. Bradley, F. Feldman, and J. Johansson (eds.) *The Oxford Handbook of Philosophy of Death*. Oxford University Press.
- Holtug, Nils (2010). *Persons, Interests, and Justice*. Oxford University Press.
- Horne, Malcolm (2009). "Are People in a Persistent Vegetative State Conscious?" *Monash Bioethics Review* 28 (2): 1–12.
- Huemer, Michael (2008). "In Defence of Repugnance." *Mind* 117 (468): 899–933.

- Hurka, Thomas (1993). *Perfectionism*. Oxford University Press.
- Johansson, Jens (2013). "Past and Future Non-Existence." *The Journal of Ethics* 17: 51–64.
- Johansson, Jens (2014). "Actual and Counterfactual Attitudes: Reply to Brueckner and Fischer." *The Journal of Ethics* 18: 11–18
- Kagan, Shelly (1993). "Me and My Life." *Proceedings of the Aristotelian Society* 94: 309–324.
- Kagan, Shelly (2012). *Death*. Yale University Press.
- Kahane, Guy (2011). "Evolutionary Debunking Arguments." *Noûs* 45 (1): 103–125.
- Kamm, F. M. (1993). *Morality, Mortality: Volume I: Death and Whom to Save From It*. Oxford University Press.
- Kamm, F. M. (1996). *Morality, Mortality: Volume II: Rights, Duties, and Status*. Oxford University Press.
- Kamm, F. M. (2007). "Review Article: The Ethics of Killing: Problems at the Margins of Life." *The Philosophical Review* 116 (2): 273–280.
- Kamm, F. M. (2017). "The Purpose of My Death: Death, Dying, and Meaning." *Ethics* 127 (3): 733–761.
- Karhu, Todd (2019). Non-Compensable Harms. *Analysis* 79 (2): 222–230.
- Kolak, Daniel and Martin, Raymond (1987). "Personal Identity and Causality: Becoming Unglued." *American Philosophical Quarterly* 24: 339–347.
- Kripke, Saul (1980). *Naming and Necessity*. Harvard University Press.
- Lewis, David (1976a). "The Paradoxes of Time Travel." *American Philosophical Quarterly* 13 (2): 145–152.



- Lewis, David (1976b). "Survival and Identity." *The Identities of Persons*, ed. A. Rorty. Berkeley: 17–40.
- Lippert-Rasmussen, Kasper (2007). "Why Killing Some People is More Seriously Wrong than Killing Others." *Ethics* 117 (4): 716–38.
- Lippert-Rasmussen, Kasper (2001). "Two Puzzles for Deontologists: Life-Prolonging Killings and the Moral Symmetry Between Killing and Causing a Person to Be Unconscious." *The Journal of Ethics* 5 (4): 385–410.
- Mackie, J. L. (1985). "Parfit's Population Paradox." In J. Mackie and P. Mackie (eds.) *Persons and Values*. Oxford University Press.
- Mackie, J. L. (1974). "What Is De Re Modality?" *Journal of Philosophy* 71 (16): 551–561.
- Mackie, Penelope (2006). *How Things Might Have Been: Individuals, Kinds, and Essential Properties*. Oxford University Press.
- McMahan, Jeff (1995). "Killing and Equality." *Utilitas* 7 (1): 1–29.
- McMahan, Jeff (2002). *The Ethics of Killing: Problems at the Margins of Life*. Oxford University Press.
- McMahan, Jeff (2005a). "The Lucretian Argument." In K. McDaniel, J. Raibley, R. Feldman, and M. Zimmerman (eds.) *The Good, the Right, Life and Death: Essays in Honor of Fred Feldman*. Routledge: 213–226.
- McMahan, Jeff (2005b). "The Basis of Moral Liability to Defensive Killing." *Philosophical Issues* 15 (1): 386–405.
- McMahan, Jeff (2008). "Debate: Justification and Liability in War." *Journal of Political Philosophy* 16 (2): 227–244.
- McMahan, Jeff (2009). *Killing in War*. Oxford University Press.

- McMahan, Jeff “Proportionality in Defense Against Inflictors of [Many] Small Harms.” Unpublished.
- Miller, David (1995). *On Nationality*. Oxford University Press.
- Montague, Phillip (2010). “Self-defense, Culpability, and Distributive Justice.” *Law and Philosophy* 29 (1): 75–91.
- Nagel, Thomas (1970). “Death.” *Noûs* 4 (1): 73–80.
- Noonan, Harold (2019). *Personal Identity, Third Edition*. Routledge.
- Nozick, Robert (1981). *Philosophical Explanations*. Harvard University Press.
- Olsaretti, Serena (2009). “IX—Responsibility and the Consequences of Choice.” In *Proceedings of the Aristotelian Society Vol. 109, No. 1pt2*. Blackwell.
- Olson, Eric (1997). *The Human Animal: Personal Identity Without Psychology*. Oxford University Press.
- Otsuka, Michael (2017). “Personal Identity, Substantial Change, and the Significance of Becoming.” *Erkenntnis* 83 (6): 1229–1243.
- Parfit, Derek (1984). *Reasons and Persons*. Oxford University Press.
- Parfit, Derek (1986a). “Comments.” *Ethics* 96 (4): 832–872.
- Parfit, Derek (1986b). “Overpopulation and the Quality of Life.” In *Applied Ethics*, P. Singer (ed). Oxford University Press.
- Parfit, Derek (2007). “Is Personal Identity What Matters?” Unpublished; available online: [http://www.stafforini.com/docs/parfit\\_-\\_is\\_personal\\_identity\\_what\\_matters.pdf](http://www.stafforini.com/docs/parfit_-_is_personal_identity_what_matters.pdf)

- Parfit, Derek (2008). "Persons, Bodies, and Human Beings." In J. Hawthorne, D. Zimmerman, and T. Sider (eds.) *Contemporary Debates in Metaphysics*. Blackwell.
- Parfit, Derek (2012). "We Are Not Human Beings." *Philosophy* 87 (1): 5–28.
- Perry, John (1972). "Can the Self Divide?" *Journal of Philosophy* 69: 463–488.
- Pitcher, George (1993). "The Misfortunes of the Dead." In J. Fischer (ed.) *The Metaphysics of Death*: 159–68.
- Raz, Joseph (1986). *The Morality of Freedom*. Clarendon Press.
- Raz, Joseph (2001). *Value, Respect, and Attachment*. Cambridge University Press.
- Robson, Jon (2014). "A-Time to Die: A Growing Block Account of the Evil of Death." *Philosophia* 42: 911–925.
- Quong, Jonathan (2015). "Proportionality, Liability, and Defensive Harm." *Philosophy and Public Affairs* 43 (2): 144–173.
- Scanlon, T. M. (2015). "Kamm on the Disvalue of Death." *Journal of Medical Ethics* 41 (6): 490.
- Scheffler, Samuel (2018). *Why Worry About Future Generations?* Oxford University Press.
- Schopenhauer, Arthur (1851/1981). "On the Sufferings of the World." In E. Klemke (ed.) *The Meaning of Life*. Oxford University Press.
- Sher, George (1979). "Compensation and Transworld Personal Identity." *The Monist* 62 (3): 378–391.

- Sher, George (2005). "Transgenerational Compensation." *Philosophy and Public Affairs* 33 (2): 181–200.
- Sidelle, Alan (2011). "Parfit on 'the Normal/a Reliable/any Cause' of Relation R." *Mind* 120 (479): 735–760.
- Shoemaker, Sydney (1970). "Persons and Their Pasts." *American Philosophical Quarterly* 7 (4): 269–85.
- Shoemaker, Sydney (1984). "Personal Identity: A Materialist's Account." In S. Shoemaker and R. Swinburne (eds.) *Personal Identity: Great Debates in Philosophy*. Blackwell: 67–132.
- Snowdon, Paul (1991). "Personal Identity and Brain Transplants." In D. Cockburn (ed.) *Human Beings*. Cambridge University Press.
- Soto, Carlos (2013). "Killing, Wrongness, and Equality." *Philosophical Studies* 164 (2): 543–559.
- Street, Sharon (2005). "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127 (1): 109–166.
- Sullivan, Meghan (2018). *Time Biases: A Theory of Rational Planning and Personal Persistence*. Oxford University Press.
- Slavny, Adam (2014). "Negating and Counterbalancing: A Fundamental Distinction in the Concept of a Corrective Duty." *Law and Philosophy* 33 (2): 143–173.
- Slote, Michael and Pettit, Philip (1984). "Satisficing Consequentialism." *Proceedings of the Aristotelian Society* 58: 139–176.
- Smilansky, Saul (2013). "A Difficulty Concerning Compensation." *Journal of Moral Philosophy* 10: 329–337.

- Steinhoff, Uwe (2016). "The Liability of Justified Attackers." *Ethical Theory and Moral Practice* 19 (4): 1015–1030.
- Stemplowska, Zofia (2009). "Making Justice Sensitive to Responsibility." *Political Studies* 57 (2): 237–259.
- Tadros, Victor (2011). *The Ends of Harm: The Moral Foundations of Criminal Law*. Oxford University Press.
- Tännsjö, Torbjorn (1992). "Who Are the Beneficiaries?" *Bioethics* (6): 32–55.
- Taurek, John (1977). "Should the Numbers Count?" *Philosophy and Public Affairs* 6: 293–316.
- Thomson, Judith Jarvis (1986). "Some Ruminations on Rights." In W. Parent (ed.) *Rights, Restitution, and Risk*. Harvard University Press.
- Thomson, Judith Jarvis (1990). *The Realm of Rights*. Harvard University Press.
- Unger, Peter (1990). *Identity, Consciousness, and Value*. Oxford University Press.
- van Inwagen, Peter (1990). *Material Beings*. Cornell University Press.
- Velleman, David (1991). "Well-Being and Time." *Pacific Philosophical Quarterly* 72 (1): 48–77.
- Vallentyne, Peter (2009). "Responsibility and Compensation Rights." In De Wijze, S., Kramer, M. H., and Carter, I. (eds.) *Hillel Steiner and the Anatomy of Justice: Themes and Challenges*. Routledge.
- Vincent v. Lake Erie Transportation Co.* (1910). 109 Minn. 456, 124 N.W. 221.

- Wiggins, David (1998). "Claims of Need." In his *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford University Press: 1–58.
- Williams, Andrew (2006). "Liberty, Equality, and Property," in J. Dryzek, B. Honig, and A. Phillips, (eds.) *The Oxford Handbook of Political Theory*. Oxford University Press.
- Williams, Bernard (1970). "The Self and the Future." *Philosophical Review* 79 (2): 161–180.
- Williams, Bernard (1979). "Internal and External Reasons." In R. Harrison (ed.) *Rational Action*. Cambridge University Press: 101–113.
- Zeman, Adam (2001). "Consciousness." *Brain* 124: 1263–1289.