Introduction

Our actions have the power to change the world. Now, of course, our actions usually only change the world in small, fairly meaningless ways. As I wrote this introduction, I chose to drink a cup of tea. In doing so, I brought about a whole new world – or *state-of-affairs* – one in which my kitchen cupboard was down one bag of tea, there was a little less milk in the fridge, and I was caffeinated and satisfied. But, sometimes, our actions can bring about wildly different worlds.

Walking down the street, I often meet homeless people who are asking for money to help them pay for shelter for the night. Indeed, I suspect many people share this experience. The choices before me, to donate or to keep my money, will bring about very different worlds. If I choose to keep my money, those asking for my assistance may lack a safe shelter for the night. If I donate, I will lose £10 – a sum which, in the bigger picture, means little to me – and a fellow person will have warmth and safety for the evening. To many, it seems obvious that the world in which I am £10 poorer and another person has shelter for the evening is surely the *better* one. This thesis, however, is about the moral reasons we have to bring about worse worlds.

1. What Does it Mean to "Do Best"?

Common sense tells us that our actions can bring about different states-of-affairs, and these different states-of-affairs can be better or worse than one another. Better or worse in what respect? Well, as with most things in philosophy, there is considerable disagreement over the answer to this question. In the following pages, I am going to gesture at the account of betterness which I have assumed throughout my thesis. I will also point to the places at which other philosophers may depart from me.

When philosophers claim one state-of-affairs is better than another, what many of them mean (or, at least, what I think they mean) is that one state-of-affairs contains more *value* or *goodness* than the other. Call this the *goodness account* of betterness, in which betterness merely is the comparative of goodness – better is to good what heavier is to weight. Such an account is heavily inspired by the

methods of economics, with the betterness relation often being treated as – in some ways – analogous to the preference relation (Broome, 2016). Betterness can be shown to have a variety of properties fairly easily. Firstly, it is thought of as a *transitive* relation; if world A contains more goodness than world B, and world B contains more goodness than world C, then surely world A contains more goodness than world C. Secondly, betterness is *asymmetric*; if world A contains more goodness than world B, then it cannot be the case that world B contains more goodness than world A. Thirdly, betterness may be *complete*. It is not implausible that all states-of-affairs are such that they can be compared to any other states-of-affairs in terms of whether they contain more, less, or the same amount of goodness. Combining such properties, betterness provides us with a total order over the set of states-of-affairs in terms of the amount of goodness they contain.[1]

Whilst this may be a common view of betterness, it certainly is not the only one. Another prominent view – that I will call the *reasons account* – understands betterness as a claim regarding the reasons we have to want a state-of-affairs to be instantiated. Perhaps the most famous proponent of this account, Larry Temkin, analyses betterness as such: when a state-of-affairs, A, is said to be better than a state-of-affairs, B, that means that one would consider A to have more reason to want it to be realized than B (2012: p.13).[2] Rather controversially, in Temkin's view, betterness is not a transitive relation (2012: ch.2, ch.6-7).[3] As such, betterness would not readily allow us to order states-of-affairs.

What I have just said is but the shallowest gloss of the debate between the goodness account and the reasons account of betterness, alongside those over the logical properties of the betterness relation. Throughout the thesis, I will be assuming the goodness account of betterness – luckily, for my purposes, I do not believe anything truly substantial hangs upon it.

[1] As a proponent of this view, John Broome has produced much work on the logical properties of betterness under this conception (2004; 1993).

[2] The conception of betterness can be found in the work of other philosophers, including Derek Parfit, who claims "good, in . . . the reason-implying sense, we mean roughly that there are certain kinds of fact about this thing's nature, or properties, that would in certain situations give us or others strong reasons to respond to this thing in some positive way, such as wanting, choosing, using, producing, or preserving this thing." (2011: 38)

[3] For discussion of the transitivity of betterness, see Stuart Rachels (1998), John Broome (2004), Michael Huemer (2008), Alex Voorhoeve (2013), Toby Handfield (2014), Shelly Kagan (2015), Jacob Nebel (2018), and Gerard Vong (2018).

On the goodness account, betterness is really a rather uninteresting notion. More interesting is what accompanies such an account – an *axiology*, a theory of goodness. There are many proposed axiologies, but there are a number of features which are common across them. First, they are frequently *welfarist* – value consists entirely in the wellbeing of individuals.[4] Second, they tend to be *impartial* – that is, equal improvements to any person's wellbeing are valued equally.[5] Third, they tend to be *aggregative*.[6] That is, most axiologies analyse the overall value of a state-of-affairs in terms of sum form of aggregate of the wellbeing held across individual entities in said state-of-affairs.[7] Typically, the aggregation method used is simple addition.

None of these characteristics are uncontroversial. Whilst almost all accept that wellbeing contributes to value, many reject the claim that wellbeing is *only* thing that contributes to value. Such *value pluralists* might believe, for example, that entities which lack wellbeing can non-instrumentally contribute to the goodness of a state-of-affairs.[8] The existence of the most brilliant piece of art, even if no one ever saw it or knew of it, would somehow make the world better.[9]

Others reject the impartial conception of goodness; some might, for example, reject the claim that it is sensible to talk of a state-of-affairs being impartially good, *simpliciter*. Rather, such philosophers might emphasise speaking in terms of states-of-affairs being *good for* someone in particular.[10] Likewise, one might believe that

[4] See L. W. Sumner (1996), Andrew Moore and Roger Crisp (1996), Robert Shaver (2004), and Simon Keller (2009) for outlines of, and arguments for, welfarism. For further discussion of the nature of welfare – or, rather, wellbeing – see Richard Kraut (2007).

[5] For discussions and proposals of impartial ethics, see William Godwin (2001[1793]), John Stuart Mill (1998[1861]: ch.5), Iris Murdoch (1970).

[6] For a discussion of the aggregative nature of goodness, Larry Temkin (2012), Gustaf Arrhenius and Wlodek Rabinowicz (2015), Wilkinson (2023).

[7] Theories of value are also sometimes, though not always, prioritarian – that is, they claim that improving the wellbeing of someone with antecedently lower wellbeing level is more valuable than improving the wellbeing of someone antecedently doing better off by the same amount. Parfit's seminal discussion of prioritarianism can be found in his paper 'Equality and Priority' (1997). For criticisms and defenses, see Parfit (2012), Ingmar Persson (2001), and Wlodek Rabinowicz (2001), Thomas Porter (2012), Michael Otsuka and Alex Voorhoeve (2009; 2011; 2018), and Roger Crisp (2011). In this thesis, I don't make any general assumptions about whether goodness has this property.

[8] Notable value pluralists include W.D Ross (1930), Jonathan Dancy (2004), and Judith Jarvis Thomson (1997). For work on how different types of values come together to contribute to goodness, see Michael Stocker (1989), Ruth Chang (2004), and Brian Hedden and Daniel Muñoz (2023).

[9] Susan Wolf has a particularly interesting discussion of this view (2010).

[10] For discussions of the relationship between 'good simpliciter' and 'good for', see John Rawls (1971), Philippa Foot (1985), and Judith Jarvis Thomson (2008).

there are morally robust forms of partiality – such as partiality to one's friends and family – which make the world a more valuable place.[11]

Finally, some reject the aggregative nature of goodness. This rejection can come in multiple forms, ranging from merely rejecting the straightforward additive conception to rejecting any form of aggregation at all. One example which motivates some to reject aggregative accounts of goodness comes from Derek Parfit:

> "For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living" (1984: 388).

If the total goodness of a state-of-affairs simply is the sum of the wellbeing of all its members, then, rather trivially, this implies that a world where there are fantastically many people in existence, all living lives of near-but-not-quite dreadful quality, is going to be better than a world in which a fewer number of people are living flourishing, happy lives. It implies that there is more value or goodness to be found in the first world. Most people have the intuition that this is just simply wrong, and, thus, our analysis of goodness is wrong. To avoid such a *repugnant conclusion*, one must either reject the welfarist conception of goodness, or amend their method of aggregating wellbeing into a measure of value.[12]

Without, hopefully, undermining the importance of the debates sketched above, I will, for the most part, ignore them. I will be assuming a straightforward welfarist, impartial, and aggregative account of goodness throughout the thesis. I believe the arguments that follow can be readily translated into competing axiologies, *mutatis mutandis*.

[11] For detailed criticisms of impartiality see Bernard Williams (1973), Michael Stocker (1976), Thomas Nagel (1986), Simon Keller (2013).

[12] Amongst the most popular proposals are average wellbeing accounts (R.I. Sikora, 1975), variable value accounts (Thomas Hurka, 1983; Ted Sider, 1991), and critical level accounts (John Broome, 2004; Derek Parfit, 2016). I briefly discuss problems for anti-aggregative axiologies in *Chapter Five*.

2. Goodness in action

Our ability to shape the world for good, or for bad, has retained a grip upon the imaginations of philosophers and non-philosophers alike. So much so that in recent years a variety of social movements have developed centred around the wish to use the resources available to us to bring about the best worlds possible.

Aptly named, *effective altruism* is a social movement concerned with the project of using evidence and reason to determine how to do the most good (MacAskill, 2019: 13). This social movement, however, is underpinned by a philosophy consisting of two key claims: firstly, that we are morally obligated to do far more to help those who are badly off than most currently do; and, secondly, that we are morally obligated to help optimally.[13] That is, in helping, one is required to do so in a manner that brings about the most good.

Effective altruists point out that there is a great variance in cost-effectiveness amongst the interventions available to us to help others – for instance, with the same amount of financial investment, you can expect bed nets to save many more lives from malaria than some of the most cost-effective cancer treatments.[14] As such, a choice to place your resources into a less cost-effective charity or intervention represents a choice to bring about significantly less goodness than you could have otherwise.

Developing as an offshoot of effective altruism, *longtermism* is the latest development in this line of philanthropic thinking. Longtermists, in general, highlight the overwhelming importance of positively shaping the far future. They point out that by improving the prospects of those living in the very far future – say, one thousand, ten thousand, or even one hundred thousand years from now – we can, in expectation, bring about a phenomenal amount of good.[15] Indeed, in

[13] For the first of these claims, see, in particular, Peter Singer (1972), Peter Unger (1996), Travis Timmerman (2015), Anton Markoč (2020), Jordan Arthur Thomson (2021). For presentations, critiques and defenses of the second claim see, Will MacAskill (2015), Iason Gabriel (2016), Alexander Dietz (2019), Antonin Broi (2019), Brian Berkey (2021).

[14] For a more detailed overview of the different in cost-effectiveness between available health interventions, see Toby Ord (2019).

[15] Proponents of longtermism include: Nick Bostrom (2003, 2013), Nick Beckstead (2013), Toby Ord (2020), Andreas Mogensen (2020), and Hilary Greaves and Will MacAskill (2021).

expectation, long-term interventions bring about far more good than those interventions that tend to those living now or soon.

The view that we can bring about the most good by focusing on improving the prospects of those living in the distant future has been motivated, in large part, by the observation that humanity's future might be vast in size. Humankind's existence could, for example, be extended by colonizing other planets, which we could inhabit for up to one hundred trillion years (Adams, 2008: 39; Beckstead, 2019: 82). Crucially, given the potential vastness of humanity's future, in expectation, the number of future persons to come will be enormous. Hilary Greaves and Will MacAskill suggest that under very conservative calculations there are at least one hundred trillion people to come, in expectation, and that a more reasonable estimate would be at least one quadrillion (2021: 9). It is this observation which gives rise to the claim that those interventions which most improve the prospects of future people will bring about the most good. Consider:

> **Catastrophic Risk.** Arthur the philanthropist is interested in two possible interventions. With the money available to him, he could fund a medical treatment for ten patients. All of the patients have exhausted all other treatment options. If they do not have this procedure, they will all certainly die. The procedure has a high success rate, making it almost certain that all ten will survive. Arthur's other option is to put his money into artificial intelligence (AI) security research. Arthur has been advised that a country of one hundred million people are currently at a one in a million risk of an AI-related fatal catastrophic event happening within their lifetimes. He has also been advised that his donation will reduce the risk of such an eventuality to one in two million.

In Catastrophic Risk, Arthur is faced with a choice. One option available to him is to treat a small number of people to substantially reduce their risk of a harm – that is, the group of ten. A second option is to reduce the risk of a harm for a vast number of other people by a very small amount – the group of one hundred million people. Even though by intervening in the large group, Arthur would only help

each individual by a tiny amount, this would, in expectation, do a lot of good. Indeed, Arthur would save fifty people – forty more lives than he would be investing in the treatment of the patients.

The choice between short-term and long-term interventions, in many cases, is very much like the choice faced in Catastrophic Risk between treatment and AI security research. Long-term interventions, like AI security research, generally only seem to improve the prospects of any far-future person by a tiny amount. This is due to the relative uncertainty with which we can shape the far-future; we are just not terribly good at bringing about far-future outcomes with much certainty. So, when it comes to long-term interventions, we can only ever raise the likelihood of bestowing a benefit to any future person by a tiny amount. On the other hand, at least some – if not a great many – of the short-term interventions available to us can substantially improve the prospects of their beneficiaries, given the relative certainty with which we can shape the present.

But long-term interventions clearly have something in their favour. Given the enormous number of individuals in the future who would receive a tiny increase in the chance of benefit, in expectation a *vast* amount of good would be bestowed by improving their prospects. This has two important implications. First, in expectation, the amount of good long-term interventions bring about is many orders of magnitude greater than the good brought about by short-term interventions. And, second, note that even the *slightest* increase to the prospects of future people generates a vast amount of good, in expectation. As such, long-term interventions which improve the prospects of future people by only a tiny amount more than alternative interventions are still vastly superior in terms of the amount of good they will bring about, in expectation.

So, it seems that longtermism provides us with reason to believe that we will bring about the most good, by an enormous margin, by investing in those interventions which *most* improve the prospects of future people. This claim, and claims similar to it, have been called *axiological longtermism*.[16] Given axiological

[16] Axiological strong longtermism has been questioned, spawning a literature on expected utility theory and fanaticism, see: Nick Bostrom (2009), Dylan Balfour (2020), Nick Beckstead and Teruji Thomas

longtermism, many argue we have compelling, if not decisive, moral reason to choose such interventions.[17]

3. From the Good to the Right

What about those of us who do not simply care about bringing about the most good? Surely, only philosophers would be content to tinker around with orderings of worlds. Most people, I presume, do not particularly care about states-of-affairs. What they do care about, amongst a multitude of things, is their actions and what it takes to live a morally decent life. Such people are not going to be satisfied by axiological longtermism alone. They might care about *deontic longtermism*, the claim that we have a moral obligation to invest in those interventions which most improve the prospects of future people.

Most people – though not all – admit that the consequences of our actions matter, morally speaking. I agree with this sentiment. A moral theory which was totally detached from the outcomes of our actions would seem absurd. Where there is considerable disagreement, however, is in *how* the consequences of our actions matter. Consequentialists, famously, hold that the deontic status of our actions – whether they are *permissible*, *impermissible*, *obligatory* – is wholly determined by their consequences.[18] Perhaps the most famous, and the most execrated, consequentialist theory is classical utilitarianism, which tells us that in a decision situation, the option which brings about the most goodness is morally obligatory. All other options are impermissible.

Classical utilitarianism moves easily from axiological longtermism to deontic longtermism; if doing that which most improves the prospects of future people brings about the most goodness, then we have an obligation to do so. Classical

(forthcoming), and Hayden Wilkinson (2022). David Thorstad, likewise, questions the value of risking existential risk (forthcoming).

[17] Longtermism is typically supported by considerations of beneficence. For discussions of alternative reasons to aid future generations, see Samuel Scheffler (2018: ch.2 & ch.3).

[18] This is a very permissive definition of consequentialism, which is supported by Frances Kamm (2007) and John Broome (1991). Some advocate for a narrower definition, including, for example, an agent-neutrality requirement (Frances Howard-Snyder, 1994; Philip Pettit, 1997), though this is controversial (Douglas Portmore, 2001; 2003). There are also concerns that the permissive definition used in this thesis renders consequentialism meaningless, as alternative theories can be 'consequentialised' – see, Douglas Portmore (2009) and Campbell Brown (2011).

utilitarianism has, however, received a barrage of criticisms over the decades[19], and, in response, a series of new consequentialist theories have been proposed to sidestep such problems.[20] But their key components remain the same: the good and the right are tightly intwined – and the majority of consequentialist theories, for that reason, will easily arrive at deontic longtermism from axiological longtermism.

Non-consequentialists, on the other hand, deny that we can derive the deontic status of our actions from their consequences alone. For the non-consequentialist, two actions can have the exact same consequences, yet differ entirely in their deontic status. Nonconsequentialists claim that actions which fail to bring about the most goodness can be morally permissible. In such cases, those actions which bring about the most goodness might be *supererogatory*, that is they are both good and permissible, but not obligatory.[21] Indeed, some actions which bring about the most goodness might fail to even be morally permissible – so called *optimific wrongings*.[22]

The reasons that non-consequentialists point to, to explain this departure of the right from the good, range widely. Some point to *agent-centred prerogatives* – permissions to do less than best grounded in the agent's interests.[23] Often such interest revolve around avoiding certain agential costs. Take for example:

Falling stranger. A stranger is falling from a cliff, and you are below watching the horrible scene. You are alone, and only you have the power to do something to help the stranger. You know that if the stranger hits the ground unimpeded,

[19] For details of classic objections to utilitarianism, see John Rawls (1971), Bernard Williams (1973; 1981), Robert Nozick (1974), Susan Wolf (1982), Peter Railton (1984).

[20] There is, for example, rule utilitarianism (Allan F. Gibbard, 1965; John C. Harsanyi, 1977; Brad Hooker, 2000), motive utilitarianism (Robert Adams, 1976), virtue consequentialism (Julia Driver, 2001; Ben Bradley, 2005; Dale Jamieson, 2007; Robert J. Hartman, 2015), satisficing consequentialism (Michael Slote and Philip Pettit, 1984; Tim Mulgan, 1993; Ben Bradley, 2006; Richard Yetter Chappell, 2019; Joe Slater, 2020), and consequentialism with prerogatives (Douglas W. Portmore, 2011; 2019).

[21] There are many competing accounts of what makes an action supererogatory, though the most popular ones ground supererogating in excessive agential costliness, see Alfred Archer (2016), Claire Benn (2014, 2018), Iskra Fileva and Jonathan Tresnan (2021), Elizabeth Drummond Young (2015), Dale Dorsey (2013). For problems facing accounts of supererogation, see Frances Kamm (1985, 1996), Claire Benn (2017), Daniel Muñoz (2021), Joe Horton (2017), and Alfred Archer and Michael Ridge (2015).

[22] For recent discussion about the status of optimific wrongings, see Andreas Mogensen (2016) and Thomas Sinclair (2017).

[23] The seminal defence of such prerogatives comes from Samuel Scheffler (1982; 1985). See Shelly Kagan (1989) for further discussion.

> he will certainly die. However, you can fling yourself into his path, cushioning his landing. Doing so would, remarkably, allow the stranger to survive injury free. You, on the other hand, will survive, but have your legs and arms broken and crushed.

Most agree that saving the falling stranger is not morally obligatory; surely you would not be an appropriate subject of moral disapprobation if you failed to seriously maim yourself to save another. This is despite the fact that it would be *best* if you were to throw yourself in the way of the stranger; a world in which one person is non-fatally injured is, surely, better than a world in which one person dies. Saving the stranger is supererogatory.

In such a case, this permission to do less than best is grounded in the costs that doing best would impose upon the agent – even though the cost to the agent is strictly less than the impartial opportunity cost of doing less than best. It seems that agents are morally permitted to place a little more weight onto their own interests.[24] This weight, however, is limited; it certainly would be impermissible to fail to save another's life if it cost the agent merely a papercut to do so.

It has been argued that such permissions can be grounded in considerations other than agential cost. Theron Pummer, for example, claims that we have permissions to do less than best based on agential autonomy (2023).[25] This captures the compelling idea that morality ought not tell us what to do in every situation. It seems that for morality to be compatible with a flourishing life, it must permit agents to make meaningful choices. Permissions to do less than best grant us such choices. Next, consider the following:

> **Saving a friend**. You are at an almost empty train station with your friend, and just two other strangers. All of the sudden, a violent gust of wind pushes both

[24] This is but one explanation of what is going on in such cases – sometimes referred to as a multiplier account (Scheffler, 1982). On such accounts, an agent is morally permitted to assign some coefficient greater than 1, by which to multiply the costs to them when assessing their interests in comparison to others who have competing interests. For criticisms and alternative proposals, see Garrett Cullity (2004), Theron Pummer (2023, ch. 1), and Brad Hooker (2009).

[25] Or in Pummer's vernacular, there are permitting reasons based on autonomy (2023: pp.27-29).

your friend and the two strangers onto the tracks, in the way of an oncoming train. Your friend is tall, heavy man, and in the time it will take you to pull him off the tracks, you could pull both of the strangers off the track. You do not have time to save both your friend and the two strangers before the train fatally collides with them.

It seems to me – indeed, it seems to many – that if you were to choose to save your friend in this situation, it would be wholly morally permissible. Yet, it would not be the best thing to do. Like Falling Stranger, such a verdict might be explained by an appeal to the agent-centred prerogatives; an agent's partiality to their friend means that failing to save them generates *agent-relative costs*. The prerogative may permit the agent not to incur these costs. Again, this prerogative is limited – partiality cannot justify, for example, failing to save one hundred strangers instead of your friend.[26]

Some may disagree that in this case was have an instance of a moral permission. Such individuals might intuit that rather than having a mere permission to save our friend, we have an obligation. It would seem to many that deciding to save the two instead of your friend would be abhorrent – a cruel betrayal of your friendship. This brings us onto a second nonconsequentialist justification to do less than best: we seem to have *special obligations* to others by virtue of the relationship we have with them. Special obligations, sometimes called *associative duties*, are not merely owed to friends.[27] We plausibly have filial obligations towards our parents,[28] and our parents have parental obligations towards us.[29] Likewise, you seem to have more general familial obligations, held to your wider family. Medical doctors have special obligations to their patients, lawyers have special obligations to their

[26] This is by no means the only justification of partiality. John Cottingham gives a justification of partiality based on its contribution to something intrinsically valuable, love (1986). It must be noted that such a justification makes use of a generally value pluralistic position, which I place to one side throughout the thesis. Diane Jeske provides a contractual account of partiality (1998). Other influential accounts come from Philip Pettit and Robert Goodin (1986). Thomas Douglas discusses how partiality might apply to future people, namely future children (2019).

[27] For a general discussion and defence of associative duties, see Jonathan Seglow (2013).

[28] See Simon Keller (2013) and Brynn Welch (2012), and Saba Bazargan-Forward (2019).

[29] For discussion, see Diane Jeske (1998), David Archard (2010), and Nellie Wieland (2011). Such parental obligations might extend to future children as well – see Hillvard Lillehammer (2009.

clients.[30] More controversially, fellow countrymen may have special obligations to one another.[31]

Nonconsequentialist also highlight moral *constraints* and *prohibitions* as reasons to do less than best. There seem to be some actions which are morally impermissible, regardless of the goodness they might bring about. [32] A well-used example is the torture of an innocent; most intuit that it would be impermissible to torture an innocent person, even if that was the only way to prevent a catastrophe which killed hundreds.[33] Another constraint nonconsequentialists point to is *the mere means principle* which tells us that it is impermissible to treat individuals as mere means, rather than an end in themselves.[34]

What do such nonconsequentialist considerations mean for longtermism? Let us begin by considering the principal argument for deontic longtermism is the *stakes-sensitivity* argument (Greaves and MacAskill, 2021: 27). The argument, generally speaking, goes as such:

1. *Axiological longtermism*. In a large range of decision contexts, the option which most improves the prospects of future people will bring about the most good, of all available options, by a significant degree.
2. *Stakes-sensitivity principle*. In a decision context in which one option brings about the most good, of all available options, by a significant degree, then we are morally obliged to choose that option.
3. *Conclusion*. In a large range of decision contexts, we will be morally obliged to choose the option which moves improves the prospects of future people (*deontic longtermism*).

[30] Rosamond Rhodes and Michael Danziger (2017).
[31] See Margaret Moore (2009), Richard Child (2011), and Seth Lazar (2013).
[32] A related but distinct discussion turns around how to conceive individual's rights, such as the right not to be killed, and so forth. Robert Nozick famously argues that such rights ought to be thought of as side constraints – things which limit of morally permissible behaviour – as opposed to as goods to maximise (1974, pp.48-50). As such, it is immoral to kill to, for example, prevent a greater number of murders.
[33] Indeed, many believe the prohibition on torture extends not non-innocents. For discussion, see Vittorio Bufacchi and Jean Maria Arrigo (2006), Jessica Wolfendale (2006), Bob Brecher (2007), Jamie Mayerfeld (2008), and Michelle Farrell (2013).
[34] Originally taken from Immanuel Kant, discussion of the principle can be found in Derek Parfit (2011: ch. 9), Victor Tadros (2011), Samuel Kerstein (2009), Warren S. Quinn (1989).

The general idea underpinning premise two is that the reasons we have to do less than best are sensitive to the size of the opportunity costs. Consider again agent-centred prerogatives and special obligations; I might be permitted to save my mother's life instead of two strangers life, but this permission seems to evanesce when I am presented with the option to save, say, one hundred strangers. It seems that, non-consequentialist considerations can justify doing slightly less than best, or even moderately less than best even. But, according to the stakes-sensitivity argument, nonconsequentialist conditions do not have the power to justify doing significantly less than best – and failing to improve the prospects of future people as much as possible is exactly that, according to the longtermist.

There is, of course, at least one obvious problem which premise two. All but a few paragraphs ago, I explained that nonconsequentialists believe that are reasons not to engage in certain actions, regardless of how much good they bring about. These reasons take the form of constraints and prohibitions. If engaging in those interventions which most improves the prospects of future people violated such constraints, then nonconsequentialists would have grounds for resisting deontic longtermism. For the most part, however, longtermists are keen to point out that unless we think long-term interventions systematically violate moral constraints, then we are able to carve out a set of non-constraint-violating long-term interventions which we are obligated to engage in.[35]

If the conclusions of longtermists are correct, then our picture of beneficence ought to be dramatically revised. Longtermism might, for example, tell us to ignore those suffering around us in favour of attending to the needs of those who are yet to exist. This would be the case if those interventions which most improve the prospects of far-future people do not benefit individuals in the short term or do so to a significantly lesser extent than alternatives. Even if the goals of long-term and short-term interventions align, such that long-term interventions also substantially help those in need now, longtermism would suggest that it was the

[35] Whilst this is a sorely underexplored issue, there is some work on whether long-term interventions are constraint-violating, see Andreas Mogensen and Will MacAskill (2021) and Charlotte Unruh (2023; forthcoming).

value that such interventions brought about in the far-future which justified them, *not* their impact on those currently in need.

4. Overview

This thesis is about the reasons we have to do less than best – the reasons we have to bring about worlds far worse than we might have otherwise. It is also about what such reasons can tell us about longtermism, and indeed, what longtermism can tell us about those reasons. Longtermism, as I see it, is but another, albeit startling, *reductio* from the consequentialist camp. By taking arguments for longtermism seriously, along with the moral costs associated with embracing them, we can learn as much about nonconsequentialist ethics as nonconsequentialist ethics can shed light upon longtermism.

In Part I, I outline, extend, and defend a series of nonconsequentialist reasons to do less than best. In Part II, I criticise the ethical case for longtermism, exploring how nonconsequentialists might resist the stakes-sensitivity argument. I argue that there are a range of plausible nonconsequentialist considerations which significantly undermine the longtermist case.

In Chapter One ("*The Conditional Obligation*"), I present recent arguments proposed by Joe Horton and Theron Pummer that supposedly demonstrate that even on nonconsequentialist grounds there is a conditional obligation to bring about the most good. Taking each in turn, I object to, and ultimately reject, these arguments. In Chapter Two ("*Agglomeration and Agent-Relative Costs*"), I move onto a form of aggregation that happens within lives. I argue that *intrapersonal agglomeration* extends the force of agent-centred prerogatives, allowing us to do less than best more than typically thought. In Chapter Three ("*Risk and Aggregation*"), I discuss and motivate deontic scepticism about *inter-personal aggregation* and, in particular, partially-aggregative moral theories. I outline its application to cases involving risk, alongside the *ex-ante/ex-post* distinction. I then defend it from a pressing recent objection from Joe Horton. By way of doing this, I propose an argument for *ex-post* moral theories.

In Chapter Four ("*Listening to Future People*") I criticise a recent argument from Andreas Mogensen for the claim that the reasons pertaining to improving the wellbeing of future people cannot ground obligations. Following from this, I claim that even admitting obligations to future people, we have reason to believe that we should prioritise those currently in existence – something I will demonstrate throughout Part II of the thesis. To begin, I build upon the work presented in Chapter Two to demonstrate that morally permissible generational partiality can justify prioritising those in the short-term. In Chapter Five ("*The Complaints of Future People*"), I extend the reasoning offered in Chapter Three, arguing that from both an *ex-ante* and *ex-post* perspective, any degree of scepticism about aggregation is incompatible with deontic longtermism. Finally, in Chapter Six ("*Leaving the Present Behind*"), I discuss the moral significance of *dooming*, whilst also arguing that considerations of *ex-ante* equality provide us with at least a strong *pro-tanto* reason to prefer short-term interventions over long-term ones.

5. Methodological Notes

Readers may have already noticed that I often make use of thought experiments in my writing. There are at least two methodological notes I must make with regard to them. First, not all of my thought experiments are intended to "pump intuitions". Indeed, many are merely meant for the purpose of illustration. Any thought experiment which elicits a particular intuition for argumentative purposes will be highlighted as such.

Second, as is suggested by what I have just said above, I do take considered moral intuitions seriously. Whilst it is beyond discussion within the scope of this thesis, I am of the belief that any ethical theorising will contain intuitions at some level. Though some philosophers do not agree with me, many do. Much of the argumentation found in this thesis takes as its premises firm intuitions I have about the thought experiments presented. For those who disagree with the use of intuitions, or with the particular intuitions themselves, the arguments of the thesis can still be viewed as a series of useful conditionals – perhaps, in some cases, by working via *modus tollens*.