Utilitarianism and the Social Nature of Persons

Nikhil Venkatesh, UCL

PhD Thesis in Philosophy

Front Matter

Declaration

I, Nikhil Venkatesh, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis defends utilitarianism: the view that as far as morality goes, one ought to choose the option which will result in the most overall well-being. Utilitarianism is widely rejected by philosophers today, largely because of a number of influential objections. In this thesis I deal with three of them. Each is found in Bernard Williams's 'A Critique of Utilitarianism' (1973).

The first is the Integrity Objection, an intervention that has been influential whilst being subject to a wide variety of interpretations. In Chapter Two I give my interpretation of Williams's Integrity objection; in Chapter Three I discuss one common response to it, and in Chapters Four and Five I give my own defence of utilitarianism against it. In Chapter Six I discuss a second objection: the problem of pre-emption. This problem is also found in Williams, but has received greater attention through the work of other authors in recent years. It suggests that utilitarianism is unable to deal with some of the modern world's most pressing moral problems, and raises an internal tension between the twin utilitarian aims of making a difference and achieving the best outcomes. In Chapter Seven I discuss a third objection: that utilitarianism is insufficiently egalitarian. I find this claim to be unwarranted, in light of recent social science and philosophy.

My responses to Williams's objections draw upon resources from the socialist tradition – in particular, that tradition's emphasis on the importance of social connections between individuals. Socialists have often been hostile to utilitarianism, in part for socialist-inflected versions of Williams's objections. Thus, in responding to these objections I aim to demonstrate that socialist thought contains the means to defuse not only mainstream philosophy's rejection of utilitarianism but also its own, and thus to re-open the possibilities for a productive engagement between the two traditions.

Impact Statement

This thesis addresses the questions of how we should live, how we should think, and how we should organise society. As such it should be of interest to everyone.

More specifically, this thesis could be impactful both within and outside of academic philosophy. It should aid philosophers' understanding of utilitarianism, which is one of the most prominent philosophical theories, taught in almost all ethics courses and widely referred to in ethics research – though also widely rejected. In providing a novel defence of utilitarianism, it may encourage reflection upon, and maybe revision of, this rejection. It also contributes to several live philosophical debates: on how we should understand Williams's Integrity Objection, on the morals of collective action, and on the currency of egalitarianism. And it contributes to a growing body of literature that brings analytic philosophy into contact with socialist ideas.

Outside of academic philosophy, there is widespread interest in both utilitarianism and socialism. Utilitarianism is particularly influential in public policy and philanthropy. Socialism is advocated, in one way or another, by governments representing billions of people, prominent opposition parties in other large countries, and very many private individuals. Contributing to a better understanding of the two, and a productive engagement between them, therefore has the potential to change many lives.

Acknowledgements

Relatively speaking, writing a philosophy thesis is not a *tough* job. There are no physical dangers, no angry customers, no domineering bosses, and an uncommon degree of autonomy over how one spends one's time. When we are tempted to complain, we should remember this. At the same time, writing a philosophy thesis is *difficult*. It is a long project and a personal one, involving setbacks and self-doubt. It requires moments of inspiration followed by tedious periods of graft, sometimes distancing oneself from the project, and sometimes examining it in painstaking detail. It is emotionally taxing and it can be lonely. This means that it requires the support of other people.

As I will argue later in this thesis, any project depends upon countless others (mostly with much tougher jobs) with whom we interact through the global economy. I hope to have done some justice to their labour. I depended, too, on many people who are known to me, and whom it is important to acknowledge by name.

Firstly, it is unlikely that one would write a thesis about moral philosophy without being brought up in a household that encouraged care for others and nurtured free thinking, as well as providing examples of good lives, good things and good people. For this, and many other things, I am eternally grateful to my parents Katherine and Ashok Venkatesh, and my sister Jothi. And of course, such a household would not have existed without my wise and loving grandparents: Victor and Sushila Venkatesh, and Frank and Barbara Southwell.

Secondly, it is unlikely that I would have pursued graduate study in philosophy without the inspiration of two people who are sadly passed. Josh Parsons taught me as an undergraduate, and the moment that he encouraged me to read Derek Parfit's *Reasons and Persons* was crucial in drawing me down this path. Parfit's influence will be apparent in the text, but Josh, in the spirit of philosophical inquiry he taught me, was at least as important. A third more recently departed philosopher who influenced my work is Joseph Raz.

Another early influence was Teruji Thomas, who answered Josh's call to supervise my undergraduate thesis, and instilled in me a patient, careful approach to research. At UCL, I have benefited from supervisions with Douglas Lavin, Ulrike Heuer, Joe Horton and Han van Wietmarschen, and for one semester I was fortunate enough to visit the University of Michigan and work with Peter Railton. Each of them has brought a different and valuable angle to this project and to my overall development as a philosopher. The single most important influence on this thesis has been Véronique Munoz-Dardé, my primary supervisor, with whom I have been arguing for six years (and doubtless will for many more to come), and whose advice and support has been immensely beneficial. Other academics who have kindly given advice and comments on this thesis include Sarah Buss, Christopher

Woodard, Alexander Dietz, Mark Kalderon, Kacper Kowalczyk, Eliot Michaelson, Nilanjan Das and Alex Voorhoeve (who, with Ulrike Heuer, examined the MPhilStud thesis that is an ancestor to this one). I have also benefited from the generous professional guidance of Robert Simpson, and administrative support from Richard Edwards and Sophie Grant (at UCL) and Carson Maynard (at Michigan). I would also like to acknowledge various anonymous reviewers for journals who have commented on papers related to parts of this thesis – as well as, in advance, my examiners, Jonathan Wolff and Mike Otsuka, who voluntarily agreed to read the whole of it, and who had extremely interesting and constructive things to say about it.

For all the kind contributions of these people, one of the last things I recall Josh saying to me was that one learns more from one's fellow graduate students than from faculty. I was extremely lucky to arrive in a department filled with brilliant and friendly research students at UCL, and another at Michigan. I have learnt from all of them, but in particular I would like to mention: Bex Rowson, Kirstine La Cour, Luke Devereux, Catherine Dale, Lizzy Holt, Jessica Fischer, Showkat Ali, Polly Mitchell, Hannah Carnegy-Arbuthnott, Rowan Mellor, Victor Braga Weber, Filippa Rohnquist, Alice Harberd, Charles Jansen, Patrick Nogoy, Naomi Alderson, Jack Hume, George Britten-Neish, Hannah McHugh, Kasim Khorasanee, Michael Markunas, Julian Bacharach, James Laing, David Olbrich, Alex Murphy, Christabel Cane, Léa Bourguignon, Euan Allison, Benjamin Davis, Robin Wisse, Kenta Sekine, Kim Leontiev, Fiona Whittingham, Matilda Carter, Alec Hinshelwood, Leora Sung, Karri Heikkinen, Daniel Guillery, Lukas Fuchs, Pete Faulconbridge and Laura Silva (at UCL), and Ariana Peruzzi,[1] Sophia Wushanley, Margot Witte, Abdul Ansari, Malte Hendrickx, Jason Byas, Rebecca Harrison, Paul de Font-Reaulx, Adam Waggoner, Sumeet Patwardhan, Gillian Gray, Laura Soter and Angela Sun (at Michigan). I would also like to mention my UCL UCU comrades, especially Saffron East, Naomi Oppenheim and Catherine Dale (for a much-deserved second appearance in these acknowledgements). Beyond UCL and Michigan, I would like to thank participants in the transatlantic 'ethics network' (especially its organiser, Michal Masny of Princeton), and Cain Shelley (LSE), who provided valuable input on Chapter Six in particular.

My fellow graduate students not only informed and improved this thesis, they also provided much-needed moral support, sounding boards, solidarity and fun. In this last aspect, they have been joined by various excellent friends (you know who you are) and by my teammates at the King's Road Cricket and Social Club. Opportunities to concentrate on

[1] Ariana also did something supererogatory (according to any plausible moral theory) and allowed me to live with her whilst I was in Michigan.

something other than one’s thesis are vital and ‘the Road’ has provided many of these – even if my focus tends to be on the teas rather than the runs and wickets.

The state of academia being what it is, this thesis depended not only on people, but also on money. Alongside an embarrassing deal of unearned privilege, I have relied on funding from the London Arts and Humanities Partnership and a grant facilitated by Longview Philanthropy – whose grantmaker, Tyler John, has also been a valuable philosophical interlocutor. I was also supported through being awarded a Forethought Fellowship for Global Priorities Research by the Forethought Foundation. I also benefitted – in cash terms and more – from working as an Administrative Assistant at the Aristotelian Society, and I must thank my colleagues there: Rory Madden, Guy Longworth, Josephine Salverda and Holly de las Casas.

Lastly, I would like to thank my most important source of support and encouragement, someone who has heard far too many sighs, witnessed far too many crises of confidence and suffered far too many explanations of what ‘self-effacing’ means. She has kept me going in difficult times, and shown me the joys of a kind of commitment that (despite Chapter Four) I endorse whole-heartedly. Without her, writing this thesis would have been significantly more painful, and my life would be significantly more dull. This is my partner, Hannah Lovell. The circumstances of the past two years meant that, as she likes to say, she has been ‘more present in my thesis’ than she anticipated. Hannah: I wouldn’t want it any other way.

Table of Contents

Epigraph

‘this it is which, when once the general happiness is recognised as the ethical standard, will constitute the strength of utilitarian morality. This firm foundation is that of the social feelings of mankind; the desire to be in unity with our fellow creatures, which is already a powerful principle in human nature, and happily one of those which tend to become stronger, even without express inculcation, from the influences of advancing civilisation.’

- John Stuart Mill (2008c, 164)

Chapter One: Introduction

0. Abstract

In this introductory chapter, I define utilitarianism and articulate its appeal. Utilitarianism is not popular amongst contemporary philosophers, in large part due to several influential objections. In this thesis I will respond to three of them: Williams's Integrity Objection, the problem of pre-emption, and the objection from equality. My answers have a common theme, emphasising the social connectedness of persons and often drawing on socialist thought. I close this chapter by addressing some of the affinities and tensions between the utilitarian and socialist traditions: my hope is for a reconciliation.

1. What is utilitarianism?

In our lives, we are faced with choices between different options. An option, as Pettit defines it, 'is a possibility which the agent is in a position to realise or not' (1993, 232), such as one's performing an action. Making these choices is unavoidable. When I wake up, I have the option of showering before breakfast, and the option of eating breakfast before I shower. I must choose one if I am not to remain in bed forever (and the latter would also be a choice to realise an option). It is sometimes appropriate to consider such choices from the moral point of view. Perhaps I had promised my flatmate that she could shower first. Would it be immoral of me to break that promise? What if it would better prepare me for a very important meeting? Utilitarianism, as I define it, is the view that as far as morality goes, one ought to choose the option which will result in the most overall well-being.

Some more on that definition. I will sometimes talk of the option which is most conducive to well-being, or which maximises well-being, or is optimific. These are equivalent to 'the option which will result in the most overall well-being'. *Overall* well-being is the sum of each person's individual well-being, weighted equally. Sometimes uncertainty will be salient, in which case the most natural (though not uncontroversial[2]) extension of utilitarianism is that one ought to choose the option which maximises *expected* well-being. For most of this thesis, uncertainty – important as it is – will be bracketed for the sake of clarity. Lastly, my definition of utilitarianism excludes 'rule utilitarianism', 'motive utilitarianism' and the like, which hold (roughly speaking) that one ought to choose the option that would be permitted by the optimific set of rules, or performed by the person with optimific motives. It is more akin to what is typically called 'act utilitarianism'; however, my

[2] See Buchak (2017) for an alternative utilitarian treatment of uncertainty.

‘options’ ranges over not only acts, but also rules, dispositions, projects and so on[3] – though, again for simplicity, I will typically focus on acts.

2. Why utilitarianism? And why not?

I am attracted to utilitarianism for a number of reasons, some more philosophically respectable than others. I suspect that a large part of its appeal for me is due to its function as a stalking horse in many philosophical discussions, in particular those of introductory ethics courses. One upshot of this is that calling oneself a utilitarian is likely to provoke both understanding and disagreement in discussion with anyone who has studied philosophy. If one likes arguments, endorsing utilitarianism is a good way to get them. More philosophically respectable reasons for taking utilitarianism seriously are its theoretical virtues of simplicity (it reduces the whole of morality to a single principle), scope (it ranges over all moral questions) and fruitfulness (it generates an interesting set of new theoretical questions). More respectable still is the support utilitarianism has from a ‘veil of ignorance’ argument (Harsanyi 1977), and several other more recent formal results (Ng 2000; Gustafsson 2021; Easwaran 2021). Perhaps it is due to these virtues and arguments that even its detractors in philosophy admit themselves to be ‘haunted’ by it (Foot 1983, 273; Scanlon 1982, 267), and to think it worth discussing, if only to resist it.

But the main reason for my sympathies towards utilitarianism is that it corresponds with what I care about. I care about the real world, not some idealisation of it, and about the external world rather than the inner one. I care about changing that world, and in particular about changing how people in that world fare. I care about each and every person and want them to be happy. I am sceptical of the existence of the supernatural and the value of tradition, and of the moral import of anything beyond cause, effect and experience. Thus utilitarianism’s focus on overall well-being, and on the outcomes that options realise, is attractive to me. I am hopeful for progress towards a better way of living together: not as a law of nature but as a live possibility that we can produce. Thus I find utilitarianism’s adaptability to changed circumstances, its practical orientation and its promise of a standard relatively detached from parochial prejudices from which to evaluate societies and changes in a very general way appealing.

Of course, one could share these inclinations whilst rejecting utilitarianism, and many do. But the theory and the inclinations share an emphasis and basis, such that for someone with those concerns utilitarianism is at least *prima facie* attractive. However, utilitarianism

[3] Greaves (2020) calls this approach ‘global consequentialism’; Railton (1988) ‘valoric utilitarianism’.

is widely rejected by philosophers – beyond a few institutional bastions, it is more likely to be mentioned as a mistake to be avoided than a view to be endorsed.[4] The rejection is largely due to a number of influential philosophical objections. In this thesis I deal with three of them. Each is found in some form in Bernard Williams's 'A Critique of Utilitarianism' (1973).

The first is what has come to be known as the Integrity Objection, a notoriously elusive but highly influential intervention. In the Chapter Two I give an exegesis of Williams's version of this objection; in Chapter Three I discuss one common response to it, and in Chapters Four and Five I give my own defence of utilitarianism against it.

In Chapter Six I discuss a second objection: the problem of pre-emption. This problem is also found in Williams, but has received greater attention through the work of other authors in recent years. It suggests that utilitarianism is unable to adequately deal with significant global challenges in the modern world, and also raises an internal tension between difference-making and achieving the best outcomes. In Chapter Seven I discuss a third objection: that utilitarianism is insufficiently egalitarian.

The guiding thought of my responses is that recognising the social nature of persons makes utilitarianism more defensible. The result is a defence of a kind of utilitarianism imbued with insights from the socialist tradition.

3. Utilitarianism and socialism

I am attracted to socialism by similar considerations that attract me to utilitarianism. Whilst utilitarianism constructs moral value out of the lives of each and every person, equally weighted, socialism promises to do the same for the economy. And in its Marxist form at least, socialism's emphasis on the economy, broadly conceived as the material, over the supernatural and superstructural, is consonant with my concern for the real external world: the world of cause, effect and experience. Socialists tend to share utilitarianism's scepticism of the supernatural and the traditional, and aspire to progress towards something very different, in which we live together in ways that are better for all of us. By holding out this

[4] In recent years, utilitarianism has been somewhat resurrected via the 'effective altruism' movement, associated with such people as Peter Singer and William MacAskill. It is a sign of utilitarianism's low stock, however, that effective altruists – despite very often being utilitarians themselves – tend to couch their arguments in terms that may appeal to non-utilitarians, and to brand themselves not as utilitarians but as effective altruists. A recent survey suggests that less than a third of academic philosophers identify with – or even 'lean towards' – consequentialism, and utilitarians will be a subset of this group (Bourget and Chalmers ms).

hope, socialism also offers a wide critique of what exists. By struggling for that future, socialists have matched utilitarianism's focus on doing things, on changing things.

Both traditions, then, fit broadly into the philosophical programme described by Peter Railton – perhaps the contemporary philosopher most attuned to my temperament, and a fellow-traveller, at least, of both utilitarianism and socialism.

> 'From my early days of protest and making the transition from outsider to insider, I acquired a sense that moral issues are not only real, but palpable. By trying to make things different, one can learn not only where power lies and how it is exercised, but also who is hurt, how they're hurt, and what might make things better. I was drawn, therefore, to a form of realism that was not minimal or quietist, but naturalistic and reforming. It called for a metaphysics in which value can be found at the intersection of mind and world, and an epistemology in which such values can be learned through experience, good and bad. And it was a realism that gives a central role to consequences—to the differences made in people's lives. We're a long way from justice, and an ideal theory cannot tell us how to get from there to here. We need a theory that permits large-scale critique while also contribute to guiding the struggles that make for moral progress—always hard won and never complete or secure. And rarely, if ever, wholly by the rules.' (Railton 2015)

Why, then, is there often hostility between utilitarians and socialists? Both theories arose (in their modern forms at least) in the heat of the Industrial Revolution in northwestern Europe. In those early years of industrialisation, the affinity between the two was strong. Both were sceptical of custom and religion, and both held out the hope of combining the greater productivity of industry with the new social science to create a better future for all who lived in it. Early utilitarians and socialists often found themselves on the same side of policy debates: Bentham was in general for the toleration of minorities and democratisation of politics, for instance, and his followers such as Francis Place were instrumental in the legalisation of trade unions. Early British socialists such as William Thompson, Robert Owen and John Wade drew from Bentham's greatest happiness principle a critique not only of the old aristocratic society on which Bentham tended to focus his ire, but also of the emerging imperial industrial capitalism. Ultimately, though, the marriage was not to last. Utilitarianism became the theory of the ever more powerful bourgeoisie, and socialism of the working-classes – at the same time, the conflicts between these groups intensified. Intellectually, utilitarianism came under attack from critics of industrialisation, whose stances looked ever more attractive to those on the sharp end of it. As E. P. Thompson writes in *The Making of the English Working Class*, by the 1830s:

> ‘Such [working-class] men met Utilitarianism in their daily lives, and they sought to throw it back… In these same years, the great Romantic criticism of Utilitarianism was running its parallel but altogether separate course. After William Blake, no mind was at home in both cultures, nor had the genius to interpret the two traditions to each other… Hence these years appear at times to display, not a revolutionary challenge, but a resistance movement, in which both the Romantics and the Radical craftsmen opposed the annunciation of Acquisitive Man. In the failure of the two traditions to come to a point of juncture, something was lost. How much we cannot be sure, for we are among the losers.’ (1980, 915)

The failure was not total, and subsequently there have been some figures – John Stuart Mill, Harriet Taylor, Bertrand Russell, and many associated with the British Fabians and the Vienna Circle – who continued utilitarian-socialist engagement. But today most socialists are as (if not more) suspicious of utilitarianism than most philosophers. I follow Thompson regarding this as a loss, and this thesis aims to go a small step towards recovering some of it.

The loss, I think, was twofold. Firstly, as utilitarianism grew in establishment influence, it lost its radical social aspect. The Victorian utilitarians achieved some progressive ends, particularly in Britain and its empire. As Wiggins notes:

> ‘[Utilitarianism] involved itself… in campaigns for law reform, prison reform, adult suffrage, free trade, trade union legislation, public education, a free press, secret ballot, a civil service competitively recruited by public examination, the modernisation of local government, the registration of titles to property in land, safety codes for merchant shipping, sanitation, preventive public medicine, smoke prevention, an Alkali Inspectorate, the collection of economic statistics, anti-monopoly legislation… In sum, philosophical utilitarianism played a leading part in promoting indefinitely many of the things that we now take for granted in the modern world.’ (2006, 145)

Throughout the twentieth century utilitarianism provided the normative background of mainstream economics, and through it influenced policy across the world. Again, this has not been without success, even viewed from a socialist perspective. But on the whole utilitarianism has accommodated itself very well – as was already clear to the English proletariat of the 1830s – to the dominant structures of capitalism. In Bentham’s day this was, to a degree, a form of radicalism, as the remnants of feudalism were to be swept away. But now it appears, as Marx alleged, more like an apology for the existing state of things. Wiggins is able to associate twentieth-century utilitarianism with neo-colonialist trade policies that hampered the development of the Global South (2009, sec. 10), whilst

Srinivasan charges twenty-first century effective altruists with ‘speak[ing] in the proprietary language of the illness – global inequality – whose symptoms [they] propose to mop up’ (2015). Effective altruism, the contemporary offspring of utilitarianism, is very closely associated with the very rich, for all the good it has done for the very poor. Owen’s vision of a new moral world (1970) has slipped from view, as utilitarians have focussed their energies on gradualist, unthreatening improvements to the existing one.

The more important loss, probably, was in the development of socialist thought. The split with utilitarianism preceded Marx’s condemnation of morality in general. The socialist tradition was left without a well-developed moral theory, even as it attracted adherents who were motivated to alleviate suffering, challenge injustice, and create a better society. When socialists found themselves with state power, they faced difficult moral decisions; the kind of high-stakes trade-offs to which utilitarianism is especially good at giving answers. Without utilitarianism, they often fell back on the traditional moral values they had previously condemned, or on liberal theories of rights and justice which hampered their radical visions – or they invested authority not in morality but in a person or party or power itself – with disastrous results. The irony, as we have seen, is that socialist and utilitarian temperaments have a high degree of affinity. Utilitarianism’s emphasis on the material over the ideal should make it less vulnerable to Marxist attacks on morality as mere ideological fluff. Moreover, as Trotsky (1938) recognised, though utilitarianism may seem to be ‘the ethics of bourgeois bookkeeping’, its flexibility to circumstance, and in particular its capacity to justify means by their ends, make something suspiciously close to utilitarianism the best candidate for ‘our [i.e. Marxist] morals’. Today, socialist movements out of power are often in the grip of non-utilitarian moralities emphasising so many and such absolute rights and duties that they are paralysed by moralising internecine disputes.[5]

The split between socialism and utilitarianism has historical roots in politics, class and other social forces. But insofar as contemporary socialist suspicion of utilitarianism is motivated by philosophical reflection, my sense is that it is largely based on socialist-inflected versions of Williams’s objections. The Integrity Objection may be read as the charge that utilitarianism alienates people in much the same way capitalism does; collective action problems are held to show that in focusing on the individual as the unit of action (rather than, say, the class, state or party) utilitarianism neglects large-scale structural changes; utilitarianism’s alleged neglect of equality makes it to hospitable to a system that, for all the wealth it has created, generates deep and growing inequalities.

[5] Trotsky, of course, was the victim of an internecine dispute, but not a moralising one – rather, the kind of brute power conflict that emerges in the absence of moral conviction.

My responses to Williams's objections draw upon resources from the socialist tradition – in particular, that tradition's emphasis on the importance of social connections between individuals. Thus, I aim to demonstrate that socialist thought contains the means to defuse not only mainstream philosophy's suspicion of utilitarianism but also its own, and thus to re-open the possibilities for the kind of productive engagement between the two for which Thompson grieved.

Chapter Two: Williams's Integrity Objection

0. Abstract

In this chapter I reconstruct Williams's Integrity Objection. On what I find to be the best interpretation of it, Williams charges that the utilitarian agent's psychology makes commitment impossible for them. I consider various other interpretations of Williams and introduce concepts that will be important in later chapters, including the problem of pre-emption.

1. Williams's cases and the notion of a project

Williams's objection is made through two hypothetical cases (1973, 97–99). In one, a recently graduated chemist, George, is offered a job in a chemical and biological warfare laboratory. He decides that he cannot accept, since he is opposed to chemical and biological warfare (CBW). He cannot accept even though his unemployment causes him and his family to suffer, and even when he is told that the person who would be hired in his place would pursue the research in such a way that more dangerous chemical weaponry would result. In the second case, Jim, in a foreign land in the aftermath of an uprising, is made an offer by Pedro, an army captain. Pedro will execute twenty innocent prisoners as a warning to dissenters unless Jim agrees to shoot one himself, in which case the other nineteen will be released.

The structure of Williams's argument invites the interpretation that he aims to dismiss utilitarianism on the grounds that it gives Jim and George the wrong advice.[6] Utilitarianism recommends, given some tacit assumptions (that George wouldn't be so depressed by taking the job that he and his family are caused more suffering; that the development of chemical weapons is bad for well-being; that Pedro's prisoners have lives worth living) that George takes the job and Jim shoots the prisoner. Those who believe that one should never assist with CBW research, or kill, or 'sell out one's principles' will disagree. These recommendations alone will persuade such people against utilitarianism. But Williams does not want to persuade only people with such moral beliefs – and is not one of them himself, remarking that 'the utilitarian is probably right' in Jim's case (1973, 117). The cases are

[6] This is Hare's interpretation of Williams (1981, 49, 130–46). Williams himself disclaims it, writing: 'the objection did not, however, take the form of my trying to disprove a theory by counter-example, as much of the discussion has assumed.' (1995, 212) Thanks to Showkat Ali for bringing this remark to my attention.

meant to make salient a certain feature of moral life, consideration of which reveals utilitarianism to be defective. That feature is integrity.

Williams introduces his discussion of integrity by considering

> 'the idea, as we might first and simply put it, that each of us is specially responsible for what he does, rather than what other people do. This is an idea closely connected with the value of integrity. It is often suspected that utilitarianism makes integrity as a value more or less unintelligible. I shall try to show that this suspicion is correct.' (1973, 99)

He goes on:

> 'the reason why utilitarianism cannot understand integrity is that it cannot coherently describe the relations between a man's projects and his actions.' (1973, 100)

We can make out Williams's claim of a 'close connection' between integrity and responsibility, and hence his notion of integrity, in terms of these relations. There is, for Williams, a deep difference between how I relate to my actions and to those of other people, even when I can prevent or encourage the latter. This difference is at play in the cases of George and Jim:

> 'The situations have in common that if the agent does not do a certain disagreeable thing, someone else will, and in Jim's situation at least the result, the state of affairs after the other man has acted, if he does, will be worse than after Jim has acted, if Jim does. The same, on a smaller scale, is true of George's case.' (1973, 108)

Jim could shoot one person or reject Pedro's offer in which case Pedro will shoot twenty. George could advance CBW or reject the job, allowing his rival to advance it in more dangerous directions. Whether the second state of affairs is realised, in both cases, is in the gift of Jim and George. It is, in the sense introduced in the previous chapter, an option for them. But it would be misleading, according to Williams, to think of Jim and George as having brought about these states of affairs if they are realised. It would be misleading to think that they will merely have had an 'effect on the world through the medium… of Pedro's [or the unnamed rival chemist's] acts.' (1973, 109) In Williams's view their responsibility for those states of affairs is therefore lesser and qualitatively different; morality respects a distinction between my actions, and actions that are not mine but whose occurrence I have control over. What could account for this distinction? Williams points to the relationship between actions and projects. If George takes the job, he adopts the development of chemical weapons as a project, and accordingly conducts the relevant research. If he doesn't, the other chemist would adopt the same project, and pursue the same research, but there would not be the same relationship between George's projects and the research. (Indeed, George could retain his project of opposing CBW.) If Jim were to reject Pedro's offer, twenty people would die. But this would not be because Jim had a project that aimed at their deaths, but because Pedro did. Their deaths in this case would thus be best described as a killing by Pedro, not by Jim, for all the opportunity Jim had to save them. We are 'identified', as Williams says (1973, 116), with the actions that 'flow' from our projects.

What if Jim were to accept Pedro's offer? Although Jim would pull the trigger, it would seem wrong to hold him responsible. Pedro's coercion effectively turns Jim into a medium through which Pedro affects the world. This intuitive description of the case is reflected at the level of projects: the killing is the aim of Pedro's projects, not Jim's. When we perform actions which do not flow from our projects – as Jim does here – we are not identified with them: rather, we are alienated from them. Our responsibility for such actions is attenuated.

That 'each of us is specially responsible for what he does', then, seems to mean this: each of us is specially responsible for the actions that flow from our projects. They must flow from our projects in the right way: if my pursuing some innocent project outrages you so much that you lash out at me, I am not responsible for this, even though my project is part of the cause of your lashing out.[7] Rather, actions we are identified with must flow from our projects in a way that is directed by those projects, as Jim's shooting of a prisoner would be directed by Pedro's aim of intimidating dissenters. This – not the fact that we had the opportunity to determine whether the action was performed or not – is what makes those actions ours and not someone else's. To neglect this connection is to attack our integrity. 'Integrity' here is meant in the sense of wholeness or unity – an agent's integrity is the unity between them, their projects and their actions.

Utilitarianism seems to neglect this connection. As I have defined it, utilitarianism provides a criterion of choice between options: what one should do, according to utilitarianism, is determined by the effects on well-being of each option available to you. An agent's options are whatever they are able to realise: this is not limited to actions flowing from their projects. Furthermore, utilitarianism is indifferent between different paths to the same sum of well-being, as reflection on Williams's cases demonstrates.

Jim has two available options: (1) accept the offer and shoot one prisoner; (2) reject the offer and see Pedro shoot twenty. Choosing (1) will lead to more well-being than choosing (2), in normal circumstances (the prisoners will have lives worth living if they survive,

[7] Thanks to Veronique Munoz-Dardé for this caveat and example.

bereavement reduces well-being, and so on). So utilitarianism recommends (1). But notice that the very same reasoning would apply if Pedro were not in the picture. Imagine Jim had a choice between (1) shooting one person and (2') shooting twenty people himself. The effects on well-being are equivalent (except perhaps for differences in guilt felt by Jim and Pedro) across (2) and (2'). For utilitarianism, therefore, these choices are equivalent: Jim's choosing (2) is as bad as his choosing (2') would be. This is so even though in (2) the killings would flow from Pedro's projects, and in (2') they would flow from Jim's. Utilitarianism therefore seems to neglect the significance of the agent's identification with their actions through their projects.

What is a project for Williams? He gives no explicit definition. He gives examples (1973, 110–11): desires for oneself, one's family and one's friends to have the basic necessities of life, and for the 'objects of taste'; 'pursuits and interests of an intellectual, cultural or creative character'; political causes such as Zionism; 'projects that flow from some more general disposition towards human conduct and character, such as a hatred of injustice, or of cruelty, or of killing'; the utilitarian project of maximising well-being. A project, to be something from which action may flow, must be capable of motivating the agent who has it. The motivational aspect of projects is reaffirmed by Williams when he says that if we 'step aside' from our projects, we are alienated 'from [our] actions and the source of [our] actions in [our] own convictions.' (1973, 116) But projects cannot be whatever motivates action – a habit or addiction would not be a project. We are conscious that our projects guide our actions (unlike mere habit) and approve of them (unlike addictions).

Desires for basic necessities motivate us all, but political causes or moral convictions are more individual: only some of us are motivated by Zionism, or by justice. These things are projects for some people, but not for others. But all projects must be such that they are had in some unique way by each individual who has them, to do the work Williams puts them to in explaining integrity and responsibility. If there are two people who have Zionism as a project, and one, motivated by that Zionism, performs some action, then the other is not responsible for that action as if it were theirs. (We might think that they have some responsibility to condemn such an action if it is wrong or to defend it if it is right; perhaps one cannot think that actions done by others for the sake of a project you share are wholly 'none of your business'. But Jim also, as Williams says (1973, 110), cannot take Pedro's actions to be none of his business, and this does not make those actions his.) So, when he says that each of us is specially responsible for the actions that flow from our projects, Williams means that each of us is specially responsible for the actions that flow from our having our projects, not for actions that flow from other people's having projects we happen to share.

2. Utilitarianism and commitments

These two passages form the crux of the Integrity Objection:

> 'how can a man, as a utilitarian agent, come to regard as one satisfaction among others, and a dispensable one, a project or attitude round which he has built his life, just because someone else's projects have so structured the causal scene that that is how the utilitarian sum comes out?'
>
> 'It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own project and decision and acknowledge the decision which utilitarian calculation requires.' (both passages Williams 1973, 116)

In the second quoted passage Williams contrasts the agent's 'own project and decision' with the utilitarian recommendation. This is a false dichotomy. An agent could adopt utilitarianism itself as a project – indeed, Williams himself considers this possibility just one page before. For such an agent, actions performed as utilitarianism requires would flow from one of the agent's projects, and so would really be the agent's own. As long as such an agent is possible, there is no necessary opposition between identified action from one's own project and decision and acknowledging utilitarian recommendations.

However, as the cases of George and Jim show, utilitarian recommendations can conflict with other, non-utilitarian projects. Williams's protagonists cannot simultaneously follow the utilitarian recommendation and their projects of opposing CBW and refraining from killing. Williams alleges an absurdity in demanding that someone step aside from their projects. If to 'step aside from' a project is simply to perform some action antithetical to it, then what seems absurd to demand is that morality never ask one to step aside from one's projects. A project, as we have seen, could be a simple desire or taste. There are surely occasions in which we ought to forego satisfying one of our desires to help someone else from a greater suffering. Williams's objection would be implausible if it rejected 'any morality not based on the accident of the agent's own projects' (Raz 1986, 287).

But not all projects can be so easily put aside. Williams is especially interested in the subset of projects he calls 'commitments'. What distinguishes commitments from other projects is left vague, but has to do with the greater strength of the attitude one has towards them, hinted at by words like 'thorough', 'deep and extensive' and 'serious'. 'One can be committed', Williams writes, 'to such things as a person, a cause, an institution, a career, one's own genius, or the pursuit of danger.' (1973, 112) A commitment is not simply a very strong desire, though; it is a project which in some way defines the person who has it. Consider the desire to eat: when one is very hungry it may be overwhelmingly strong, but it is hardly something that defines one's character and shapes one's life. Williams writes that one could treat a cultural pursuit as a commitment. One's relationship to that pursuit would be 'at once more thoroughgoing and serious than their pursuit of various objects of taste, while it is more individual and permeated with character than the desire for the basic necessities of life.' (1973, 111) Enjoying the tune of some aria does not count as a commitment, even if it motivates you to go to an opera. Being an opera-lover, on the other hand, which involves educating oneself about the history and subtleties of the form, keeping oneself abreast of current productions, watching and listening to opera frequently, defending its value in argument, and so on, could be a commitment. Insofar as there is a distinction between an opera-lover and someone who enjoys the opera, it seems that for the former their relationship with opera has permeated their character, such as to become partly constitutive of their identity. If being an opera-lover is related to us in this way, and essentially involves certain actions, then performing those actions is essential to our being who we are. This means that a different level of integrity is at stake in the actions flowing from our commitments. Actions flowing from our projects are ours; actions flowing from our commitments are not only ours, they are us.

Is the integrity objection, then, that utilitarianism asks us to step aside from our commitments, and that this is impossible or too demanding, given their connection with our identity? If it is impossible to step aside from a commitment, this offers no objection to utilitarianism. To see why, consider the following objection to utilitarianism. Well-being would be maximised if Jeremy pushes a button that is labelled 'maximise well-being', and does as it its label suggests. But Jeremy does not have access to such a button, because no such button exists. So utilitarianism demands the impossible of Jeremy, and therefore is false. This argument is obviously ridiculous. It is ridiculous because utilitarianism does not ask us to push non-existent buttons. It does not do this because it does not ask us to do the impossible. Therefore, if stepping aside from one's commitments truly is impossible, utilitarianism does not ask us to do it. In this case there would be no conflict between utilitarianism and commitment.

Perhaps, though, we *can* put our commitments aside, but they are such that our doing so would be an unreasonable demand of a moral theory. This is Elizabeth Ashford's (2000) interpretation of Williams's objection. She responds that any plausible moral theory will ask us to step aside from our commitments in emergencies, and that – as Williams himself suggests elsewhere (1985a, 186) – such situations are common in the actual world, given the extremes of poverty and wealth in an ever more closely connected global economy. Thus, she says, though utilitarianism might ask us to put aside our commitments to save lives, this is not an unreasonable demand. I am very sympathetic to Ashford's view, and return to it in Chapter Four. However, I think Williams's objection goes beyond it. Utilitarianism is incompatible with commitment in other, deeper ways. Chief among these is a problem of the psychology of the utilitarian agent: this will be drawn out next, and the following two chapters will provide a response to it. Other problems suggested by Williams's objection: what I will call the 'normative', 'value-theoretic' and 'coherence' problems, are addressed more briefly below, as will the 'pre-emption problem', a distinct problem for utilitarianism arising from Williams's cases, and the subject of Chapter Six.

3. The psychological problem

Williams believes, I think, that if an agent accepts utilitarianism they will be incapable of having commitments. This is because of two facts about the psychology of the utilitarian agent: they will regard their projects impartially, and they will regard them as dispensable.

To regard projects impartially is to refrain from valuing one project more than another simply because of whose it is. Most importantly, someone who is impartial does not value their own projects more than those of others simply because they are their own. This captures Williams's description that the projects of a utilitarian agent are, to them, 'one satisfaction among others'.

Why does utilitarianism require such impartiality? As Scanlon puts it (1998, 95–100), to value X is to take oneself to have reasons for certain attitudes and actions towards X. In Scanlon's view these attitudes and actions may be several, and may vary depending on what X is. With projects, very generally, we can say that a crucial part of valuing them is taking oneself to have to pursue them (when they are one's own), and to assist in them (when they belong to others). To value project Y over project Z, then, would be to take oneself to have stronger reasons to pursue/assist in Y than Z. For a utilitarian (as far as morality goes) one has stronger reasons to do one thing than another if and only if it would result in greater overall well-being. I should do what leads to more well-being, whether that involves acting on my projects or those of a stranger. But if I value my own projects more, simply because they are mine, I take myself to have more reason to pursue them than to assist in others, even when doing the latter would result in greater well-being. Such an action would be wrong, according to utilitarianism.

Insofar as one employs utilitarianism as a decision-procedure, then, one regards one's projects impartially. Whether a utilitarian agent takes themselves to have stronger reason to pursue/assist in one project rather than another does not depend on whose projects they are, but simply on the well-being that would result from one's pursuing or assisting actions. But if I deliberate in this way, how is my project my project? It seems obvious that if X is my project and Y is not, I must regard X in a different light to Y (typically as more valuable) and be generally disposed to act on X rather than Y. So for the agent who accepts utilitarianism, their projects seem to have a double life: they are both that agent's projects, special to her, and they are, according to utilitarianism 'one satisfaction among others'. The utilitarian agent's actions do not flow from these projects, but rather from well-being calculations that take everyone's projects into account on an equal basis. This, I think, is what Williams means by alleging that utilitarianism 'cannot coherently describe the relations between a man's projects and his actions'.

Now, Williams writes that 'in the case of many sorts of projects' it is 'perfectly reasonable' to weigh the utility gains of your satisfying your project against the gains of someone else satisfying theirs when the two conflict (1973, 115–16). This not only permits a moral theory to ask us to abandon our projects on occasion, it also affirms the utilitarian's impartial method of counting one's projects as 'one satisfaction among others'. But Williams thought such impartiality was impossible with respect to commitments, even if reasonable for other projects. As he puts it concerning a subset of commitments, moral convictions:

> 'we… cannot regard our moral feelings merely as objects of utilitarian value… to come to regard those feelings from a purely utilitarian point of view, that is to say, as happenings outside of one's moral self, is to lose a sense of one's moral identity; to lose, in the most literal way, one's integrity.' (1973, 103–4)

I will argue in Chapter Five that such impartial attitudes are in fact consistent with a healthy attitude towards the self. Williams's objection in the first instance, however, is that they are incompatible with commitment: if one is committed to a project, one cannot regard it impartially – one must value it more than those of others, simply because it is one's commitment. This I will grant.

So much for impartiality (for now). What is it to regard a project as dispensable – and why does utilitarianism require it, and commitment preclude it?

It is important to note that regarding is an attitude, not an action or disposition to action. It is not that one *can* never dispense with a project to which one is committed. (Though a commitment may give rise to what Williams terms 'moral incapacity' (1992) or 'practical necessity' (1981c), which is the sense in which it is true that George *cannot* take the job – see next section for more discussion.) Nor is it that one *should* never dispense with one's commitments (Williams thinks Jim should do just that). The modality involved is about how agents see their possibilities: as Williams says, his cases show 'most importantly of all, what would be implied by certain ways of thinking about the situations'. (1973, 96)

What is it to regard a project as dispensable? For Williams, I think, *to regard a project as dispensable is to entertain as alternatives outcomes in which one dispenses with it*.

What is it to dispense with a project? One might imagine that it involves ridding oneself of positive attitudes towards it and the associated patterns of motivation. But that would not address the cases of George and Jim. They are, Williams thinks, asked by utilitarianism to dispense with their commitments. But they are not prevented from continuing to believe in the wrongness of CBW, or of killing, nor from living their lives, beyond these tragic episodes, in accordance with those beliefs. Thus for Williams to dispense with a project does not necessitate fully ceasing to believe in and pursue it. Performing certain one-off actions that are to a sufficient degree at odds with it – such as killing, with respect to the commitment not to kill ('stepping aside' from the project, as Williams puts it) – also counts as dispensing.

Now, if agents were unable to conceive of circumstances in which they would dispense with their commitments, commitments would be obviously morally unattractive. For any project you have, I can ask you to imagine that Satan has promised to wreak untold suffering on humanity if you do not dispense with it. If you held that even in such circumstances you would not dispense with their commitments, you would appear to be not principled but dangerously fanatical.

Williams's position was not as implausible as this. By 'entertaining as alternatives' he does not mean merely conceiving of them. He writes:

> 'it could be a feature of a man's moral outlook that he regarded certain courses of action as unthinkable, in the sense that he would not entertain the idea of doing them… Entertaining certain alternatives, regarding them even as alternatives, is itself something he regards as dishonourable or morally absurd. But, further, he might equally find it unacceptable to consider what to do in certain conceivable situations. Logically, or indeed empirically conceivable they may be, but they are not to him morally conceivable, meaning by that that their occurrence as situations would represent not a special problem in his moral world, but something that lay beyond its limits.' (1973, 92)

If Jim and George are committed to their respective projects of not killing and of opposing CBW, Williams suggests, the situations in which they find themselves require a very different way of thinking to that which they employ in other situations. It is not that they cannot, or don't like to, think about situations in which they have to kill, or do military research. It is that their commitments circumscribe a set of alternatives that they are willing to entertain, and this is partly constitutive of their outlooks on life. (In the case of moral commitments, entertaining such alternatives may be 'dishonourable', the situations 'morally inconceivable'. But not all commitments are moral ones, for Williams: situations may perhaps be inconceivable with respect to an agent's politics, aesthetics, or other values.) Their commitments are usually inputs, or constraints, on their deliberation. In the kind of situations in which Jim and George are placed, they find themselves required to deliberate without them, as the commitments themselves are up for debate. What was solid in their thinking melts into air; they are compelled to question what was previously bedrock. This is what makes taking the job, for George, seem 'absurd'.[8]

So for Williams, entertaining some outcome as an alternative is not merely conceiving of it. It is being willing to conceive of it within the constraints set by one's outlook on life. Commitments set such constraints: a committed agent is unwilling to conceive of outcomes in which they dispense with their projects. This does not mean they never do, but that when they do, a novel and (to them) unsettling mode of deliberation is required.

This is not the case for the utilitarian agent. It is a distinctive (and, to some, attractive) feature of the utilitarian outlook on life that it does not shirk difficult decisions, applying one simple formula to all moral choices. The cases of Jim and George are to utilitarians, like all cases, cost-benefit problems – with the sad fact that one of the costs is the agent's dispensing with a project. The only inputs to utilitarian deliberation are facts about the well-being that a course of action will produce, and, as the cases of Jim and George suggest, there will always be possible alternatives in which dispensing with a project maximises well-being. Therefore, the utilitarian must entertain alternatives in which they abandon any of their projects.[9, 10]

[8] For a similar (though more general and less negative) account of absurdity, see Nagel 1971.

[9] Frankfurt disagrees (1988, 180–81). He argues that a utilitarian may be so sure that a project of theirs will never be inimical to well-being that they do not entertain such outcomes, and that even if they did they may be sure that in such circumstances they would not be able to bring themselves to dispense with it. Though both phenomena are possible, I don't think they save utilitarianism. An agent who is as Frankfurt describes would probably not be complying with utilitarianism (because such surety is unlikely to be warranted, in either case). When evaluating utilitarianism here we should take the ideal utilitarian agent as our test subject.

[10] It may be thought that one could be committed to the utilitarian project of maximising well-being whilst regarding it impartially and as dispensable. Regarding it in these ways would, perhaps, simply be part of one's commitment, rather than in tension with it. (On the other hand, one might argue that this shows commitment to utilitarianism to be impossible.) I do not think that Williams, nor those persuaded by his Integrity Objection, would be satisfied with this response to it. They typically insist

My interpretation of Williams is similar to Edward Harcourt's. He writes:

> 'utilitarianism requires agents to be prepared at any time to set their (first-order) commitments aside. But it is partly constitutive of having a commitment that one is not prepared at all times to set it aside. This is not to say that commitments can never be sacrificed, but rather that if one's attitude to any of one's projects is that it could at any time be set aside, then setting it aside when it comes to it will hardly be a sacrifice. It follows that there is no room for both utilitarianism and commitments: as Williams says [(Williams 1981b, 51)], "one cannot both have the world containing these dispositions, and its actions regularly fulfilling the requirements of utilitarianism".' (Harcourt 1998, 190)

Harcourt has a second interpretation of the Integrity Objection, too, which I will deal with in a later subsection.[11]

It is crucial to note that Williams is not arguing that a committed agent *ought not* regard their projects impartially and as dispensable, in the utilitarian way. Rather, it is that they *cannot*. An 'ought not' would simply invite the utilitarian to dispute Williams's moral intuition. They could say, 'Well, I think that agents should regard their commitments in such a way.' Williams does not use straightforwardly moralistic language in this part of his critique. He asks, 'how can a man…?' and calls the utilitarian demand not wrong, but 'absurd'. The suggestion is that it is not immoral but rather impossible to have a commitment and regard it in a utilitarian manner. One cannot view one's commitments as 'happenings outside of one's moral self', impartially and as dispensable 'when the sums come in'. If I were to regard my relationship with someone as dispensable, it could not be a committed relationship. If I were to regard opera as valuable just insofar as it made people happy, rather than for its own sake, I would not be an opera-lover. If I were to regard the fact that I am a Zionist as one more fact to be considered when deciding on the optimific resolution to the situation in Palestine, just as much as the fact that someone else is an anti-

on the need for commitments other than to utilitarianism. Therefore, I will set aside this possible exception.

[11] Other interpretations of Williams that are similar to mine include Mulgan (2001, 15–16) and Tanyi (2015, 502–5). Both of these authors only mention the Integrity Objection briefly, as they focus on the demandingness objection from which they (correctly) distinguish it. Another interpretation in the neighbourhood is that of Joseph Raz (1986, chaps 11–13). On Raz's view, it is constitutive of commitment that one does not regard that to which one is committed as commensurable with other goods. Commensurability and dispensability may be related. However, I see no evidence that Williams's argument depends on incommensurability, or that he thinks that it is the commensurability, rather than the dispensability, of projects which condemns utilitarianism.

Zionist, rather than my ‘way of seeing the situation’ (Williams 1988, 190) bound up with who I am, I would not be a committed Zionist. The agent who looks at things in a utilitarian way cannot have such commitments. Thus utilitarians cannot argue that committed agents ought to regard their projects impartially and as dispensable, but could (and, I will claim in Chapter Four, should) argue that agents ought not have commitments in the first place.

The psychological problem, then, is this: if we accept utilitarianism then we regard our projects in a way (impartially and as dispensable) that is impossible for us insofar as we are committed to them. Schematically:

1. Having a project as a commitment is incompatible with regarding that project impartially or as dispensable.
2. If one accepts utilitarianism, one regards all projects impartially and as dispensable.
3. By 1 and 2, those who accept utilitarianism cannot have commitments.

There is a well-worn response to the psychological problem, which is to argue that accepting utilitarianism is not optimific (either for the reasons above or for others). In this case, the response goes, utilitarianism would direct agents not to accept utilitarianism: it would be ‘self-effacing’. Therefore, it is claimed that utilitarianism makes no recommendation that is incompatible with commitment, even if *accepting* utilitarianism is incompatible with having commitments. In Chapter Three I cast doubt on this response. I think that utilitarians would be better off confronting the psychological problem head-on, and arguing that utilitarianism does endorse a psychology incompatible with commitment, but that it is right to do so. An argument to that effect is given in Chapters Four and Five.

4. The normative problem

The psychological problem reconstructs the Integrity Objection in terms of thought: how utilitarianism requires us to regard our commitments, and how commitments, according to Williams, must be regarded. But commitments require action as well as thought (one cannot be an opera-lover without going or listening to the opera, or a friend to someone without ever lifting a finger to help them). What I will call ‘the normative problem’ is the allegation that utilitarianism is in tension with the reasons that committed people have for action.

There are of course many situations in which utilitarianism recommends actions which are incompatible with the actions required by some commitment. I might have a choice between going to see a friend in hospital, which would honour our friendship but not be much fun for either of us (he would not be good company for me, and my presence would make him feel guilty) and going to watch a cabaret which would be highly enjoyable. Utilitarianism recommends the latter (at least on some accounts of well-being). A committed vegetarian, presented with meat by an easily upset host, might be asked by utilitarianism to eat it – and to appear to enjoy it. Whilst, as I said above, it is not a blow to utilitarianism that it sometimes asks us to perform actions contrary to our desires, it is more worrying that it may ask us to act at odds with our commitments. The permeation of character by commitments means that actions stemming from commitments are part of how we maintain our distinctive selves – our integrity, in the sense of the term associated with wholeness and unity. However, once again we can point out that utilitarianism itself could be a commitment. There are people, for example in the 'effective altruist' movement, of whom it can be said that they have built their lives around the maximisation of well-being. For them, if utilitarianism asks us to act contrary to another commitment this is a conflict between commitments. Not only do such conflicts seem possible without rendering either commitment defective, that we may sometimes breach one commitment for the sake of other moral considerations is entailed by Williams's judgment that Jim should accept Pedro's offer, as we have already seen, as well as by Ashford's argument in her response to Williams. The normative problem is not simply that utilitarianism sometimes requires action at odds with other commitments.

Rather, the normative problem is that commitments alter 'the normative landscape' (Owens 2012) in a distinctive way that utilitarianism is unable to acknowledge. One way of thinking about such alteration is along the lines of obligation. That one is obliged to do something, many believe, does not merely mean one has a very strong reason to do it. Rather, obligations affect the force of other reasons: they might exclude certain kinds of reasons from one's consideration, for instance. As Scanlon puts it:

> 'The fact that it would be slightly inconvenient for me to keep a promise should be excluded as a reason for [not] doing so. Even if I am in great need of money to complete my life project, this gives me no reason to hasten the death of my rich uncle or even to hope that, flourishing and happy at the age of seventy-three, he will soon be felled by a heart attack. Against this, it might be claimed that I do have such reasons and that what happens in these cases is that I conclude that an action (breaking the promise or hiding my uncle's medicine) would be wrong and that the normative consequences of this conclusion then outweigh the very real reasons I have to do it. But this does not seem to me, intuitively, to be correct.' (1998, 156–57)

Utilitarianism, it might be thought, is committed to the position that Scanlon rejects. For the good my uncle's money could do, or the minor inconvenience of keeping a promise, may well affect the sum total of well-being and are thus weighed against the harms of death and of promise-breaking respectively in the utilitarian calculus. On Scanlon's view, the fact that we are obliged not to kill our uncles or break our promises makes such weighing inappropriate. (For slightly different views that accord with this general point, see Raz 1999, 1975; Owens 2012.) The Scanlonian claim is not that there could never be any reasons to breach an obligation. It is that some considerations do not count as reasons, in the light of obligation. Utilitarianism, it seems, does count such considerations as reasons, so long as they bear on the sum of well-being.

Something similar could be true of commitment. When some action is required by a commitment, this is not simply a strong reason for one to do it. It affects the normative landscape. Other reasons – reasons that utilitarianism takes to be relevant – become irrelevant. Thus, the actions of a committed person are robust: that is, they are performed across a range of circumstances (Pettit 2015). This is because they are not sensitive to small changes in the balance of reasons – commitment makes countervailing considerations that would ordinarily be reasons irrelevant. This accords with a second meaning of integrity, roughly synonymous with incorruptibility; someone is 'a person of integrity' insofar as they stick to their principles in a range of circumstances. A judge with integrity, for example, will deliver fair trials, however much money she is offered to do otherwise. That the money would be nice is not a reason that figures in her deliberation (nor even is the good it could do in famine relief). Similarly, someone committed to a romantic relationship is prepared to stick with their beloved 'for richer, for poorer; in sickness and in health' - that their beloved becomes sick or poor is not a reason for them to end the relationship.

This line of thought relates to what Williams called 'moral incapacity'. This is 'the kind of incapacity that is in question when we say of someone, usually in commendation of him, that he could not act or was not capable of acting in certain ways.' (1992, 59) George, as Williams puts it, 'cannot' take the job. The idea is not simply that he should not, nor that he will not in these circumstances – although both of these are true – but that it is not possible for him. The impossibility is not unbounded. If George signed the contract when a gun was put to his wife's head and a pen in his hand, one could not say that this undermines his claim to be committed to opposing CBW. Williams writes:

> 'It is plausible to say, with the pessimist, that if having a moral incapacity implies that there are no circumstances at all in which the agent would knowingly do the thing in question, then there are no moral incapacities. Ingenious coercion or brutal extremity can almost always produce such circumstance.' (1992, 69)

For Williams, a moral incapacity is at least 'proof against rewards' (1992, 69) – if a greater salary were offered, George would still refuse the job. Just as inconvenience and money were no reasons, for Scanlon, to breach obligations, rewards are no reason, for Williams, to breach commitments. The normative problem is that utilitarianism recognises considerations that, in the light of commitment, are irrelevant – and is thus incompatible with the robust action that commitment demands.

I believe that utilitarianism can, contrary to appearances, accommodate the normative landscaping of commitment (or at least, the most attractive aspects of it). Utilitarianism acknowledges that commitment can make some facts that would otherwise provide reasons irrelevant. This is the case when commitment makes responding to those facts impossible. Commitments cannot always be disposed of at will. It may be that George really cannot, not simply as a matter of moral incapacity, but physical incapacity too, bring himself to accept the CBW job simply for the promise of a higher salary. That is not to say that he could not physically accept the job, in situations of 'ingenious coercion or brutal extremity', and perhaps in situations like the one from Williams's original example, where it is unemployment and the prospect of his rival conducting more dangerous research that George is asked to regard as reasons to accept. Rather, it is that he could not accept in some range of circumstances short of that, where the putative reasons for doing so are things such as salary. In these cases it may well be impossible for George to accept, and this would be because he has a commitment to opposing CBW. As we saw above, utilitarianism does not ask us to do the impossible – and it need not acknowledge any seeming considerations in favour of the impossible as reasons.

In the case of such a commitment, George's refusal would also be robust: he would refuse in any circumstances except those of coercion and extremity. How about cases in which George *can* dispense with his commitment? Then, accepting the job is possible for him. However, it is only possible at the cost of his commitment. Fully spelt out, accepting the job involves George not only signing a contract, clocking in, performing experiments and collecting his salary, but also dropping his commitment.

This significantly affects the utilitarian calculation regarding the case. The utility costs of dropping a commitment are likely to be very high. Firstly, consider the disutility to the agent of breaching their commitments. As Railton puts it: 'Commitments… may be very closely linked to the self, and a hedonist who knows what he's about will not be one who turns on his self at the slightest provocation.' (1984, 142) One aspect of well-being, or precondition for it, is a secure sense of identity. The questioning of one's identity – being told that one is not a real philosopher, learning that one was adopted, being misgendered – is distressing. In abandoning a commitment, one calls one's own identity into question, which may be similarly painful. Secondly, consider the disutility to the world, given that the commitment is conducive to well-being. Commitments determine repeated actions across a long period of time. Some commitments (say, to a spouse) lead us to do things (say, to help her in her projects, please her, lessen her burdens, and so on) every day, for the foreseeable future. George's commitment to opposing chemical warfare might not be so frequently active in determining what he does. But it will regularly influence him – every election, every protest march, every conversation he has in which the subject is brought up – for as long as he has the commitment, which could be many decades. If his commitment is conducive to well-being, then it is highly likely that the actions it leads him to perform are themselves conducive to well-being. And if each of these actions is conducive to well-being, then the sum of well-being generated by the commitment that leads George to perform them over and over again is likely to be very large.[12] If utilitarianism were to recommend that he drop the commitment for the sake of performing just one action (taking the job), then that action must have an even larger positive impact on well-being, outweighing the sum of all the possible actions dependent on the commitment.

So utilitarianism will only recommend that George take the job when (a) he is able to give up his commitment, and (b) taking the job has very high benefits for overall well-being. This may mean that George can, consistent with utilitarianism, refuse the job in a wider range of possibilities than it seems: his action can still be fairly robust. (Recall that Williams's incapacities are not fully robust: they except 'brutal extremity'. Whether the committed utilitarian agent acts more or less robustly than the one with a Williamsian moral incapacity depends on what counts as 'brutal extremity', and (as it should) on the value of the commitment in question.)

Because utilitarianism can accommodate the ways that commitment changes the normative landscape: in making some reasons irrelevant, and in making some actions robust – it can also accommodate the notion of normative powers, that is, that agents can voluntarily change the reasons they have. For instance, when one makes a promise, one places oneself under an obligation, altering the normative landscape accordingly. Likewise, for those commitments we choose to take on, when we take them on we change the reasons we have according to utilitarianism. If George had no commitment to opposing chemical

[12] In the previous subsection I noted that it is possible to dispense with a commitment, in Williams's terms, whilst retaining some of the attitudes and patterns of action associated with it. Thus George could take the job and still oppose CBW. That this is conceptually possible is important; however, it may not be in actual fact likely. Once he takes a CBW job, George will probably be exposed to more arguments for CBW, be barred from engaging in anti-CBW activism, and above all come to have a vested interested in CBW. This will likely at least temper his opposition to it, and he could foresee this likelihood at the point of deciding whether to accept. As Upton Sinclair found: 'It is difficult to get a man to understand something, when his salary depends upon his not understanding it!' (1994, 109)

warfare he would be directed by utilitarianism to accept the job in a much wider range of circumstances, because the costs of dropping a commitment would not feature in the utilitarian calculation. And if we can take on commitments that become impossible for us to dispense with, this excludes reasons that would otherwise apply to do things that would involve dispensing with the commitment.

Of course, the reasons that utilitarianism takes commitment to make irrelevant, and the specification of the extreme circumstances in which utilitarianism recommends breaching commitments will differ from the reasons and specification given by other moral theories. But now we are just back to trading intuitions. The deeper normative problem, that utilitarianism could not acknowledge that commitment shapes the normative landscape, can be avoided.

It is worth mentioning briefly here the view that, in the light of some commitments, moral considerations of any kind should not figure as reasons. This is Williams's famous 'one thought too many' case (1981a). This is not part of the Integrity Objection – though it is related. Williams presents the case as a problem not just for utilitarianism but for moral theory in general. Section Four of Chapter Four responds to a similar concern drawing on the socialist focus on our connections with others. But for now let me mention two simpler points a utilitarian can make in this area. Firstly, that morality is not the whole of normativity – this is the lesson Susan Wolf (2012) takes from Williams, and one with which the utilitarian can whole-heartedly agree. It may be thought that utilitarianism is less open to such a pluralism than many other moral theories; after all, it is supposed to ground all reasons on one thing, overall well-being. However, utilitarians must qualify this as all *moral* reasons, and recognise other kinds of reasons. Utilitarians tend to assert that it is rational for individuals to maximise their own well-being – indeed, anti-utilitarians often criticise them for doing so[13] – and use this claim in arguments for utilitarianism, making morality a generalised version of prudential rationality (Harsanyi 1977). So they must recognise reasons that are not about maximising overall well-being, and are not moral. They may also endorse reasons that are neither moral nor prudential, such as reasons stemming from roles, identities and projects. The utilitarian need not deny that a footballer has a reason to shoot at goal, even if doing so would not maximise overall well-being (or even her own). They simply deny that this is a moral reason, or even a prudential one, but one grounded in the sport they are playing. Similarly, those involved in a particular project or relationship might

[13] Consider Marx's identification of the marketplace of classical economics with 'Bentham, because each looks only to his own advantage. The only force bringing them together, and putting them into relation with each other, is the selfishness, the gain and the private interest of each.' (1990, 280)

have non-moral, non-prudential reason to do certain things. From the perspective of that project or relationship it might be inappropriate for moral concerns to be considered.

Secondly, that there is good utilitarian reason for the agent to sometimes not consider morality when acting – doing so could impede other aspects of the act which are conducive to overall well-being, such as the joy of someone acting or being acted upon for no reason but love. Now we are back into the psychological realm: discussing not what reasons agents have, but how they ought to deliberate. The thought that utilitarian agents sometimes ought to ignore some reasons will reappear in Chapter Three.

5. The value-theoretic problem

I noted above that Harcourt takes Williams to be making the objection I have labelled 'the psychological problem'. He goes on to make a further argument against utilitarianism, which is inspired by, but extends beyond, Williams's. This objection alleges that utilitarianism is unable, by its own lights, to provide verdicts about which actions are right and which are wrong. I will call this 'the value-theoretic problem'. As Harcourt introduces it:

> 'If... there were some evaluatively relevant feature of alternatives which utilitarian calculation is necessarily unable to recognise, it would follow... that wherever such a feature was instantiated, the verdict obtained by applying the utilitarian test of rightness was inadmissible by utilitarian lights; that is, applying the utilitarian test of rightness was necessarily unable to tell us which of the alternatives was the utilitarianly right one. It would then follow that the utilitarian test of rightness was incoherent.' (1998, 192)

Harcourt then aims to show that there is some evaluatively relevant feature of alternatives which utilitarian calculation is necessarily unable to recognise. There are, he says, things whose value to people cannot be represented by utilitarianism's tools. Harcourt focuses on a species of utilitarianism, which defines well-being as the satisfaction of preferences. As a test of rightness, preference-satisfaction utilitarianism takes all of the preference-orderings of affected people over all available outcomes, and commends as right the action that satisfies the most preferences (weighted by the intensity with which these preferences are had). Harcourt points out that preferences cannot represent certain aspects of the way in which people value things – for instance, Williams's moral incapacities. In such cases, it is not that one strongly prefers one course of action to another, but rather that the fact that some course of action is available rules all others out. Williams's cases of George and Jim could be cases of this kind, with the fact that a commitment is at stake excluding consideration of their accepting their offers. In some such cases, one might not form any attitude towards some course of ruled-out action (and therefore not give it a place in one's preference-ordering).

Although later I will argue (as Harcourt himself suspects (1998, 198)) that Harcourt's argument only has bite against preference-satisfaction utilitarianism, I think such narrowing of scope is not forced upon him at this stage. It seems to me that any ordering – in terms of preference, or hedons, or whatever – cannot represent features of practical deliberation such as moral incapacities. Orderings compare options, and these sorts of features involve constraints on what can be compared.

But does this mean that utilitarianism is unable to take some aspects of *well-being* into account, and therefore cannot assess what is 'utilitarianly right'? To show this, Harcourt needs the additional premise that the contribution of things that are valued in such non-ordered ways contributes to well-being, and does so in a way that itself cannot be represented by an ordering. Harcourt is close to suggesting this when he posits 'an internal relation between the value which things have to an agent and the shape of the agent's practical deliberation' (1998, 192). Therefore, if a person's valuation of a choice cannot be represented by an ordering, then neither can the potential effects of their well-being, and thus utilitarianism, with its reduction of value to orderings, cannot acknowledge all aspects of well-being. But why should this premise hold?

This premise can be split into two parts. The first is the claim that those things that people value in non-orderable ways are conducive to their well-being. This may come from a more general thesis that whatever people value is conducive to their well-being. The second is the claim that the non-orderable ways in which people value those things are reflected in their contribution to well-being being non-orderable. This may come from a more general claim that the way in which people value something determines how it affects their well-being.

Utilitarians may reject either of these claims. People value things other than well-being, and value things that do not contribute to their well-being. Harcourt understands that not all of the valuations an agent makes are relevant to their well-being. But he thinks that people value commitments in non-orderable ways, and that they do contribute to their well-being. He gives as examples political commitments, such as to Zionism, and 'an athlete's commitment to excel at a sport' (1998, 197).

Now, it seems true that people have such commitments, and value them, or things connected with them, in non-orderable ways. But it is less obvious that these valuations have anything to do with well-being. It seems perfectly consistent to be committed to something, to view it as indispensable and build one's life around it, whilst still thinking that, as far as well-being goes, it is simply one way of living, better for you than some but maybe worse than others. You might think that you have to do such-and-such to be a good athlete, or a good Zionist, and care deeply about being these things, without thinking that doing such-and-such will improve your well-being. In fact, most committed people think about their well-being very little when considering their commitments (Scanlon 1998, 126–33). Thus, the valuations to which Harcourt draws attention are a feature of practical deliberation, but not of practical deliberation about well-being.[14]

The second claim was that the non-orderable ways in which people value things are reflected in their contribution to well-being being non-orderable. Even if we grant that whatever people value, including whatever they value in non-orderable ways, contributes to their well-being, we may still deny that their contribution to well-being follows the non-orderable shape of their own valuations. It may be that some commitment is very important to a person, and this is reflected in certain practical modalities in their valuations of it, but its contribution to their well-being can also be represented by a cardinal or ordinal well-being score. An athlete might have a commitment to sporting excellence that leads them to think some things must be done, that other options are excluded or incommensurate, but it may also be the case that their well-being in a sporting career is directly comparable in a straightforwardly orderable way (by a morally motivated outsider, rather than by them), with their well-being in another lifestyle.

So Harcourt's argument is resistible, if we reject that idea that what people value is what matters to their well-being and/or the idea that their well-being is affected by something according to the shape their valuation of that thing takes.

Preference-satisfaction utilitarians might not be able to coherently reject these ideas. They are committed to the views that whatever someone prefers is what matters to their well-being, and that the intensity of their preferences for something determines how much their well-being is affected by it. These views are more specific versions of those that I have just said utilitarians should reject to avoid Harcourt's objection. However, preference-satisfaction utilitarianism faces other serious problems, such as those that arise from adaptive preferences (Elster 1983; Sen 1995, 259–84). I do not defend preference-satisfaction utilitarianism in this thesis, but rather forms of utilitarianism that escape the value-theoretic problem.

[14] This is not to say that having such commitments may cause one to have greater well-being than not doing so. But this doesn't make the valuations involved in them valuations *of* well-being.

6. The coherence problem

The psychological problem was this:

1. Having a project as a commitment is incompatible with regarding that project impartially or as dispensable.
2. If one accepts utilitarianism, one regards all projects impartially and as dispensable.
3. By 1 and 2, those who accept utilitarianism cannot have commitments.

One response to this problem is to argue that utilitarianism is self-effacing; it does not require acceptance of itself. I will argue against this response in the next chapter. I believe:

4. Utilitarianism requires individuals to accept utilitarianism.

At points, Williams suggests a further problem for utilitarianism. This is that commitments are so important for well-being that:

5. Utilitarianism requires agents to have commitments.

If this were the case, then, given 3 and 4, utilitarianism would be incoherent. It would require agents to both have and not have commitments. Call this the 'coherence problem'.

How could 5 be supported? In one section of his *Critique*, Williams suggests that utilitarianism would be nonsensical if people did not have projects, of which commitments are a subset. Rejecting hedonism, he takes it 'that in talking of happiness or utility one is talking about people's desires and preferences and their getting what they want or prefer, rather than about some sensation of pleasure or happiness.' (1973, 80). For utilitarianism to be meaningful, therefore, there must be preferences to satisfy. Williams assumes that this requires people to have projects from which those preferences arise. And this must include some projects that are not the utilitarian project itself, since conceived as a project of maximising preference-satisfaction, it is 'vacuous' unless there are 'other more basic or lower-order projects.' (1973, 110) Commitments are one class of those projects.

This argument fails to establish that utilitarianism requires us to have commitments. For one thing, the 'requirement' in question is more like a presupposition than a moral prescription. If utilitarianism would be vacuous without commitments, then the defender of utilitarianism might be glad that some people have them, for this makes her theory meaningful, but it does not follow that utilitarianism says that agents *should* adopt commitments. For another, even if we grant that utilitarianism requires us to have preferences, it does not follow that utilitarianism requires us to have commitments. Imagine (with Parfit 2016, 118) a world whose inhabitants had only the drabbest of things in their lives – muzak and potatoes. Suppose that they only have two preferences: that there should be muzak rather than silence, and potatoes rather than gruel. It is not that they have any deep affection for muzak or potatoes: they desire and enjoy them no more than we do. They simply lack the means to form preferences for anything else. They do not have commitments, in Williams's sense. Yet it is obvious what they prefer and therefore what utilitarianism recommends for this world: more muzak and more potatoes. Lastly, Williams's argument here, like Harcourt's, would only apply to preference-satisfaction utilitarianism (or closely related alternatives), which is problematic on other grounds.

Williams makes a better argument for 5. He proposes, as an empirical hypothesis, 'that many of those with commitments, who have really identified themselves with objects outside themselves, who are thoroughly involved with other persons, or institutions, or activities or causes, are actually happier than those whose projects and wants are not like that.' (1973, 113–14) Whether this hypothesis is true or not is an interesting and important question, and the answer is not obvious, although the prevailing view is that it is (see Calhoun 2009 for dissent). Williams's hypothesis does not depend on any particular account of well-being, but some accounts of well-being – for instance, Joseph Raz's – make it closer to tautological than empirical. Raz writes that that our typical notion of a good life

> 'is of a life well spent, of a life of achievement, of handicaps overcome, talents wisely used, of good judgment in the conduct of one's affairs, of warm and trusting relations with family and friends, stormy and enthusiastic involvement with other people, many hours spent having fun in good company, and so on.' (1986, 306)

Trusting familial relationships and friendships paradigmatically involve commitments, and achieving things, overcoming adversity and wisely using one's talents may also do so. A life without these things might include good company and sound judgment (as well as sensory pleasures), but we might resist calling a truly happy life – or at least think that it would have been better with respect to well-being had it involved commitments. This tells in favour of Williams's hypothesis. On the other hand, some of the worst lives tend to involve commitments as well: loving marriages and friendships break up, dreams are unfulfilled, martyrs are made in defence of lost causes. It is likely that those with commitments that are fulfilled have happier lives than those without commitments, but this group is sadly only a subset of those with commitments. Nevertheless, I will grant that it seems likely there are some kinds of commitment such that for most of us, if we adopted them, our lives would be happier.

However, this still does not entail 5. Utilitarianism does not require agents to make their lives go as well as possible. It requires agents to maximise overall well-being – that is, to make the lives of everyone, taken as an equally weighted sum, go as well as possible. And it could be that whilst it would reduce *my* well-being for me to reject commitment, doing so would free me to do things for others that would maximise well-being *overall*.

At this point one might object that if we all rejected commitment, there would be no commitment in the world, and this would reduce overall well-being (to zero, perhaps, if something like Raz's account of well-being is true). But as far as utilitarianism is concerned, what matters is not what would happen if we all did something, but the marginal effect of my doing so.[15] And these things can come apart. Consider this case from Feldman (who used it to make a different point):

> 'Suppose a group of adults has taken a group of children out to do some ice skating. The adults have assured the children and their parents that, in case of accident, they will do everything in their power to protect the children. Each adult in the party is a good skater and swimmer. Suppose, finally, that, while they are out skating, it just so happens that all the adults are spread out around the edge of the pond. A lone child is skating in the middle, equidistant from the adults. Suddenly, the ice breaks, and the child falls through. There is no time for consultation or deliberation. Someone must quickly save the child. However, since the ice is very thin, it would be disastrous for more than one of the adults to venture near the place where the child broke through. For if two or more were to go out, they would all fall in and all would be in profound trouble.' (Feldman 1980, 171)

Now, imagine that you are one of the adults and you know that no other adult will go to save the child. Utilitarianism then directs you to save the child: this would result in the best outcome your actions could produce. This is true even though if every adult did the same, a disaster would occur.

Rejecting commitment could be like going to save the child. That is, it could be that both (a) if we all rejected commitment, this would be suboptimal, and (b) that the expected effects of my rejecting commitment, given the likely behaviour of others, would be optimal. In such cases, utilitarianism would direct me to reject commitment. (Cases with a similar structure reappear in the next chapter.) So even if commitment is necessary for well-being, this does not mean that utilitarianism requires commitment from agents.[16]

[15] This is what gives rise to the pre-emption problem which we will meet in the next section, and in Chapter Six.

[16] Wolf expresses a similar thought: 'if the utilitarian wants to influence more people to achieve more good, then he would do better to encourage them to pursue happiness-producing goals that are more

7. The pre-emption problem

To reiterate, Williams's Integrity Objection is not an objection to utilitarianism's recommendations for how Jim and George should act. Rather, Williams primarily uses the cases to (a) emphasise the importance of commitment, and (b) show utilitarianism's incompatibility with it. However, utilitarianism's recommendation in cases of this kind do, I think, pose a serious problem for the theory. Though this is not the focus of Williams's objection, it is worth mentioning here – and will be discussed in detail in Chapter Six.

Return to Williams's description of the structure of his cases:

> 'The situations have in common that if the agent does not do a certain disagreeable thing, someone else will, and in Jim's situation at least the result, the state of affairs after the other man has acted, if he does, will be worse than after Jim has acted, if Jim does. The same, on a smaller scale, is true of George's case.' (1973, 108)

Since not doing so leads to a worse outcome, utilitarianism is bound to recommend that Jim and George (and others in similar situations) do the disagreeable thing. We may add to Williams's description that in these cases if nobody did the disagreeable thing – that is, if Pedro refrained from shooting anyone, and if George's rival (and indeed all chemists) refused to take the job – the outcome would be better than if Jim or George did so, notwithstanding the fact that their doing so would be better than Pedro or George's rival acting as expected. Further, we can generalise such cases so they do not depend on being certain about what others will do, to get the following description of cases akin to Jim's and George's:

> There is some action available to me that would be the proximate cause of a bad outcome. However, were I not to take such action, some other individual would probably take a similar action which would be the proximate cause of as bad or worse outcomes. The likelihood with which they would take such an action, and the worseness of the outcomes they would produce are such that the expected value of my taking the action is greater than or equal to the expected value of my not doing so. This is so even though if nobody took the action, a better outcome would result.

attractive and more within a normal person's reach. These considerations still leave open, however, the question of what kind of an ideal the committed utilitarian should privately aspire to himself. Utilitarianism requires him to want to achieve the greatest general happiness, and this would seem to commit him to the ideal of the moral saint.' (1982, 427–28)

Call cases matching this description ‘pre-emption cases’.

Pre-emption cases pose a serious problem for utilitarianism. It is not just that utilitarianism, as Williams notes, will recommend that agents do disagreeable things in such cases. It is also that there is a tension between two intuitive motivations for utilitarianism. One guiding light of utilitarianism is that moral requirements are determined by outcomes, ranked in terms of overall well-being. But in pre-emption cases, utilitarianism recommends actions (George taking the job, Jim shooting a prisoner) which are incompatible with the best possible outcomes (CBW is eliminated, nobody is shot). This is a result of utilitarianism’s other guiding light, that the difference an action makes to outcomes is crucially important to its moral status. ‘What difference would it make?’ is often a very good question to ask in moral deliberation. And indeed, this is part of utilitarianism’s appeal: many non-utilitarians also consider the degree to which an action would make a difference to an outcome significant to whether, and how urgently, it ought to be done.

Pre-emption cases are widespread. As I explain further in Chapter Six, they are common in the workplace and in markets, and are connected with very serious social problems such as weapons of mass destruction, climate change, and the exploitation of the global working-class. It is also worth noting that – unlike in Williams’s cases as he describes them – pre-emption cases can occur even when every individual involved is trying to bring about the best outcome they can.

I will propose a utilitarian solution to these cases in Chapter Six. This solution emphasises the importance of the connections between individuals and others in groups. Before we get there, we will see in the next three chapters that a similar emphasis is part of the solution to the psychological problem – that is, to the Integrity Objection proper.

8. Conclusion

In this chapter I have explored Williams’s Integrity Objection, and the diverse range of problems it is alleged to pose for utilitarianism. I rejected two interpretations of the objection: that it offers counterexamples in which utilitarianism’s recommendations are at odds with moral intuition, and that it aims to show that utilitarianism is too demanding by making agents abandon commitments. Williams’s objection has to do with the way utilitarianism treats the relationship between agents and their commitments. The most significant problem that the objection poses is what I called ‘the psychological problem’, that agents who accept utilitarianism cannot have the attitudes towards their projects that they must if they are to be properly committed to them. There are other problems in the neighbourhood: that utilitarianism does not acknowledge that commitments can change the reasons agents have; that it cannot accommodate their special value; that it incoherently asks agents to both hold and not hold commitments. I briefly showed how utilitarians might respond to these three problems. I also introduced the pre-emption problem, which springs naturally from Williams's cases, though it is not central to his discussion of integrity. This problem will be discussed at greater length in Chapter Six. In the next three chapters I offer a response to the psychological problem.

Chapter Three: Is utilitarianism self-effacing?

0. Abstract

The claim that utilitarianism is self-effacing offers one response to Williams's Integrity Objection. If utilitarianism is self-effacing, it does not recommend that agents accept utilitarianism, and thus makes no recommendation that is incompatible with commitment – even if Williams is right that accepting utilitarianism is incompatible with commitment. However, Williams believed that being self-effacing is a theoretical vice for moral theories.

In this chapter I describe and rebut three arguments for the claim that utilitarianism is self-effacing for individuals. I then note that they might suggest that utilitarianism is self-effacing for collectives. If utilitarianism is not self-effacing for individuals the Integrity Objection cannot be easily avoided: utilitarianism does direct agents to accept itself, and thus makes commitment impossible for them. However, self-effacingness for collectives may mitigate some of the problems associated with accepting utilitarianism for individuals.

1. Introduction

A normative theory T is self-effacing for an agent X iff it implies that X ought not accept T. To accept a theory, broadly, is to be disposed to use it in certain ways in practical reasoning, and to take certain practical attitudes towards it (for accounts of acceptance along these lines see Harman 1986; Van Fraassen 1980; Ross 2006).[17]

Utilitarianism is self-effacing for an agent X iff X's not accepting utilitarianism would be more conducive to well-being.[18] Williams's Integrity Objection, as we have seen, holds that accepting utilitarianism is incompatible with commitment to one's projects (Williams 1973). If utilitarianism is self-effacing, though, it implies that individuals ought not accept utilitarianism, and so cannot be blamed for any problems that arise when they do so.

[17] On this notion of acceptance, one can accept a theory by believing it, but acceptance need not entail belief. In the next section I temporarily grant that this entailment might hold in the case of moral theories. My arguments do not, I think, hang on whether acceptance implies belief.

[18] Parfit holds that a theory is self-effacing iff it tells agents not to believe itself, 'but some other theory' (Parfit 1984, 40). Thus, utilitarianism would be self-effacing for X iff X's accepting some non-utilitarian theory instead of utilitarianism would result in more well-being. My definition is more inclusive. It may be that more well-being is produced if some agent accepts no theory at all: according to my definition, but not Parfit's, this would make utilitarianism self-effacing. Note that Parfit also puts self-effacingness in terms of belief rather than acceptance, though this is not a distinction he defines or seems to find important.

Commitment might be off the table for those misguided souls who accept utilitarianism, but they are not problems for the theory itself. Both Parfit (1984, chap. 1) and Railton (1984) make arguments of this form.

Williams believed that being self-effacing is a theoretical vice for moral theories. He objects that a self-effacing utilitarianism would 'usher itself from the scene' (1973, 134) – failing to serve the action-guiding role to which moral theories aspire. So utilitarians seem to face a dilemma: if utilitarianism is self-effacing, it may avoid the Integrity Objection but face this other objection; if it is not self-effacing, it avoids this other objection but not the Integrity Objection.

In this chapter I describe and rebut three arguments for the claim that utilitarianism is self-effacing for individuals. The first is 'the argument from other motives', which holds that since acting from non-utilitarian motives will sometimes be more conducive to well-being, utilitarianism tells agents not to accept utilitarianism. The second is 'the argument from collective self-defeatingness', which holds that since a world in which we all accepted utilitarianism would be suboptimal, utilitarianism tells each of us not to accept utilitarianism. I argue that both of these inferences are invalid. The third is an argument from the end of Derek Parfit's last volume of *On What Matters*. Parfit's argument suffers from similar flaws to the argument from collective self-defeatingness.

I then note that the argument from collective self-defeatingness suggests that utilitarianism may be self-effacing for collectives, and explore the meaning and implications of this possibility. I then claim that if utilitarianism is self-effacing for groups but not individuals this need not make utilitarianism objectionably 'esoteric'.

The upshot is that utilitarianism is unlikely to be self-effacing for individuals, but is more likely to be self-effacing for groups. The former means that the Integrity Objection cannot be easily avoided: utilitarianism does direct agents to accept itself, and thus makes commitment impossible for them. However, self-effacingness on a group level may mitigate some of the problems associated with accepting utilitarianism. Foreshadowing the chapters to come, this chapter shows how reflection on the distinctions and relations between individuals, collectives and groups can help us to make progress in evaluating utilitarianism.

A preliminary note: all plausible moral theories are self-effacing for some agents in some circumstances. Parfit demonstrates this with examples involving Satan (Parfit 1984, 43–45: see also Eggleston 2013, and Lazari-Radek and Singer 2010). For any theory, we can imagine that the Devil credibly threatens to unleash great amounts of suffering on the world if some agents accept that theory. If that theory is plausible, it will tell agents to do whatever it takes to avoid this outcome (it is not only utilitarians who care about avoiding disastrous outcomes). Therefore, it will tell the agents concerned not to accept itself: it will be self-effacing for them. Objections to self-effacingness, then, cannot rule out moral theories just because they would be self-effacing for some agents in some circumstances. This would leave us with only implausible theories. But perhaps utilitarianism would be objectionable if it were self-effacing for ordinary agents in the actual world, or ones close to it. In this chapter, that is what I will take the claim that utilitarianism is self-effacing to be.

2. The argument from other motives

It is a commonplace that individuals who constantly, consciously attempt to maximise well-being are unlikely to be successful. We will sometimes produce better outcomes by acting from other motives. In these cases, utilitarianism will recommend that we act on those other motives. As Sidgwick said:

> 'if experience shews that the general happiness will be more satisfactorily attained if men frequently act from other motives than pure universal philanthropy, it is obvious that these other motives are to be preferred on Utilitarian principles.' (Sidgwick 1884, 409)

To act from other motives, one will probably have to accept some moral principles other than utilitarianism, including principles that issue, in some circumstances, different recommendations. Utilitarianism might tell us to accept such principles. The 'argument from other motives' holds that this makes utilitarianism self-effacing: it will not only tell us to accept those principles, but to not accept utilitarianism.

Williams seems to make something like this argument, writing that his utilitarian interlocutor JJC Smart is right to want to recommend non-utilitarian motives, but that 'once that has started, there seems nothing to stop, and a lot to encourage, a movement by which [utilitarianism] retires to the totally transcendental standpoint from which all it demands is that the world should be ordered for the best, and that those dispositions and habits of thought should exist in the world which are for the best, leaving it entirely open whether those arc themselves of a distinctively utilitarian kind or not'. (1973, 134–35) If his arguments for the importance of non-utilitarian motives are correct, he says, 'it is reasonable to suppose that maximal total utility actually requires that few, if any, accept utilitarianism.' (1973, 135). I think, however, that utilitarianism is likely to recommend acting from other motives much of the time without moving itself 'to the totally transcendent standpoint': it will still recommend accepting itself.

Before I make that argument, let us look at when utilitarianism might require action from non-utilitarian motives. One situation is decision-making under time constraints. We could not get by if, as Hare puts it, we had to perform a cost-benefit analysis on the spot each time we wanted to make a decision (1981, 36). We would miss opportunities to produce well-being simply because we were trying to work out how best to produce it. To effectively fulfil the aims of utilitarianism, then, we need to draw on other, non-consequentialist principles that are simpler to apply: for instance, general rules such as 'drive within the speed limit'. The generality of such rules is what makes them more useful than utilitarianism in time-constrained circumstances; it is also what brings them into conflict with utilitarianism in exceptional ones. For there will be occasions in which it is optimific to break the speed limit. But these occasions being rare, and making constant predictions about the consequences of your speed as you drive being inadvisable, it may still maximise well-being for you to accept the general rule. Therefore, utilitarianism might recommend, as Railton says, that one 'should have (should develop, encourage, and so on) a character such that [one] sometimes knowingly and deliberately acts contrary to [one's] objective consequentialist duty' (1984, 159).

There are also cases in which the fact that one does something from a particular non-utilitarian motive is itself conducive to well-being. That we do some things out of love for a partner, or devotion to a cause, or as an expression of our identity might be a component of our well-being – or it might be necessary for the action to have the kind of meaning to ourselves and others that makes us all better off. Another example of such motives is those involved in competitive games, where the purpose, and thus a good deal of the value of those games would be lost if one played them with the aim of maximising overall well-being rather than of winning. Acting from these motives might be accompanied by the endorsement of non-utilitarian principles such as the general rules mentioned in the last paragraph, but it might be more instinctive still, involving no reference to principles. Either way, these are actions that are not motivated by reflection on the utilitarian calculus – and may thereby be more conducive to well-being.

I am happy to grant that utilitarianism recommends the adoption of other motives and principles.[19] But it does not imply that it is self-effacing, as I have defined it. We could accept these other principles, and thus act from these other motives, *and also* accept utilitarianism.

[19] For a more detailed argument for this position, see Crisp (1992). Crisp seems to share my position that it would maximise overall well-being if agents sometimes acted from utilitarian motives, 'on certain special occasions' (1992, 154, 156), though he also writes that: 'Utilitarian moral agency requires the sort of insight into what makes life worth living that can be gained only by forswearing utilitarianism as a decision procedure' (1992, 159). This forswearing, in my view, can only be partial.

Only if it prevents us doing the latter does the need to act from other motives in order to maximise overall well-being show that utilitarianism is self-effacing.

This would be so if it were psychologically impossible to accept two inconsistent moral principles. Williams seemed to think that it is, at least in some relevant cases. He writes of Hare's 'two-level utilitarianism', which asks agents to endorse both utilitarianism and non-utilitarian principles:

> 'you cannot combine seeing the situation in that way, from the point of view of those dispositions, with seeing it from the archangel's way, in which all that is important is maximum preference satisfaction [read: overall well-being], and the dispositions themselves are merely a means towards that.' (Williams 1988, 190)

But why can we not make such a combination? There are cases in which we do combine different ways of seeing the world, with one way being a means towards the goal of the other. Railton gives the case of a tennis player (Railton 1984, 144–45), who has the aim of winning as many matches as he can, but is underperforming. He is advised he will play better, and hence win more frequently, if he forgets about this aim whilst he played. He should play for the love of the sport, rather than to win. In taking the advice, he accepts, on the one hand, that winning is of paramount importance. On the other hand, he pushes this attitude from his mind whilst he plays, and accepts that the love of the sport is of paramount importance. These principles are inconsistent. But their combination seems possible, and similar cases are common.[20]

As the tennis player does winning, the moral agent could take maximising overall well-being to be of paramount importance. But she could also push this attitude away in certain cases, and act from non-utilitarian motives (acting, perhaps, for the love of her wife, rather than the love of tennis). I do not take this to show that we must take such a divided approach, accepting incompatible principles. The point is that we do take such an approach in other parts of our thinking, seemingly unproblematically, and therefore the burden of proof is on Williams's claim that it is impossible with respect to moral principles.

Perhaps it is not that it is psychologically impossible to accept both utilitarianism and principles that sometimes conflict with it, but that to do so would violate some epistemic norm. It may be (at least on some views) that morality does require the violation of some

[20] Nagel gives further cases of sophisticated attitudes: 'The only way to run downstairs is not to try, you cannot make her love you by doing what you think will make her love you, you will not impress the interviewer unless you stop trying to impress him.' (Nagel 1970, 132)

epistemic norms. But this would be worrying. The epistemic norm that seems threatened here is the norm of consistency: one ought not believe both P and not-P.

Now, that we can *accept* both P and not-P seems to have been proved by cases like Railton's tennis player. Unless we think this tennis player is epistemically irrational, there seems no norm that one ought not accept inconsistent principles. But it is more plausible that one ought not *believe* inconsistent principles. Such a norm seems to be the basis of the force of many philosophical and everyday arguments. And accepting a moral principle might entail believing it, in a way that accepting a principle to help you win tennis matches does not. I do not endorse the claim here that accepting a moral principle entails believing it; but I'm happy to grant it for the sake of argument. It is not implausible. It is less obvious that Railton's tennis player must believe that the love of the sport is sufficient motivation for his play than that he must simply be motivated by it; whilst Railton's devoted husband might believe in the rightness of non-optimific actions performed out of love for his wife, a belief that is inconsistent with a belief in a utilitarian criterion of rightness.

Could utilitarianism that recommend agents hold inconsistent beliefs? We all have some inconsistent beliefs, so it is not psychologically impossible. Nor is it true that it is always irrational to hold inconsistent beliefs. Imagine that one has good reason to believe each of three propositions. They are, however, jointly inconsistent. One can infer that at least one is false, but has no evidence as to which one. There does not seem to be sufficient reason to reject any one of the three propositions; since one has good reason to believe each of them, it seems rationally permissible to go on believing all of them, despite the inconsistency.[21] Now, the case of utilitarianism and non-utilitarian principles could be similar. If utilitarianism is true, as I have already noted, this would provide good reason to accept certain non-utilitarian principles: doing so would produce more well-being. There may also be (again, if utilitarianism is true) good reasons to accept utilitarianism – I will argue for this below. Both being the case, I should accept both. This might lead me to inconsistent beliefs. But this is not sufficient reason to reject either. If it were, which should I reject? Utilitarianism, which provides the reasons for holding onto the other principles – but then why should I keep the latter? Or the other principles, which utilitarianism implies that I should have – but then wouldn't it be more irrational to go on believing utilitarianism whilst not having them?

If, as I have suggested, utilitarian and non-utilitarian motives can be accepted in combination, the mere fact that we need non-utilitarian motives does not imply that utilitarianism is self-effacing for individuals. In fact, there is good reason to think that the need to act from other motives demonstrates the need for utility-maximising agents to accept

[21] Thanks to Mark Kalderon for this example.

utilitarianism. Call the set of principles and motives that an agent accepts 'M'. According to utilitarianism, each agent ought to cause themselves to accept the M such that, of all Ms they could possibly cause themselves to accept, their acceptance of that M would be optimific. How is one to hit upon the optimific M? The obvious way is to try and work out which M would be optimific for them to accept, and accept it on those grounds. But only someone who accepts utilitarianism would be motivated to do this: if one didn't think that one ought to do whatever is optimific, why would one select one's principles based on the amount of well-being that accepting them would produce?

So even though utilitarianism will recommend accepting non-utilitarian principles and motives, in order to establish which of these one ought to accept it may be necessary for agents to accept utilitarianism alongside them. Thus, utilitarianism would not be self-effacing for these agents. However, we might think that it will fast become so. Once an agent has used utilitarianism to motivate their acceptance of the optimific M, will utilitarianism then suggest that they ought to jettison utilitarianism and accept only M?[22] If so, utilitarianism is self-effacing for agents who accept the optimific M – that includes, given the process above, all agents who have in the past used utilitarianism correctly to find M. I don't think utilitarianism can be jettisoned, however. For M will need updating over time. The M such that your acceptance of it produced the best outcomes five years ago probably will not be your optimific M today. You might become more able to deliberate accurately with more complex principles (consider how the simple rules we were encouraged to adopt as children gradually acquire caveats). You might be able to do more things – or fewer things. External circumstances might change: perhaps your wife leaves you, so you do not do so much good by having the character of a devoted husband. If you do not change your M in the light of such changes, what you do will diverge from what produces the best outcomes, that is, what utilitarianism says you ought to do. So utilitarianism will say you ought to change your M – but again, how? The obvious way is by employing utilitarianism just as before, and retaining a standing disposition to employ it in decisions about M.

Now, it is conceivable that employing some other theory in this role would also cause us to firstly accept and then to maintain an optimific M. But given the considerations about updating M over time it seems unlikely. No theory substantially different from utilitarianism will respond to every possible change by directing us to make precisely the same changes in M that utilitarianism will. And since getting M right, given that M largely determines one's actions and therefore on the well-being one produces, is of great importance to

[22] As Wittgenstein says of the propositions in his *Tractatus*, the reader 'must, so to speak, throw away the ladder after he has climbed up it' (6.54).

utilitarianism, utilitarianism will suggest we use the tool that gets M as precisely right as possible – that is likely to be utilitarianism itself.

A further reason to think that utilitarianism is not self-effacing is that successfully guiding one's actions by the optimific M, though it may be easier than successfully and ubiquitously employing utilitarianism itself, is hard. Any plausible candidate for my optimific M will demand a significant amount of behaviour that is not in my immediate interest, and so it will often be tempting to defect. To keep us on the right path (or at least as close to it as possible), we will need to be able to justify this M to ourselves. Why, we find ourselves asking, should I follow this set of rules and dispositions? The obvious answer is that this is the optimific M. One will only be moved by this answer if one thinks that one ought to do what is optimific; that is, if one accepts utilitarianism. Thus, by accepting utilitarianism you'll be more likely to successfully abide by this M, producing better results.

Now, other principles could play this role. For instance, one might think that one should comply with the optimific M not because it is optimific, but because it is commanded by a god whose perfection one accepts. Or one could simply wrap one's identity up in being the sort of person who employs M. Both justifications often prove successful at getting individuals to stick to some morality. However, there is reason to think that better outcomes will result from a utilitarian justification of the optimific M. Firstly, it is necessarily true that the employing the optimific M is optimific, whilst it is at best contingently true that doing so is commanded by your god or central to your identity - so utilitarianism will provide a more robustly effective motivation. Secondly, consider again the need to revise one's M over time, to prevent divergence from optimific-ness as circumstances change. If your justification for employing some M is that it is optimific, you will be disposed to make the right kinds of revision when it no longer is. This would not be true if your justification of employing that M is divine command (why would a perfect god change their mind?) or identification (it is very difficult, and costly, to revise principles one has built one's life around, as contributors to this debate frequently note).

Where does this leave us with respect to the argument from other motives – that one will do better at bringing about good outcomes if one acts from motives other than utilitarianism, and therefore that utilitarianism is self-effacing? My conclusion is that although each individual will likely do better by accepting some other motives, and therefore some non-utilitarian principles, it is unlikely that they will do better by not accepting utilitarianism at all. Without accepting utilitarianism it would be sheer luck if one accepted, maintained, revised and justified the other principles that it would be optimific for one to accept. Utilitarianism has an important role, then, at a higher deliberative level than many other principles it might tell us to accept, as Hare (1981) suggests. This is not to say that utilitarianism will never have a role in determining individual actions as well. Utilitarianism is therefore unlikely to be self-effacing for individual agents.

Holly M. Smith (1989) suggests that utilitarianism will not be able to play this role. As Smith puts the objection, it only applies to agents whose need for other motives stems from their inability to accurately make the calculations that utilitarianism requires. Wouldn't it be the case that 'the same empirical misinformation that plagues their application of [utilitarianism], now prevents them from seeing that [the optimific M] is the correct code by which to guide their actions' (H. M. Smith 1989, 131)? As the examples of relationships and games show, there are reasons other than 'empirical misinformation' that agents might need other motives to act optimifically. Perhaps Smith's worry generalises: why should agents who need non-utilitarian motives (for any reason) in order to make optimific decisions be best able to make decisions about which non-utilitarian motives to acquire and maintain by applying utilitarianism?

There are two ways to answer Smith's worry, and both are plausible. The first is that agents can be more successful with their employment of utilitarianism in some deliberative situations than others, so their application need not be equally 'plagued' in each. In time-constrained, stressful or emotional situations, it is very difficult to accurately predict consequences. In such situations, we would be more likely to act optimifically by applying non-consequentialist principles. But 'in a cool hour' (Hare 1981, 52), when we are not being forced into a decision on a particular question of action, but rather reflecting upon our general policies for answering them, we might do better by applying utilitarianism. To use a sporting analogy, players might need to deliberate in different ways on the field than they do, alongside their coaches, in a team-talk. Secondly, Smith may be right that similar limitations that lead us to need non-consequentialist motives do prevent us from using utilitarianism, to hit upon the perfect optimific M. But we have to choose an M somehow: and I have given reasons to think that using utilitarianism will help us make a better choice, with respect to maximising value, than using any other theory (or not using a theory at all).

To end this section with another worry: it might be thought that even if the need for other motives does not imply that utilitarianism is self-effacing, it is somewhat embarrassing for a moral theory that it warns against its own general application, and recommends the acceptance of conflicting principles. I think, however, that every plausible moral theory of a similar level of abstraction to utilitarianism is in the same boat. Kantians do not tend to recommend that we consult the categorical imperative before every decision. If their theory requires this, then it is implausible; people are not morally worse for using the sorts of rules of thumb we find in M. (Williams (1981a) alleges that sometimes – in his 'one thought too many' cases, they may be morally better for acting without reference to moral principles at all.) Moreover, if Kantianism is to permit competitive games, it must not only permit agents to not consult the categorical imperative before every decision, but recommend, in some circumstances, that they do not, and instead consult other principles (such as 'one should try to win') that will sometimes conflict with it. The same, I think, will go for all moral theories that are general enough to count as rivals to utilitarianism.

3. The argument from collective self-defeatingness

So much for the argument from other motives. In this section we'll examine a different argument for utilitarianism's self-effacingness. Before we come to the main argument of the section, observe that there seems to be a simple argument from the Integrity Objection to the conclusion that utilitarianism is self-effacing. The Integrity Objection showed that accepting utilitarianism precludes commitment. We may add the premise that commitment is conducive to well-being. If so, utilitarianism will recommend that we bring about commitment; therefore, it will recommend that agents do not accept utilitarianism.[23]

Again, Williams gestures at something like this argument, or at least the additional premise needed for it. He writes:

> 'It may even be that the utilitarian researcher will find that many of those with commitments, who have really identified themselves with objects outside themselves, who are thoroughly involved with other persons, or institutions, or activities or causes, are actually happier than those whose projects and wants are not like that. If so, that is an important piece of utilitarian empirical lore.' (1973, 115–16)

Let us grant Williams's piece of lore, and admit that those with commitments enjoy greater well-being than those without. Nevertheless, the argument described in the previous paragraph fails.

Utilitarianism does recommend that we act optimifically, and therefore will require us not to do things that are not conducive to our well-being, other things equal. But in the case of accepting utilitarianism, other things are not likely to be equal. Accepting utilitarianism might prevent me from having commitments, and so, granting Williams's lore, reduce my well-being. However, it might also motivate me to do other things which increase overall well-being, such as donating large amounts of money to famine relief. Taking this into

[23] Readers will notice the similarity of the arguments in this section to those of the 'incoherence objection' that featured in the previous chapter. If utilitarianism is self-effacing though, it is not incoherent in the way mentioned there, since it would not require agents not to have commitments (in virtue of not requiring them to accept utilitarianism).

account, it might still be that my accepting utilitarianism brings about more well-being than my not doing so.

Just because accepting utilitarianism prevents me from having commitments and the associated well-being, then, does not necessarily make it self-effacing for me. It does imply something else, however. It implies that if everyone accepted utilitarianism, commitment would completely disappear. Now let's grant another empirical claim: that the complete loss of commitment from the world would be disastrous for well-being, such that it could not be optimal, even considering other things un-committed people might do. (Donating money to famine relief might not do that much good, for instance, if the people it saved went on to have lives barren of commitment, or if the people involved in delivering famine relief were not committed to their cause.) If so, then given the incompatibility of utilitarianism and commitment, utilitarianism is 'collectively self-defeating' - if every individual accepted utilitarianism, its aims (greater well-being) would be worse achieved.

In some places, Parfit, Railton and Eggleston seem to suggest that it follows from the claim that a theory is collectively self-defeating that it is self-effacing for individual agents. (Parfit and Railton are talking more generally about consequentialism, which Parfit labels 'C', but presumably their arguments are supposed to apply to utilitarianism too.) Parfit writes:

> 'If we were all pure do-gooders, the outcome would be worse than it would be if we had other sets of motives. If we know this, C tells us that it would be wrong to cause ourselves to be, or to remain, pure do-gooders.' (Parfit 1984, 28)

Railton imagines someone justifying their non-consequentialist commitments on consequentialist grounds by saying:

> 'Look, it's a better world when people can have a relationship like ours and nobody could if everyone were always asking themselves who's got the most need. You'd make things worse in a hurry if you broke up those close relationships for the sake of some higher goal.' (Railton 1984, 150)[24]

Eggleston writes:

[24] Both Parfit and Railton, in these texts at least, ultimately express doubt that C is wholly self-effacing, judging it more likely that it recommends accepting some mix of consequentialist and non-consequentialist theories. But these arguments are supposed to push us towards the thought that C is self-effacing.

> ‘it can be expected that a society of act-utilitarian agents would do worse, in terms of achieving the act-utilitarian aim of maximising happiness, than would a society of agents who subscribe to other moral theories… if happier outcomes result from agents’ subscribing to some other moral theory than act utilitarianism, then—given act utilitarianism’s characteristic insistence on agents’ bringing about the happiest outcomes—act utilitarianism itself will enjoin agents to subscribe to some other moral theory than itself.’ (Eggleston 2013, 31)

These authors seem to infer from the fact that a theory is collectively self-defeating that it is self-effacing – in our case from the fact that less well-being would arise if everyone accepted utilitarianism to the claim that, according to utilitarianism, agents ought not to accept utilitarianism. But it is not in general true that if the outcome of our all doing something would be non-optimific, then utilitarianism implies that each of us ought not do it. Recall Feldman’s case from the previous chapter:

> ‘Suppose a group of adults has taken a group of children out to do some ice skating. The adults have assured the children and their parents that, in case of accident, they will do everything in their power to protect the children. Each adult in the party is a good skater and swimmer. Suppose, finally, that, while they are out skating, it just so happens that all the adults are spread out around the edge of the pond. A lone child is skating in the middle, equidistant from the adults. Suddenly, the ice breaks, and the child falls through. There is no time for consultation or deliberation. Someone must quickly save the child. However, since the ice is very thin, it would be disastrous for more than one of the adults to venture near the place where the child broke through. For if two or more were to go out, they would all fall in and all would be in profound trouble.’ (Feldman 1980, 171)

Now, imagine that you are one of the adults and you know that no other adult will go to save the child. Utilitarianism then directs you to save the child: this would result in the most well-being that your actions could produce. This is true even though if every adult did the same, a disaster would occur.

Accepting utilitarianism could be like going to save the child. That is, it could be that both (a) if we all accepted utilitarianism, this would not be conducive to well-being, and (b) that my accepting utilitarianism, given the likely behaviour of others, would maximise well-being. In such cases, utilitarianism would direct me to accept utilitarianism: it would not be self-effacing, even though it is collectively self-defeating.

The incompatibility of utilitarianism and commitment makes it quite plausible that we are in such a case. We have already seen how (a) might be met – that is, how utilitarianism might be collectively self-defeating. If accepting utilitarianism precludes commitment, then a world in which everyone accepted utilitarianism would be a world without commitment, and this might be disastrous for well-being.

And (b) could also be met. The vast majority of people do not accept utilitarianism. Many non-utilitarian moral principles are so embedded in social norms and education, and so much better suited to general biases and the interests of the powerful, that there is little prospect of this changing very much in our lifetimes. The observations of Integrity Objection give us another reason to be sceptical that utilitarianism will become a consensus position. Given utilitarianism's incompatibility with commitment, and commitment's important role in well-being (which we have granted to Williams), accepting utilitarianism means sacrificing a good degree of one's own well-being. Moreover, the nature of commitments is that they are hard to abandon, and largely people do not consider abandoning them – so embracing utilitarianism, if it involves these things, will be difficult and rare. Thus, we know that few others will accept utilitarianism, just as in my presentation of Feldman's case I stipulated that you know that no others will go to save the child.

Therefore, it is unlikely that my accepting utilitarianism will lead to a world barren of commitment. What is more likely is that I accept utilitarianism and everyone else carries on as they are – with commitments and all the well-being they bring. In this scenario, although *I* would be unable to enjoy the well-being that springs from commitment, others would continue to do so. Moreover, it is plausible that if I accepted utilitarianism, I would help them to do much more of it (for example, by expending my energies on policies that save their lives or increase their capacities) than I would if I simply pursued my own existing commitments and rejected utilitarianism. Thus it seems that precisely because commitment is conducive to personal well-being, and utilitarianism is incompatible with it, that given the likely behaviour of others it would be optimific for me to accept utilitarianism – that is, utilitarianism is not self-effacing.

To recapitulate: I am questioning the inference from collective self-defeatingness to self-effacingness. In fact, the problems associated with commitment that make it plausible that utilitarianism is collectively self-defeating also make it unlikely to be self-effacing for individual agents. There was a divergence, in Feldman's case, between what it would be expectably best for me to do and what would be best for everyone to do, because I had good reason to think that the other adults would not go to save the child. Utilitarianism's incompatibility with commitment gives us good reason to think that most other people will not accept it. It means that accepting utilitarianism is hard, as it precludes certain things that make one's life go better. This may be offset, in the person-neutral terms of the theory itself, by the good things one can bring into the lives of others. But we know that individuals do not typically sacrifice their commitments for the good of others beyond those commitments. Therefore, if accepting utilitarianism amounts to making such a sacrifice, there is little risk of everyone accepting it, and each of us has the opportunity to increase overall well-being by sacrificing some of our own to help the majority who do not accept utilitarianism to live in conditions in which they can enjoy the fruits of commitment.

The fact that utilitarianism is collectively self-defeating does imply that agents ought not (according to utilitarianism) bring it about that everyone accepts utilitarianism. However, this is not a relevant possibility for individuals in the actual world, and it is not equivalent to the claim that agents ought not bring it about that *some* agents, and in particular *they themselves*, do not accept utilitarianism. (I suspect that conflating the two is what leads Eggleston to endorse the argument from collective self-defeat to self-effacingness (2013, 42).) Therefore, this claim has no bearing on whether utilitarianism is self-effacing, as I define it.

Interestingly, Parfit and Railton seem aware, in the very works where they suggest the argument from collective self-defeat to self-effacingness, of the possibility that it could be both non-optimific if we all accepted utilitarianism (or for them, consequentialism) and optimific for each of us to accept it. Parfit writes:

> 'I know that most of us will continue to have the motives much like we have now. Most of us will love certain other people, and will have other strong desires on which most happiness depends. Since I know this, C may tell *me* to try to be a pure do-gooder. This may make the outcome better even though, if we were *all* pure do-gooders, this would make the outcome worse. If most people are *not* pure do-gooders, it may make the outcome better if a few people are.' (Parfit 1984, 30)

Railton writes that

> 'just how demanding or disruptive [complying with consequentialism] would be for an individual is a function – as it arguably should be – of how bad the state of the world is, how others typically act, what institutions exist, and how much that individual is capable of doing.' (Railton 1984, 161)

Parfit and Railton seem to recognise that which theories utilitarianism requires one to accept depends on what others accept. This being the case, there is a clear possibility that if others don't accept utilitarianism, I should, even if it would 'make things worse in a hurry' for everyone to do so. The fact that Eggleston, Parfit and Railton seem to suggest an invalid argument, and an argument that is invalid for reasons the latter two themselves note, leaves us with an interpretative puzzle. One explanation is that they were not, in these pieces, particularly interested in making the argument that utilitarianism (or consequentialism) is self-effacing; their focus is instead on what the implications would be if it were. Another is that my interpretation of their argument here is uncharitable. Later in this chapter I will suggest an alternative interpretation of their arguments. Nonetheless, what I have described here as the argument from collective self-defeatingness is flawed, whether Parfit and Railton make it or not.

4. Parfit's later argument

In later work, Parfit gives a more explicit argument for consequentialism's self-effacingness, to which he seems to have become more committed, and which is slightly different from the argument from collective self-defeat as described so far. I take it that though he directs his argument at act-consequentialism – which he often labels 'C', it is meant to apply to utilitarianism too. However, this argument suffers from similar problems to the argument from collective self-defeatingness.

Towards the end of Volume Three of *On What Matters*, Parfit writes that he finds it plausible that 'things would on the whole go better if most people accepted, not C, but some improved version of Common Sense Morality.'[25] He then says that it may also be true that 'since it would be best if everyone had the same moral beliefs, it would be best if everyone accepted some improved version of Common Sense Morality.' He finally claims that if this second statement were true, 'Act Consequentialism would imply that… everyone ought to accept, not C, but this version of Common Sense Morality.' (Parfit 2017, 415) Thus, according to my definition, C would be self-effacing.

Parfit's argument here is somewhat unclear and has so far received little attention, so it is worth unpacking. His first premise, that things would go best if most people accepted improved Common Sense Morality (let's call this theory 'CSM*') and not C, is close to the claim that C is collectively self-defeating. As I defined it, C is collectively self-defeating if, if *everyone* accepted it, the outcome would be suboptimal. Parfit's premise is the stronger claim that if *most* people accepted it, the outcome would be suboptimal – because it would be better if most people accepted CSM*. For the sake of argument, I will grant this claim.

His second premise, as stated, is really a collection of claims in some logical sequence. The first claim is that it would be best if everyone had the same moral beliefs. Another is that

[25] In this work, Parfit uses 'AC' to abbreviate act-consequentialism. I have replaced this with 'C' for consistency. This does not change the meaning, as 'C' throughout this chapter, and in Parfit's earlier work, means act-consequentialism.

it would be best if everyone accepted CSM* (and not C). To validly infer the latter from the former, as the word 'since' suggests Parfit is trying to do, we need to add the premise that a world in which everyone accepts CSM* is better than any world in which there is some other moral view that everyone accepts. This is not entailed, but is suggested, by the very first premise – that it would be better if most people accepted CSM* (and not C). (If things would be better if most people accepted something, and it is best that either everyone accepts it or no one does, then it is likely to be best that everyone accepts it.) Parfit then infers from the fact that it would be best if everyone accepted CSM* (and not C) that C is self-effacing.

This last step suffers from similar problems to the argument from collective self-defeat, as I explain below. I would also like to challenge the claim that it would be best if everyone had the same moral beliefs (note that Parfit here uses 'belief' and 'acceptance' interchangeably). At this point in Parfit's book, it seems unmotivated. Later, however, he justifies a similar claim (that 'everyone ought to have the same moral beliefs') by writing that 'Moral truths are not true only for certain people.' (2017, 420) This suggests that Parfit's motivation for his claim that it would be best if everyone had the same moral beliefs is the universal nature of moral truths: they are not true only for certain people. Since they are true for everyone, we might think, it would be best if everyone accepted them. If everyone accepted them, then everyone would have the same moral beliefs.

There are a couple of problems with this justification for the claim that it would be best if everyone had the same moral beliefs, in the context of Parfit's argument for C's self-effacingness. The first is that it depends on those beliefs being true. Perhaps it would be best if everyone accepted the truth about morality, and perhaps the truth about morality is the same for everyone. It would follow that when things go best everyone has the same moral beliefs. But this is only because everyone has true moral beliefs; it is the truth, not the sameness of their beliefs, that makes things better. So this is no argument for the view that it would be best if everyone had the same false moral beliefs. If Parfit wants to use this justification to support the view that it would be best if everyone accepts CSM*, he is assuming that CSM* is true – and therefore that C is false. But Parfit repeatedly says that the fact that C is self-effacing does not make it false – but if he motivates his argument in this way it seems that he must assume its falsehood to show its self-effacingness.

The second problem with the justification from universal moral truth is that it seems to ignore a distinction to which Parfit himself drew attention: between epistemic or intellectual oughts, and moral oughts (Parfit 1984, 43). Perhaps everyone ought to believe moral truths, and since they are true for everyone, everyone ought to have the same moral beliefs. And perhaps we can hold this in a way that doesn't make it the case that everyone ought to have the same moral beliefs simply because everyone ought to have the true moral beliefs, and so avoid the problem mentioned in the previous paragraph. For instance, we might think that because of the universality of moral truths, we should try to align our beliefs with those of others, whether their beliefs are true or false, since the fact that many others believe them, and they have decent moral intuitions about what is true for them, makes them likely to be true for us as well. But the idea that we ought to believe something because it is true, or likely to be true, is an epistemic or intellectual ought. It is, as Parfit says, another question what we ought morally to believe. Parfit's Satan examples show how these oughts can come apart: if the Devil threatened to unleash great suffering if we continued to believe certain true things, it would become immoral to believe them – but they would still be true, there would still be arguments and evidence that show that we ought (in an epistemic sense) to believe them.

To put this in terms of betterness, Parfit seems to have some reason to think that it might be best epistemically if everyone had the same moral beliefs. However, if his argument is to go through, it must be best *morally* for everyone to have the same moral beliefs. Parfit is trying to secure the claim that it would be best if everyone accepts CSM* and not C. If this is true, he says, C is self-effacing, telling everyone to accept CSM* rather than itself. But C is a *moral* theory. The claim that it would be epistemically best if everyone has the same moral beliefs, and thus if everyone accepts CSM* rather than C, implies nothing about what C requires. Parfit needs the claim that it would be *morally* best if everyone accepts CSM* rather than C. Thus he needs the premise that it would be morally best if everyone has the same moral beliefs. But he cannot motivate this by appeal to the nature of moral truths, since this could only be the basis of epistemic value.

Let's imagine that Parfit does somehow secure the claim that it would be morally best if everyone had the same moral beliefs. (Perhaps if everyone had the same moral beliefs there would be less conflict, and more time spent on constructive collaboration that improves the world.) Further, let's say that he secures his claim that the moral beliefs it is morally best for everyone to have include a belief in CSM* and not in C. He says that in that case, C would imply that everyone ought to accept CSM* rather than C. This inference, I believe, fails – for reasons that are very similar to those that invalidated the argument from collective self-defeatingness.

This is not surprising, since it is very close to the inference from collective self-defeatingness to self-effacingness. C is collectively self-defeating iff, if everyone accepted it, the outcome would be suboptimal. Parfit's premise here is that if everyone accepted CSM* rather than C, the outcome would be optimal. It follows from this premise that C is collectively self-defeating. Parfit's conclusion is that everyone ought to accept CSM* rather than C. This implies that everyone ought not to accept C – that is, that C is self-effacing for all individual agents.

The inference from collective self-defeat to self-effacingness failed because of the possibility that the outcome might be suboptimal if everyone accepts C, but given that many people will not, each of us might make things go as well as we can by accepting C. A similar possibility defeats Parfit's inference. Maybe it would be best if everyone accepts CSM* and not C. However, not everyone does accept CSM*. Perhaps, since it conforms more with common intuitions, more people accept it than accept C. But remember also that CSM*, unlike CSM, is a revisionary moral theory – by definition, it is an improvement on what most people accept. Since changing people's moral beliefs is difficult, we are a long way from a world in which everyone accepts any revisionary morality, including CSM*.

In this situation, granting all of Parfit's premises, what does C tell me to do? It tells me that if I had the power to make everyone accept a particular moral theory, I should make them accept CSM*. But I do not. I can choose, to some extent, what I accept. C tells me to accept the theory such that my acceptance of it would make things go best. But Parfit's argument does not prove that this theory is not C. It is perfectly possible that it would be best if everyone accepted CSM*, but also that given that they will not, it would be best if I accepted C. This is made most plausible when considering a practical situation Parfit was very interested in: charitable giving. Perhaps it would be best if everyone accepted CSM* and thus all the super-rich gave enough of their wealth to the poor to relieve global poverty. However, since they are not doing this, it is true for many of us that we could have a positive effect on the lives of the global poor by increasing our charitable giving above the level that CSM* demands. We might do this only if we accept C. If this is the case, it would be best (at least with respect to poverty relief) for each of us to accept C.

5. Why utilitarianism might be self-effacing for collectives

Consider the following case, adapted from Barbara Postow (1977). Fred and Mary have a garden that needs attention. Each can weed, or water, or do both, or neither. It would be best if one waters and the other weeds simultaneously. It would worst if neither task is done. Both of them watering would be disastrous, as it would waterlog their plants. However, if one of them does neither, it would be best if the other waters without weeding, since there is no time to do both in sequence, and watering is the more urgent task. In this case, it is true of each of Fred and Mary that, if the other does neither task, they ought to water without weeding. Both Fred and Mary are tired from the working week and reluctant to do either task, and each knows this is true of the other.

In sections 3 and 4 I argued that there was no valid route from the claim that utilitarianism is collectively self-defeating to the claim that it is self-effacing for individuals.

This was because of the possibility that though it would be worse if we all accepted utilitarianism, it might be optimific for each individual, given what others will accept, to accept it. In this respect, accepting utilitarianism is like watering without weeding. Whether you are Fred or Mary, given that your partner is unlikely to do either task, you should water without weeding. This is true even though it would be disastrous if you both watered, and best if both watering and weeding were done. The interest of Postow's case is this: there is a sense in which 'Fred and Mary ought not water without weeding' is true, even though it is false that either Fred or Mary ought not to water without weeding, given what the other can be expected to do.[26] This suggests that there might be a sense in which (according to utilitarianism) 'people ought not accept utilitarianism' is true, even though it is false that any particular individual ought not accept utilitarianism. That is, there could be a sense in which, consistent with my arguments above, utilitarianism is self-effacing. In this section I argue that this sense is best understood as utilitarianism being self-effacing for collectives, interpreted either as many agents jointly bearing an 'ought' or as singular group agents.

In Postow's example, 'Fred and Mary ought not water without weeding' is true when read in what we might call a non-distributive manner. According to Oliver and Smiley, 'A predicate F is distributive if it is analytic that F is true of some things iff it is true of each of them separately' (2013, 3). If 'ought not water without weeding' is read as a distributive predicate, it implies that Fred ought not water without weeding, and that Mary ought not water without weeding, which would be false according to the example. But shorn of this implication, the same sentence can be read as true, holding the facts of the example fixed. Similarly, the predicate 'ought not to accept utilitarianism' may be read distributively, in which case 'according to utilitarianism, people ought not to accept utilitarianism' would imply that utilitarianism is self-effacing for each individual. But it can also be read without this implication, that is, non-distributively. This opens the door to a sense in which utilitarianism might be self-effacing for collectives even if, as I have argued, it is not self-effacing for individuals.

According to Postow, the non-distributive reading of 'Fred and Mary ought not water without weeding' is true because Fred and Mary form a group agent that ought to see to it that the garden is both watered and weeded (see also F. Jackson 1987). But it is not necessary to posit a group agent to justify a non-distributive reading. Some philosophers believe that the predicate 'ought to X' can be borne jointly by multiple agents (T. H. Smith 2009; Estlund 2020, chap. 12; Mellor, n.d.). There are predicates like this. 'Beautiful' is one:

[26] By contrast, in Feldman's case (which is in some ways similar) there does not seem to be a sense in which the adults ought not save the child.

the multiple dots in Seurat's pointillist painting *Un dimanche après-midi à l'Île de la Grande Jatte* are beautiful, but this property does not distribute – it is not the case that each dot is beautiful (T. H. Smith 2009, 32; for broader discussion see Oliver and Smiley 2013). So we have two ways of analysing non-distributive oughts: as applying to a group agent, or as applying jointly to multiple agents.[27] Both can ascribe moral obligation to a collective without ascribing it to individual members of the collective. To return to Seurat: the art critic who is the analogue of the group agency theorist says that the painting – a single whole made up of dots – is beautiful, though each dot is not; the one who is the analogue of the joint-ought theorist says that the dots – as distinct but multiple entities – are beautiful, though each one is not.

Therefore, the claim that utilitarianism is self-effacing for some set of people could have a true non-distributive reading if either: (1) according to utilitarianism, a group agent incorporating the people in question is such that it ought not to accept utilitarianism; or (2) according to utilitarianism the people in question jointly ought not to accept utilitarianism. I will now argue that that it is plausible that some such claims are true – especially if utilitarianism is collectively self-defeating, as I have been granting.

This offers an alternative reading of Eggleston's, Parfit's and Railton's inferences criticised in Section 3. When Eggleston writes that if utilitarianism is collectively self-defeating it 'will enjoin agents to subscribe to some other moral theory than itself', 'agents' could be read not as 'each agent' – which makes the unwarranted inference to self-effacingness for individuals – but to 'agents jointly' or 'agents as a group', which makes the inference to self-effacingness jointly or for groups. Likewise, when Parfit writes that '[i]f we know [that C is collectively self-defeating], C tells us that it would be wrong to cause ourselves to be, or to remain, pure do-gooders' he might mean, not that for each of us, it would be wrong to cause our individual selves to accept consequentialism but rather that for us, jointly or as a group, it would be wrong to cause ourselves to accept consequentialism. Similarly, when Railton justifies Juan having non-consequentialist attitudes towards his wife by saying: 'You'd make things worse in a hurry if you broke up those close relationships for the sake of some higher goal', the 'you' could refer to the collective, not any individual. On this reading, Juan cannot appeal to the collective self-defeatingness of utilitarianism as a reason against his accepting it. But he can appeal to the collective self-defeatingness of utilitarianism as a reason for collectives not to accept it.

[27] I have a general preference for the group agency analysis: see Chapter Six, Section 4 for further discussion.

For the notion that utilitarianism is self-effacing for collectives to make sense, utilitarianism must be applicable to collectives (either jointly or as groups), and not just to individuals. It is, I think, within the spirit of utilitarianism to apply it as such. Many of the actions that have the most significant consequences for overall well-being are collective actions. Postow's case is a sanitised philosopher's example, but others involve climate change, structural injustice, and questions of nuclear proliferation. I return to such cases in Chapter Six, where I make a more developed case for the need to apply utilitarianism to groups – including as an answer to 'the pre-emption problem' raised by Williams's examples. For now, grant me that there are good reasons to want to apply utilitarianism collectively, either to groups or jointly.

Given this precondition, the claim that utilitarianism is self-effacing for a collective is meaningful. But is it true? The claim can be fully spelt out as 'according to utilitarianism, these people ought not accept utilitarianism'. Now we have the options introduced above, of interpreting this 'ought' jointly or as belonging to a group agent. I believe that on the joint interpretation, utilitarianism is self-effacing for us collectively, and that it is probably is self-effacing on the group interpretation – given, as I have granted throughout, that it is collectively self-defeating.

On the joint interpretation of Postow's case, Fred and Mary ought (jointly, non-distributively) not water without weeding. We can give a utilitarian justification for this: watering and weeding would be better for their garden, and therefore more conducive to well-being. Now, if utilitarianism is collectively self-defeating, then us all accepting utilitarianism is similarly less conducive to well-being than some alternative patterns of action. If this is true, it seems that in the same sense that Fred and Mary ought not water without weeding, we ought not accept utilitarianism. That is, on the joint interpretation of 'ought', utilitarianism is self-effacing for a collective for which it is collectively self-defeating.

With groups the answer is less straightforward. This is because the 'ought' is applied not to the members of the collective, but to a group, and a group's accepting utilitarianism need not entail all of its members accepting it. What is it for a group to accept a moral theory? I have so far considered acceptance as something individuals do. However, we do attribute attitudes towards moral principles to groups ('The Church believes that gay sex is sinful', 'Pakistani society has conservative views about gender', 'The University of Manchester supports academic freedom'). Sometimes we attribute non-acceptance ('The Church no longer accepts that gay sex is sinful', 'The proletariat has no ideals to realise'). Whole moral theories such as utilitarianism are less often attributed to groups in ordinary speech, but we do say things such as 'America has a Judeo-Christian morality' or 'Cuba is a Marxist state'. But we might think of anthropologists, historians or sociologists aiming to work out what general moral theories some community accepts. They might look at what some set of people seem to believe and observe their practices and from this construct a theory the group can be said to accept – see talk of 'Victorian morality' or 'Indian morality'. We might think no community is homogeneous enough in this respect for it to be straightforwardly true to say that it accepts some theory, but we can make sense of the claim. So ordinary language supports the view that groups can accept moral theories.

On many accounts of group acceptance ('summative accounts'), a group accepting something entails that most individuals who are members of the group accept it. If utilitarianism is collectively self-defeating within some group, this means that if every member accepted it that would be non-optimific. If this is true, then it is likely that if most of its members accepted utilitarianism, that would be non-optimific too. Therefore, on summative accounts of group acceptance, the group's accepting utilitarianism would likely be non-optimific, and utilitarianism would likely be self-effacing for the group.

According to other accounts of group acceptance ('non-summative accounts'), groups can accept something even if most of their members do not. There are cases motivating such accounts: a group might defer to a vocal minority, or to a leader; it might accept a compromise position that no individual member personally accepts; it might accept something simply because it has previously been decided that this is the party line. Such views about group acceptance have been used to underpin accounts of group belief (Gilbert 1989; 1994; Tuomela 1992). However, it is the nature of group acceptance that I am interested in here, rather than whether it is a necessary or sufficient condition for group belief (for discussion and answers in the negative, see Lackey 2016; 2020[28]).

On these non-summative accounts of group acceptance, the path from collective self-defeat to self-effacingness for the group is less certain. It is not entailed by the group's accepting utilitarianism that most of its members do. However, on most plausible non-summative accounts, utilitarianism's collective self-defeatingness still makes it likely that utilitarianism is self-effacing for the group. What, if not the fact that a majority of members accept some view, makes it the case that a group accepts it? Non-summative accounts tend to emphasise widespread, conventional and norm-governed behaviour within the group. For instance, the willingness of group members to let that view stand as the view of the group, to

[28] Lackey also argues that groups cannot be justified in believing something merely on the grounds of non-summative group acceptance: rather, it is a necessary condition on justified group belief of p that members of the group justifiably believe p. This is not incompatible with the view that a group ought to accept p even if its members ought not. Lackey's justification is an epistemic justification, which (crucially to her argument) aims at truth. The 'ought' in question for me is a moral ought.

publicly defend it, to rebuke group members who dissent from it or be quiet about their own dissent from it, and to allow group actions to proceed from it (Gilbert 1989; 1994; Schmitt 1994; Tuomela 1992). Thus, if a community accepted utilitarianism, the principle of utility would be proclaimed, enforced, taught to children and so on. Acting non-optimfically, or openly dissenting from C, would make one fit for rebuke or punishment by the community, or at least mark one out as 'letting the side down'. Whether individuals in this group accept utilitarianism or not, such a regime incentivises behaviour that is very similar to the behaviour of individuals who do. If utilitarianism is collectively self-defeating within the group and so it would be non-optimific if every member's accepted it, then it is likely that much of the membership acting as if they accept it would be non-optimific too. Maintaining commitment, for instance, would be difficult in such a community, as doing so sometimes requires thinking and acting in non-utilitarian ways, but such ways would be discouraged. Committed people would face criticism from others. Furthermore, such a system is likely to cause more individuals to personally accept utilitarianism. It would be the moral theory to which they are most exposed, which best explains the public actions of their community, and the costs of complying with it in such a community would be much lower than the costs in our community. Thus, the group's accepting utilitarianism, on this account of acceptance, could cause (though it does not entail) that most of its members accept utilitarianism. As argued above, if utilitarianism is collectively self-defeating within some group, it would likely be suboptimal if most of its members accepted utilitarianism. If the group's accepting utilitarianism would cause the latter, utilitarianism will likely say that the group ought not accept utilitarianism.[29]

So, if it is true that utilitarianism is collectively self-defeating (as I have assumed for the sake of argument in this paper), then it is likely self-effacing for collectives. If we interpret self-effacingness with a joint ought, then the members of the collective would not be bringing about the most well-being by accepting utilitarianism, and therefore jointly ought not to, according to utilitarianism itself. If we interpret it with a group ought, we then have a choice as to whether group acceptance should be analysed summatively or non-summatively. On summative accounts, if utilitarianism is collectively self-defeating it is probably self-effacing

[29] One might worry that if, as I have argued, efforts to persuade individuals to accept C are unlikely to be successful, but things would be better if a few individuals accepted C, groups ought, according to C, to try to persuade their members to accept it, with the expected consequence that only a few will. My response is that a large part of the reason that individuals are unlikely to accept C is that their groups do not accept it, and so disincentivise acceptance. If important groups did accept C, this may very well lead to many individuals – more than would be optimal – accepting it. Thanks to Joe Horton for pressing this worry.

for the group, since if all members accepting it would be non-optimific, then most members accepting it would probably be non-optimific too. On non-summative accounts, if utilitarianism is collectively self-defeating it is probably self-effacing for the group, though the story is more complex. People in a group that accepts utilitarianism, even non-summatively, are likely to think and act in ways very similar to those who accept utilitarianism, and many are likely to accept utilitarianism themselves as a result. Such a group would produce fairly similar results to one in which everyone accepts utilitarianism, which would be, given collective self-defeatingness, not optimific.

A variant of the argument from collective self-defeat therefore succeeds. So does a variant of Parfit's later argument, or at least that last inference in it: if things go better if most people in a group accept CSM* rather than C, then the group probably ought to accept CSM* rather than C, according to C. This argument shows that C is likely self-effacing for groups, or jointly, even if not for individuals.

In Section 2 of this chapter, I argued that utilitarianism is unlikely to be self-effacing for individuals because of the need for individuals to accept utilitarianism in order to find, maintain, revise and justify the other deliberative principles it would be optimific to accept. Groups have the same needs, insofar as they are to act optimifically. We might think that, as with individuals, this shows that utilitarianism will direct groups to accept utilitarianism. Accepting it would help them to hit upon and successfully use the principles that will be most conducive to well-being. However, I will now argue that groups have other ways of ensuring that they find and reliably employ optimific principles.

Parfit writes, of a society in which no one accepts C, but rather CSM*:

> 'Given changes in the world, and in our technology, it might later come to be true that the outcome would be better if we revised our moral beliefs. But if we no longer believe C, and now believe [CSM*], we would not be led to make these needed revisions in our morality… This suggests that the most that could be true is that C is partly self-effacing. It might be better if most people cause themselves to believe some other theory [having previously believed C], by some process of self-deception that, to succeed, must also be forgotten. But, as a precaution, a few people should continue to believe C, and should keep convincing evidence about this self-deception… If the moral theory believed by most did become disastrous, the few could then produce their evidence. When most people learn that their moral beliefs are the result of self-deception, this would undermine those beliefs, and prevent the disaster.' (Parfit 1984, 42)

I will not discuss the details of Parfit's social engineering here. The important point to note is simply that we can imagine political and social arrangements such that most people do not accept utilitarianism, and the moral beliefs that are widespread in the community can nevertheless be revised so that they are usually optimific. I think that, on either view of group acceptance, the society Parfit describes would not count as accepting utilitarianism. So, a group can avoid accepting utilitarianism and still solve the problem of revising the principles it does accept in an optimific manner.

It is plausible that this could be done by more realistic and less elitist means than Parfit's caste of consequentialists. A community could reject utilitarianism but accept the narrower principle that the community should accept the principles such that their accepting them makes things go best. Call this 'political utilitarianism'. It is not incoherent to accept this but reject utilitarianism. Many moral theories are utilitarian in some respects but not in others (think, for example, of a theory that says that agents ought to do whatever maximises well-being, apart from when this would breach a deontological side constraint, or a personal prerogative). Many moral theories make different claims about what groups should do to those about what individuals should do (think, for example, about Rawlsian theories that require states not to act on particular conceptions of the good, whilst permitting individuals to do so). Most, or even all, of its members could accept political utilitarianism. When the group's principles need revising to retain optimificness, these members of the group would lead it to do so, performing a role akin to Parfit's 'few' who accept C. This is another way in which a group might revise its principles in an optimific manner without accepting utilitarianism. It also provides a way for groups to justify their acceptance of principles to their individual members. Most members (like the group itself) accept political utilitarianism, so would understand why the group adopted the principles it did.

So utilitarianism is more likely to be self-effacing for collectives than for individuals (at least, given its collective self-defeatingness). This is in itself an interesting result. Moreover, that utilitarianism is self-effacing for collectives has implications regarding the Integrity Objection, which mitigate it to some extent.

If utilitarianism is self-effacing for collectives, it may not ask for social pressure to be brought to bear to make people to abandon their commitments for the greater good, even if it asks individuals to make such sacrifices. It may not ask for groups or individuals to rebuke or punish people for pursuing their commitments when there are mosquito nets to be bought, even whilst it asks individuals themselves to pivot to mosquito nets. The group need not educate or incentivise people into utilitarian ways of thinking, even if individuals ought to adopt them on their own account. It may, to maintain, revise and justify the optimific non-utilitarian principles it accepts, need to inculcate acceptance of political utilitarianism. But note that political utilitarianism need not conflict with commitment in the way that full-blown utilitarianism does – for it does not require attitudes of impartiality or dispensability from individuals, and if it is conducive to well-being that most people have commitments, political utilitarianism will issue principles that encourage or at least permit them to do so.

6. Esotericism

I have granted for the sake of argument that utilitarianism is collectively self-defeating. I have argued that even granting this, it is unlikely to be self-effacing for individual agents, given what others will actually do. However, collective self-defeatingness suggests that utilitarianism might be self-effacing for collectives, interpreted jointly or as group agents.

These possibilities might provoke the worry that utilitarianism is 'esoteric'. A theory T is esoteric iff it that implies that some individuals ought to accept T but others ought not. Sidgwick thought that his version of utilitarianism was esoteric, and he regarded this as regrettable (1884, 484–85). Esotericism seems problematic, I think, because it is inegalitarian: it seems somehow in tension with the idea that all individuals are moral equals. Williams's description of esotericism as 'Government House', a reference to colonial administration, speaks to this worry. Parfit's proposal mentioned at the end of the previous section might be a kind of esotericism.

We should first note that my objections to the argument from self-defeatingness and to Parfit's later argument do not rely on utilitarianism being esoteric. If utilitarianism is esoteric, it might direct some agents to accept it even if it is collectively self-defeating. I might be one of the chosen few, whose acceptance of utilitarianism would maximise well-being even though if everyone, including the un-chosen, accepted it the results would be non-optimific. However, even if utilitarianism is not esoteric it could direct me to accept utilitarianism, despite being collectively self-defeating. Utilitarianism might direct *every* individual to accept it, even if the universal acceptance of utilitarianism would be suboptimal.

To see this, consider Feldman's case again. It might be not just that *I* have good reason to think that the other adults will not go to save the child, but that *every adult* has good reason to think that no other adult will go to save the child. Utilitarianism does not direct some of them to save the child and some of them to stay where they are: it directs each of them to go to save the child, since this would maximise well-being given what the others are likely to do. It says this even though if they all do it, disaster will occur. Likewise, we all have good reason to predict that very few others will accept utilitarianism, and therefore utilitarianism might direct each individual to accept it, even though if everyone did so, the results would be suboptimal.

Neither does the notion that a theory is self-effacing for groups but not for individuals necessarily make it esoteric. It could be that every single individual in the group ought to accept utilitarianism but the group ought not. The notion does entail that there is a class of agents (groups) whose moral obligations differ from those of another class (individuals). But this is not problematic, at least not for the egalitarian reasons that we might find esoteric theories problematic. It is not intuitive that individuals and groups have equal moral standing; unlike two different individuals, an individual and a group are different in kind in a morally salient way. In fact, part of the motivation for applying moral obligations to groups is that groups can do things, and be obliged to do things, that individuals cannot.

If a theory is esoteric – it seems that those who ought to accept are also obliged by that theory to keep those who ought not in the dark, to hide the truth from them and manipulate them into false but expedient moral beliefs, as if they were colonial subjects. (For instance, the British Raj found the four *varna* caste system expedient as it divided Indian society and provided a ruling class who would collaborate with them, and so entrenched and exacerbated the principles of caste, even though few, if any, British rulers believed the Laws of Manu to be true.) Even if every individual ought to accept utilitarianism – and therefore it is not, strictly speaking, esoteric – if it is collectively self-defeating a similar worry persists. For each of us, according to utilitarianism, it would be true that although we ought to accept utilitarianism, we ought not bring it about that everyone else accept it. In fact, if things looked as if they were going that way, we might be obliged to step in and ensure that a sufficient number of people went on not accepting utilitarianism. The same goes if utilitarianism is self-effacing for groups: we ought not to bring it about that our group accepts utilitarianism, and we ought to step in to ensure that they do not, if it seems as if they might. To fulfil these obligations might involve the sorts of deceit and manipulation that make Government House rule so distasteful.

Or would it? I have already noted two things about the world we live in. The first is that none of us has the power to bring it about that everyone accepts a moral theory. The second is that – especially taking utilitarianism's incompatibility with commitment into account – it is unlikely that most people, and most groups, will accept utilitarianism. The first means that utilitarianism does not place an obligation on us not to bring it about that everyone else accept utilitarianism. Utilitarianism only tells us to bring about the most well-being we can. It says nothing about things that are out of our control. Granted, it implies that if we could bring about the universal acceptance of utilitarianism, we ought not. But since the antecedent will never be true, this obligation is never actually placed upon us. The second means that it is unlikely that we will be in a situation in which everyone, or our group, would accept utilitarianism unless we persuaded them otherwise. So we do not have this obligation either. Moreover, since most people won't be persuaded to accept utilitarianism, even if universal or group acceptance of utilitarianism would be very bad, we need not stop persuading people to accept it. We can sincerely endorse utilitarianism, trying to persuade everyone we meet to accept it (remember, in this scenario, it is true of each person that they ought to accept utilitarianism), safe in the knowledge that most people will go on rejecting it. No manipulation is necessary to encourage people not to accept a truth that it is costly for them to accept.

This response relies upon the claim that most people are unlikely to accept utilitarianism. We might think this in itself is cause for concern. If utilitarianism is true, why should it be difficult to accept? Not all true theories, of course, are likely to be widely accepted – think of true theories in physics or biology, which are accepted only by experts (some true scientific theories, of course, may be yet to be accepted by anyone, because they are yet to be considered). But the reason that these scientific theories are not widely accepted is because they are very difficult to understand without technical training, and because most people never think about them or need to use them. We might think that moral theories should not be like that. The truth about morality should not be accessible only to experts. Morality is something that almost every person reflects on, has opinions about, and uses. Therefore, if utilitarianism will never be widely accepted, does this show it is false, or otherwise inadequate?[30]

I think not. There are explanations for why utilitarianism is unlikely to be widely accepted that do not suggest that it is false. One is that accepting it is demanding, including by disregarding commitments. People are not disposed to accept claims that imply that they ought to do things that are costly to themselves, or that conflict with their commitments. This does not just go for moral theories, but empirical facts too. Wealthy people find it hard to accept that they are wealthy relative to others: together with widely held moral beliefs, this would imply that they ought to forego some of their wealth, and perhaps advocate for economic changes that would make them poorer. Scientists (and perhaps even philosophers!) find it hard to accept theories that contradict ones they have endorsed themselves. The latter example suggests another explanation for why utilitarianism is unlikely to be widely accepted: it is not widely accepted now, and moral beliefs are difficult to change. We are not educated to be utilitarians. Perhaps, if utilitarianism is collectively self-

[30] Thanks to Ulrike Heuer for pressing this objection.

defeating, this is for the best. But given this education, it is unsurprising that most of us will never accept utilitarianism. This in no way suggests that the theory is false.

It should also be noted that no moral theory enjoys universal acceptance, or is ever likely to do so. Certain vague principles might (many of them endorsed, as good heuristics, by utilitarianism). But nothing approaching utilitarianism's level of generality does: not Kant's formulae, nor Scanlon's principles, nor Aristotle's virtues. If there is something to the thought that a moral theory should be generally acceptable, then all moral theories will have to explain why they are not in fact generally accepted. I have just briefly sketched how utilitarianism might do that.

7. Conclusions

I have argued that neither the need for other motives, nor the fact of self-defeatingness, nor the premises of Parfit's later argument secure the conclusion that utilitarianism is self-effacing for individual agents. In fact, I think it is unlikely to be so, given the need for individuals to acquire, update and motivate themselves to follow optimific principles. However, collective self-defeatingness suggests that utilitarianism might be self-effacing for collectives, interpreted as multiple individuals jointly bearing an obligation or as group agents. This raises certain problems associated with esotericism, which I have suggested some ways of avoiding.

As long as utilitarianism is not self-effacing for individual agents, Williams's charge that it 'ushers itself from the scene' is not apt. It instead casts itself in a rather important role, as a tool of deliberation for individuals. I have admitted that this role will be limited: in many cases, individuals would bring about better results by employing non-consequentialist principles. But I have argued that utilitarianism will still be useful in some cases: not least in selecting the non-consequentialist principles one should accept and use in other cases. One can use a moral theory to deliberate without (a) using it constantly, and (b) wanting everyone else to accept it. This, I think, is what utilitarianism requires individuals' attitudes towards utilitarianism to be.

What are the implications of my arguments for the Integrity Objection? Well, as long as utilitarianism directs agents to accept utilitarianism, it will make commitment (as Williams defined it) impermissible. I will argue in the next chapter that, contra Williams, it is right to do so. But this chapter has already suggested how some of the negatives associated with the preclusion of commitment are mitigated. Other motives, such as those stemming from love and deeply held projects, are permitted and even encouraged by utilitarianism. Moreover, if utilitarianism is self-effacing for groups, then even if it holds that individuals ought to accept utilitarianism it does not imply that groups ought to do anything to enforce this acceptance amongst its members. Abandoning commitment might be demanding, but no group ought to demand it of its members, even if morality does.

Chapter Four: Dispensability, Mutual Interdependence and Alienation from Projects

0. Abstract

Williams condemns utilitarianism for requiring us to regard our projects as dispensable, and thus preventing us from being properly committed to them. In this chapter, I argue against commitment as Williams defines it, drawing upon insights from the socialist tradition as well as mainstream analytic moral philosophy. I show that given the mutual interdependence of individuals (a phenomenon emphasised by socialists) several appealing non-utilitarian moral principles also require us to regard our projects as dispensable. This means that those who endorse these principles cannot appeal to Williams's argument against utilitarianism. It also puts pressure on his thought that moral theories ought to permit commitment – in fact, it suggests that they ought not.

Regarding one's projects as dispensable may be alienating, and this may motivate us to hang onto commitment and reject these non-utilitarian principles along with utilitarianism. However, commitment also threatens a kind of alienation – from other people. Drawing upon the socialist tradition again, I argue that avoiding this form of alienation is necessary for proper engagement with our projects.

1. Introduction

I argued in Chapter Two that Williams's Integrity Objection is best understood as raising what I called 'the psychological problem': utilitarianism requires a psychology that is incompatible with commitment. To restate Williams's objection schematically:

1. Having a project as a commitment is incompatible with regarding that project impartially or as dispensable.
2. If one accepts utilitarianism, one regards all projects impartially and as dispensable.
3. By 1 and 2, those who accept utilitarianism cannot have commitments.

Since this is supposed to move us to reject utilitarianism as a moral theory, Williams must endorse something like that following adequacy condition on moral theories: an adequate moral theory should permit agents to have commitments.

In Chapter Three I examined one way to resist Williams's objection – claiming that utilitarianism is self-effacing, and so requires agents not to accept utilitarianism. This would mean that even if 1, 2 and 3 are true, utilitarianism need not fall short of the implicit adequacy condition – agents would be permitted (or even required) not to accept utilitarianism, and so may be able to abide by utilitarianism's requirements whilst having commitments. I argued that utilitarianism is unlikely to be self-effacing for individual agents, so this response to Williams's objection is closed. I also argued that utilitarianism might be self-effacing for groups, and this would mitigate some of the problems that might arise with the employment of a moral theory that does not respect commitments. This finding should reduce the appeal of the adequacy condition, though perhaps to a small extent.

In this chapter, I make a more explicit argument that this adequacy condition should be rejected. Moral theories need not permit commitment – at least, as Williams defines it, whereby being committed to a project entails not regarding it as dispensable (I will address impartiality in Chapter Five). I show that given the mutual interdependence of individuals, several appealing non-utilitarian moral principles also imply that one should regard one's projects as dispensable. Thus, anyone who endorses these principles cannot reject utilitarianism on the grounds that it is incompatible with commitment. Insofar as these other principles are plausible, this suggests that commitment is likely to be morally impermissible, which would make Williams's implicit adequacy condition mistaken.

Recall Chapter Two's explication of what it is to regard a project as dispensable. It is not simply that one would dispense with it in some circumstances. It is rather than one entertains as alternatives outcomes in which one dispenses with it. Entertaining some outcome as an alternative is not merely conceiving of it. It is being willing to conceive of it within the constraints set by one's outlook on life. Commitments set such constraints: a committed agent is unwilling to conceive of outcomes in which they dispense with their projects. This does not mean they never do, but that when they do, a novel and (to them) unsettling mode of deliberation is required.

In contrast, to a utilitarian agent, all moral decisions call for the same kind of deliberation – basically, cost-benefit analysis.[31] The only inputs to utilitarian deliberation are facts about the well-being that a course of action will produce, and, as the cases of Jim and George suggest, there will always be possible alternatives in which dispensing with a project maximises well-being. Therefore, the utilitarian must entertain alternatives in which they abandon any of their projects. I will argue in Section 3 that moral agents who comply with several other moral theories must join them in this, given the mutual interdependence that I explicate in Section 2.

[31] At least, insofar as they employ utilitarianism as a decision-procedure. I argued in Chapter Three that they ought to do this sometimes, though not always. Bracket this caveat for now.

Regarding one's projects as dispensable may alienate one from them, and this may motivate us to hang onto commitment and reject these non-utilitarian moral principles along with utilitarianism. However, commitment also threatens a kind of alienation – from other people. Drawing upon the socialist tradition and again appealing to the facts of mutual interdependence, I argue that avoiding this form of alienation is not only important in itself, but is also a precondition of properly engaging with our projects.

Note that I am not arguing that we ought not have the kinds of projects that can become commitments. That is, I am not arguing against moral convictions, political struggle, cultural pursuits, intimate relationships and so on. I am arguing against pursuing these things *as commitments* – where this implies not regarding them as dispensable – rather than as mere projects. I am also not arguing against many of the attitudes that may in ordinary language be referred to as 'commitment'. Williams uses the term in a particular way, according to which it is a conceptual truth that committed agents cannot regard projects to which they are committed as dispensable. It is against commitment so defined that I argue.

2. Mutual interdependence

Some of our projects – such as desires for food and water – are naturally determined. Williams does not think that such projects can be commitments. They are qualitatively identical with the projects of all other humans; they are not, in Williams's words 'individual and permeated with character' (1973, 111).[32] Thus it does not show admirable robustness, or one's distinctive way of seeing the world, for one to satisfy one's own desires for food and water at the expense of others.

The projects to which we can be committed are things such as cultural pursuits, careers, political causes and moral convictions. They can define us, and we can hold these projects in distinctive ways. They do not tend to be naturally determined. Often, we choose them. But we do not make the choice in conditions of our own making – and the choices we make affect many others. Our projects are mutually interdependent. The core of my argument is that this fact, together with some attractive moral principles, supports the view that we should regard our projects as dispensable.

Socialists traditionally emphasise the fact of our mutual interdependence, as have some mainstream analytic moral philosophers. Marx describes several aspects of mutual interdependence in this pithy sentence from the *Economic and Philosophic Manuscripts*:

[32] Typically, at least: the hunger striker's desire for food has a distinctive meaning, for instance. I thank Peter Railton for this example.

> 'Not only is the material of my activity given to me as a social product (as is even the language in which the thinker is active): my own existence is social activity, and therefore that which I make of myself, I make of myself for society and with the consciousness of myself as a social being.' (1988, 105)

The material point is the least deniable, but perhaps the most overlooked. To pursue any activity one needs certain material resources, and these resources are socially produced. Most basically, simply to live, we need food, water, shelter, healthcare and the like. Once we are alive, pursuing the projects that make our lives meaningful requires particular resources: practising medicine requires drugs and surgeries; birdwatching requires coats, notebooks and binoculars; marriage (in the conventional form) requires rings and a place to live together. Some projects require other people's presence (doctors require patients; marriage requires a spouse and celebrants at a wedding). Some projects require time; this depends on an economy that produces enough (and distributes enough of what it produces to you) to permit you to spend some of your waking hours doing more than merely pursuing survival. In the modern economy, having these resources depends on the actions of countless others across the world.[33]

Secondly, regardless of resource constraints, projects depend on the concepts that others recognise. What things are available projects for us depends upon which social forms are recognised in our society. This is one interpretation of Marx's parenthetical claim about language. Joseph Raz – neither a socialist nor a utilitarian – emphasised a similar point (1986; 1996). As he writes:

> 'one cannot practise medicine except in a society in which such a practice is recognized. Notice that in principle one may be born into a society with no medical practice or knowledge endowed with an innate knowledge of medicine. One could then cure many diseases, but one could not be a medical doctor, of the kind we have in our society. It takes more than medical knowledge or curing powers to do that. A doctor participates in a complex social form, involving general recognition of a

[33] See also Locke, writing centuries earlier in the story of our economy's increasing complexity and globalisation: ''Twould be a strange Catalogue of things, that Industry provided and made use of, about every Loaf of Bread, before it came to our use, if we could trace them; Iron, Wood, Leather, Bark, Timber, Stone, Bricks, Coals, Lime, Cloth, Dying-Drugs, Pitch, Tar, Masts, Ropes, and all the Materials made use of in the Ship, that brought any of the Commodities made use of by any of the Workmen, to any part of the Work, all which, 'twould be almost impossible, at least too long, to reckon up.' (For recent discussion of the relevance of this fact to a different set of normative questions, see Brudney n.d., where Locke's passage is quoted.)

> medical practice, its social organization, its status in society, its conventions about which matters are addressed to doctors and which not… and its conventions about the suitable relations between doctors and their patients.' (1986, 310–11)

As for careers, so for other pursuits:

> 'Bird watching seems to be what any sighted person in the vicinity of birds can do. And so he can, except that that would not make him into a bird watcher. He can be that only in a society where this, or at least some other animal tracking activities, are recognized as leisure activities, and which furthermore shares certain attitudes to natural life generally.' (1986, 311)

Raz's point is that the kinds of projects which we could adopt as commitments (what Raz calls 'comprehensive goals') are only on the table for us because of social conditions.[34] These conditions are made up of the attitudes and actions of large numbers of other people in our society: the respect they have for doctors, the way they relate to birds and nature, and so on. Without others doing certain things and being certain ways the projects that are available to us would not be.[35] Moral convictions, such as George's and Jim's, may not require material resources for their pursuit. But they require certain ways of life, conceptions, institutions and recognition, and so depend on society in this second way. Williams himself recognises a similar point in later work (Williams 1985a).

Not only are our lives materially dependent on others, then, but also the range of possible lifestyles available to us is determined by the attitudes of others in society. As Bakunin, Marx's factional enemy in the nineteenth-century socialist movement, puts it:

> 'Man becomes conscious of himself and his humanity only in society and only by the collective action of the whole society. He frees himself from the yoke of external

[34] See also Walzer (1983, 6–10). Walzer suggests, going beyond Raz, that all goods – or at least, all those whose distribution is the concern of justice – are social products.

[35] It might be thought that this only goes for a subset of the projects that could be commitments, namely, ones that accord with prevailing social norms. But Raz argues that its scope is wider. Even when our projects involve transgressing social norms, they can only involve that because those norms exist. Raz considers a couple who pursue an open marriage (1986, 309). This is not a social form that society normalises. But it is conceivable only because there is a standard social form of marriage, which the couple use as a basis for innovation. Even further, we might consider a couple who make it central to their identity that they do not marry – as an act of rebellion against patriarchal norms, perhaps. Their pursuit of this project still depends on there being an institution of marriage: you cannot rebel against what does not exist. Such transgressive projects still depend on social forms, and hence on the actions and attitudes of others.

> nature by collective and social labour, which alone can transform the earth into an abode favourable to the development of humanity. Without such material emancipation the intellectual and moral emancipation of the individual is impossible.' (1973, 236–37)

To the material and conceptual bounty of society we could add the special kind of meaning our lives – or at least many of the activities within them – attain from their effects on others. As Marx says, I make myself 'for society'. Scheffler makes a similar point, reflecting on what it would mean for us if humans were to go extinct a generation from now:

> 'many of the activities that we had previously regarded as worthwhile would no longer seem to us as appealing. We would see less reason to engage in them. Some of those activities might even seem completely pointless. We might see no reason at all to engage in them. To be sure, some activities, such as spending time with family and friends, would almost certainly continue to seem worthwhile to most people. Overall, however, our capacity to find value in our activities would be seriously eroded.' (2018, 43)

Consider someone who builds their life around producing art, or furthering scientific knowledge. If there were no other people to enjoy that art or use that knowledge, their life would be less meaningful. Similarly, for Marx, meaningful production in general is production for other people (Kandiyali 2020), and it is the tragedy of capitalist production that it obscures the others for whom we produce (we produce for our own wage, for our employer, for a market – but not usually directly for another) and thus alienates us from our own working selves. To sum up, we have three ways in which we depend on others: they produce the material resources that make our lives possible; their attitudes determine the concepts that determine the lifestyles that are available to us; doing things for them adds meaning to our lives.

Our pursuit of projects, then, is shaped by others. But, of course, I am an other to others. The implication is that just as my projects are partly produced by others, I play a role in producing theirs. That is to say, as far as the pursuit of projects goes, we are mutually interdependent (Marx and Engels 2000b, 185) with respect to one another. I believe that this fact makes tenable utilitarianism's insistence that we regard our projects as dispensable.

3. Dispensability

In this section I offer three appealing moral principles that imply that we ought to regard our projects as dispensable. An agent regards a project as dispensable if they entertain as alternatives outcomes in which they dispense with that project. Agents who accept utilitarianism regard their projects as dispensable because, as Williams points out, the utilitarian outlook entertains all outcomes as alternatives, applying the same cost-benefit standard to each. This will include – as in the cases of George and Jim – outcomes in which agents are required by utilitarianism to dispense with their projects.

In this section, I will describe three other principles, weaker and more appealing than utilitarianism, that each also require agents to entertain alternatives in which they dispense with their projects. Therefore, they each imply that agents ought to regard their projects as dispensable, and hence that commitment, as defined by Williams, is impermissible.

a. Rescue

In her response to Williams, Elizabeth Ashford (2000) argues that all plausible moral theories will ask agents to step aside from their projects in situations when others are in serious danger, immediate to the agent, from which the agent could save them at the cost of abandoning one of their projects. Though I do not endorse Ashford's reading of Williams, I adopt this element of her response. The principle at stake is something like the following:

> *Rescue: When one is able to save others in one's immediate vicinity from serious and urgent dangers one ought to consider doing so.*

Rescue should have broader appeal than utilitarianism. It is in fact a less demanding principle than Ashford seems to endorse, since it requires agents merely to *consider* saving others: one could in theory comply with Rescue without doing any actual saving. Many utilitarians endorse Rescue, since they believe that saving people from serious, urgent danger tends to be conducive to well-being, and this makes considering helping in such cases a good general strategy (though there may be rare circumstances where this is outweighed by countervailing considerations). Utilitarians typically place no weight on the immediacy of the endangered to the agent: famously, Singer (1972) extrapolates from the duty to save a child from drowning in a nearby pond to the duty to save children on the other side of the world from starvation.[36] This means that utilitarians will probably endorse a stronger version of Rescue, without the 'immediate vicinity' clause, as well as endorsing Rescue as written. Many

[36] Rescue is not the same as Singer's principle: 'if it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it.' (1972, 231) Rescue is weaker in that it does not require doing anything beyond considering, and does not apply to all bad things – but it is stronger in that it does not carry Singer's 'comparable moral importance' caveat.

non-utilitarians think immediacy *is* morally significant. Many also think that we are not obliged to help people *whenever* doing so would be more conducive to well-being. These non-utilitarians could still endorse Rescue, since 'serious and urgent dangers' is a narrower category than 'threats to the optimisation of well-being'. Ashford notes that Williams himself, in later work, endorses something like Rescue (Williams 1985a, 186).

Rescue falls foul of commitment for the same reason that utilitarianism does. In some circumstances, namely those in which it is necessary to save others immediate to one from serious and urgent danger, one ought, according to Rescue, to consider dispensing with one's projects. One ought, for instance, to at least consider missing an interview for a job about which one deeply cares to save a drowning child. As noted above, that there are some extreme circumstances in which an agent will dispense with their projects is not sufficient to undermine commitment. It is rather if such circumstances are entertained as alternatives within the constraints set by the agent's outlook on life, if they do not appear as 'beyond the limits of their moral world', that commitment is undermined. So the question is whether Rescue requires an outlook on life that makes projects dispensable in this way. I think that there is enough chance that Rescue suggests considering dispensing with projects with sufficient frequency that complying with it requires such an outlook.

Why think that we are frequently in situations in which we can save others, immediate to us, from serious and urgent dangers, at the cost of our projects? There are not children drowning in front of us on every commute. But as we saw in the previous section, our lives are intertwined with others around the world. This fact, in times of war, natural disaster and poverty, connects us to many people who are in serious and urgent danger. Our interdependence with them should make them count as 'immediate' to us. Ashford quotes Williams's own words:

> 'We should be more concerned about the sufferings of people elsewhere.... We should not banish the category of immediacy, but we must consider what for us, in the modern world, should properly count as immediacy...' (1985a, 186)

Williams does not go on to propose an answer to the question of what should properly count as immediacy. But the way he poses the question implies that what counts as immediate has expanded in modernity to cover a wider geographical range. Thus, immediacy cannot be fixed by geographical proximity. Moreover, we should expect that the determinants of immediacy explain its expansion over time. If mutual interdependence were one such determinant, this would be explanatory. Centuries ago, most people were interdependent with a small number of others in their local community. Back then, immediacy in the sense relevant to moral principles such as Rescue would be geographically limited. Nowadays, there is a global network of social relations that constitute mutual interdependence between just about every person on the planet, as I argued in the previous section. This is why the category of immediacy is to be expanded.[37] We should extend 'immediacy' to mean, not simply the child drowning in the pond on your own street, but the child scouring the refuse heap halfway across the world where your old plastic bags are shipped, and their mother who stitched your trainers. Modernity has intertwined our lives with theirs in similar ways to how our lives are intertwined with people in our hometown. If immediacy extends so far, then every day there are many people facing serious and urgent dangers, who are immediate to us.

One might worry that extending immediacy so far exhausts it of meaning. If everyone is immediate to us, nobody is – or at least, there is no need for an immediacy clause in moral principles like the one in Rescue. But though basing immediacy on mutual interdependence expands the category, it does not universalise it. We are not in such relations with uncontacted tribes, with extra-terrestrial life (should it exist) or with far future generations (should they exist). And indeed many of us do think that our duties of rescue towards these beings are less, or perhaps non-existent. To serve my purposes here, what is important is that immediacy in the modern, interdependent world is likely to put us in touch with people in serious and urgent danger. A sad fact about the world is that few of us need to expand the concept through many degrees of separation to find such people.

Can we save them from those dangers? Ashford believes so: for people towards the top of the global wealth distribution in particular, she thinks, there are cases in which one's charitable giving could save lives, and these are so frequent that attending to them will often mean abandoning one's projects. As she puts it:

> 'the current state of the world is a constant emergency situation; there are continually persons whose vital interests are threatened and, given modern communications, the relatively well-off are continually able to help them.' (2000, 430)

For instance, it might mean abandoning your vocation in philosophy to take up a more lucrative career that will enable you to donate more money to charity, or, in Railton's example (1984, 159), failing to see your spouse so that you can spend your time fundraising, thereby risking your marriage.

This claim is based on some non-obvious empirical claims about how effective charitable giving in fact is (Macaskill 2015 makes the case for giving; see Wenar 2011; and Budolfson

[37] It also constitutes an explanation other than moral progress for the phenomenon Singer calls 'moral circle expansion'.

and Spears 2019 for qualified dissent).[38] If they are right, it would not only be in extreme cases, but very often, that we could save others immediate to us from serious and urgent dangers. Therefore, Rescue will require us to entertain alternatives in which we dispense with our projects on a frequent basis; thus, it precludes commitment to our projects.

Even if we doubt that the likes of Singer, MacAskill and Ashford are correct about how frequently we are in a position to save others at the cost of one of our projects, Rescue may still require such attitudes. We ought to have some credence in their pro-giving claims: they are sincere, intelligent and well-informed. If we endorse Rescue, then their being right would alter our moral duties significantly. So we should at least entertain the possibility that they are right, try to evaluate their claims, and be prepared to respond if we become convinced by them. Rescue says that we should respond, if they are right, by considering dispensing with our projects. Therefore, when we are investigating whether they are right, we should be prepared to consider dispensing with our projects, if our inquiry vindicates them. But this preparation itself amounts to entertaining as alternatives outcomes in which we dispense with our projects; therefore, of regarding our projects as dispensable. To put it another way, if you are committed to your projects, in Williams's sense, then you hold that if Ashford *et al* turn out to be correct you will contravene Rescue, because you rule out in advance that you could dispense with those projects to which you're committed. If you truly endorse Rescue and think Ashford's claims have enough credibility to be worth investigating, this would be a strange position to hold.

So Rescue – a principle with wider appeal than utilitarianism – is also at odds with commitment, given our mutual interdependence and at least some chance that the empirical claims of people like Ashford are true. But perhaps you wholeheartedly reject those empirical claims, or Rescue itself. There are at least two further principles with similar implications for commitment.

b. Non-Instrumentalisation

The next principle has a Kantian flavour: it concerns the immorality of treating others as mere means. As Parfit puts it:

> 'we treat someone *as a means* when we make use of this person's abilities, activities or body to help us achieve some aim… we treat someone *merely as a means* if we

[38] Note too that there may be other ways to save people from serious and urgent dangers, such as political activism or scientific research. So even if giving is ineffective, those of us who can be effective in these ways might frequently face the demands of Rescue.

> both treat this person as a means, and regard this person as a mere instrument or tool: someone whose well-being and moral claims we ignore, and whom we would treat in whatever ways would best achieve our aims.' (Parfit 2011, 213)

Parfit rejects the principle that an act is wrong if it involves using someone merely as a means. I think he has good reasons for doing so. In a case he gives (2011, 231), a gangster who regards everyone but his own family as instruments uses the body of another person to save his own child's life during an earthquake, causing minor injury to this other person. This action is not wrong. But Parfit endorses the claim that one ought not to *regard* other people as mere means (2011, 232). There is something wrong with the gangster, but it is the attitude he has towards the other person, not the fact that he uses him to save his child. We may draw from Parfit this principle:

> *Non-Instrumentalisation: If one makes use of some other person's abilities, activities or body to pursue some project then one should sometimes be prepared to dispense with one's project for their sake.*

The gangster violates Non-Instrumentalisation not because he makes use of another person's body to save his child, but because he does so without considering that person's moral claims – the gangster would not have held back from his aim of saving his child, using this person's body, even if that person had a good moral claim not to be so used (for instance, that they would have been killed or worse).

Mutual interdependence entails that in pursuing my projects, I make use of countless other people's abilities, activities and bodies to help me achieve my aims. It implies, therefore, that the antecedent of Non-Instrumentalisation is true for all of our projects. We cannot avoid treating others as means. Therefore, to comply with Non-Instrumentalisation we must, for all projects, sometimes being prepared to dispense with our projects for the sake of those on whom our success in achieving those projects relies. If one accepts Non-Instrumentalisation, therefore, one must entertain as alternatives outcomes in which one dispenses with one's projects for the sake of others – not to entertain such possibilities would be to instrumentalise those others on whom one's pursuit of one's project depends.

The thought is this: we use others – this is unavoidable given mutual interdependence – but we shouldn't use them as tools. Avoiding using them as tools means entertaining possibilities in which we dispense with our projects for their sake. What those possibilities are is left undetermined, and so Non-Instrumentalisation is compatible with a wide variety of further moral principles specifying them.

Note that Non-Instrumentalisation does not prohibit using others to pursue one's projects; nor pursuing them whenever they are at odds with the interests of those you use to achieve them. It simply says that one should *sometimes be prepared* to dispense with one's projects for their sake. One could comply with Non-Instrumentalisation by simply being prepared to dispense with one's projects when pursuing them would cause very severe, irreparable damage to those on whom you rely. One would then only dispense with one's project in extreme situations. But one should have the attitude to others, that they are sufficiently important to sometimes consider dispensing with our projects for their sake, as part of one's regular outlook on life – on pain of instrumentalising them. This would mean regarding one's projects, given that they rely on others, as dispensable.

One clearly does not have to be a utilitarian to accept Non-Instrumentalisation. It has Kantian inspiration, and a good degree of intuitive plausibility. Furthermore, since the specification of the 'sake' of others that we should consider are unspecified, a wide range of moral theories could endorse Non-Instrumentalisation. It could ask us to consider other people's needs, interests, rights, or moral claims of any kind. It is consistent with utilitarianism, with 'sake' read as well-being, as we are likely to perform actions which more reliably maximise well-being when we are prepared to forego our own aims to increase the well-being of others. It might be thought that there is some tension between Non-Instrumentalisation and utilitarianism, as the former emphasises the sake of those whom we use, whilst utilitarianism weights the well-being of each person equally. However, Non-Instrumentalisation simply encourages us to consider those whom we use; it does not claim that they have greater weight than anyone else's. That said, given the kind of mutual interdependence that occurs in the modern world, Non-Instrumentalisation itself might encourage us to consider the claims of every person in the global economy.[39]

c. Responsibility for injustice

Iris Marion Young proposes the following principle as part of her 'social connection model' of responsibility (2006, 102–3).

> *Responsibility: all who contribute by their actions to the structural processes that produce injustice have responsibilities to work to remedy those injustices.*

This principle may be more controversial than the previous two, but many find it plausible. Such a principle is necessary, Young argues, given that many injustices are largely determined by structural processes rather than the direct effects of action. Young's central

[39] Onora O'Neill (1996, chap. 4) makes an argument that somewhat resembles this last thought.

example is the global clothing industry. Many of the workers who manufacture clothes face serious injustices, including overwork, precarity, low pay, unsafe conditions and restrictions on their rights to organise. Mostly these injustices are the direct effects of the actions of those who employ or manage them – usually small enterprises in poor countries, dependent on larger exporters, who are in turn dependent on large multinationals who sell to consumers in rich countries. However, Young says,

> 'In this system, each of the links in the chain believes itself to be operating close to the margin in a highly competitive environment, and usually is under heavy pressure to meet orders at low cost by firms higher up the chain.' (2006, 110)

Manufacturers can truthfully say that if they mitigated the condition of the workers, they would be outcompeted by a rival who would treat workers more harshly. Multinationals can truthfully say that if they paid manufacturers more, allowing them to improve working conditions, they too would be outcompeted. Their actions, then, do not generate the injustice. What of the consumers to whom they ultimately sell, and whose demand for cheap clothing puts downward pressure on costs throughout the chain? Well, these consumers are often themselves not wealthy, and are usually just trying to clothe themselves and their families in accordance with their own budget constraints and prevailing social norms. Furthermore, buying fewer clothes may make things even worse for the workers by putting them out of work, and paying more for their clothes may simply increase the profits of multinationals.

There seems to be, in this case, no single agent or agents on whom this injustice can be blamed. But, Young notes, we have a lingering feeling that all of the agents mentioned bear some responsibility for the injustice. The explanation for this is that structures are not some alien force: they are, as noted above, produced by people. Our actions produce structures, and this gives us some responsibility to do something about them when they cause injustice, namely, to work to remedy it.

What does this have to do with projects and commitment? If the argument of Section 2 is correct, our pursuit of our projects contributes to structures. This is largely how they affect (amongst other things) the projects of others. There are, according to any plausible view, many unjust structures in the world. Because there are so many plausible views about justice, we cannot be confident which these are; nor, given the complexity of the social world, can we be confident to which structures our pursuit of projects contributes. Often, as Young says, these contributions are unwitting and many degrees removed from identifiable harms. (Pursuing a career writing for fashion magazines may contribute to structures that exploit workers on the other side of the world, that are enmeshed with local class and gender structures, which affect the political system of that country…). Therefore, if we accept Responsibility as a principle, we should acknowledge that our pursuit of our projects will give us widespread responsibilities to work to remedy injustices.

This work to remedy injustices will often involve entertaining as alternatives outcomes in which one dispenses with one's projects. It is not that when we realise that our pursuit of some project contributes to unjust structures we ought to give up that project. Young emphasises that in cases of structural injustice, the remedy will tend to be collective, rather than individual, action. One consumer abandoning their project of trying to keep up with the latest affordable fashion will not in itself remove the injustice; this is part of what makes the injustice structural. The work of remedying such injustice, therefore,

> 'is ultimately political responsibility… [it] involves joining with others to organize collective action to reform unjust structures… Thus, discharging my responsibility in relation to sweatshop workers might involve trying to persuade others that the treatment of these workers is unacceptable and that we collectively can alter social practices and institutional rules and priorities to prevent such treatment.' (Young 2006, 123)[40]

However, engaging in that political struggle, I think, will necessitate entertaining possibilities in which we give up our projects. If one struggles for a change in the social structures to which one's project contributes, one aims for a world in which one's project would be radically altered. If the clothing industry really did reorganise to remove injustices in its supply chain, it is likely that the project of following fashion would be very different to what it is today. If the activist fashionista is serious about such change, they will entertain the possibility of their project no longer being available to them, as a result of their success in the work that Responsibility sets them. Even if one thinks such complete success is unlikely, and work to remedy injustices will usually result in piecemeal change compatible with retaining one's project, that it is one's aim means that one must entertain it as an alternative, and a desirable one.

Therefore, a moral framework including Responsibility, given that the pursuit of our projects is likely to contribute to unjust structures (as the fact of mutual interdependence suggests), implies that agents ought to entertain as alternatives possibilities in which they dispense with projects. Responsibility is, like Rescue and Non-Instrumentalisation, a principle with broader appeal than utilitarianism. Utilitarians may agree with it. As long as

[40] Pogge (2002) also makes a compelling case for the responsibilities of citizens of wealthy countries to engage in such struggles.

unjust situations instantiate less than optimal well-being, utilitarians believe that everyone ought to do whatever they can to remedy injustice: including those who contribute to the processes that cause it. Utilitarians do not believe moral responsibility ends there, of course, and this is one reason that it is sometimes thought too demanding. But those who think this can endorse Responsibility. Responsibility could be true if nobody has a responsibility to remedy injustice apart from those who contribute to the processes that produce it. It could also be true on a variety of conceptions of justice, and on a variety of conceptions of what remedying such injustices in fact involves – its demands could therefore be much weaker than the demands of utilitarianism (although they need not be).

d. The upshot

In this section I have argued that (like utilitarianism) Rescue, Non-Instrumentalisation and Responsibility imply that we ought to regard our projects as dispensable, given the mutual interdependence of human lives and projects. This means that any moral framework incorporating any of these principles will be inadequate, according to the condition suggested by Williams's Integrity Objection to utilitarianism. Anyone who endorses any of those three principles, therefore, cannot dismiss utilitarianism on the grounds of this objection. Moreover, insofar as it is likely that at least one of utilitarianism, Rescue, Non-Instrumentalisation and Responsibility is true, Williams's view that an adequate moral theory should permit commitment is false.

We do not just have an argument against this adequacy condition for moral theories. We also have an argument against the permissibility of individuals having commitments as Williams defined them. Williams's argument showed that utilitarianism made commitment impermissible. This was meant as a *reductio* against utilitarianism. However, we now see that three other principles with broader appeal than utilitarianism also make commitment impermissible. This strengthens the case for thinking that commitment is indeed impermissible: otherwise, not only utilitarianism but also each of these three other principles is mistaken.

There are some ways of modifying the Integrity Objection to respond to the arguments of this section. One would be to note that, whilst utilitarianism implies that anyone in any context should at least sometimes regards their projects as dispensable, Rescue, Non-Instrumentalisation and Responsibility only imply that we should regard our projects as dispensable contingently, in light of the facts of mutual interdependence. One might hold that a moral theory is inadequate if it *necessarily* precludes agents from having commitments, but not if it does so only contingently. Such a view would count utilitarianism as inadequate but not make the same judgment against all theories implying Rescue, Non-Instrumentalisation or Responsibility. But it is hard to see how such a modified condition could be motivated: we are making moral theories for the actual world, after all.[41]

Alternatively, one could modify the definition of commitment such that agents could count as committed to some project whilst regarding it as dispensable, but not whilst regarding it as dispensable *for utilitarian reasons*. Thus, non-utilitarian theories implying Rescue, Non-Instrumentalisation or Responsibility would not preclude commitment, though they would have us regard our projects as dispensable. Again, there is a need to motivate this definition. It reduces the Integrity Objection to a special version of the claim that utilitarianism does not specify the correct moral reasons. That may be so, but it is a more general objection to utilitarianism, and it is unclear why Williams would focus his discussion on commitments if this were his claim.

Ultimately, I think that Williams himself would have responded to my arguments so far by rejecting Rescue, Non-Instrumentalisation and Responsibility on the grounds that they alienated agents from their projects. It is to this challenge that I now turn.

4. Alienation from projects

After launching the Integrity Objection in his 'A Critique of Utilitarianism', Williams went on to make somewhat similar objections to Kantianism (1981a), and to morality in general (1985a, chap. 10). So he may have been aware that commitment was in tension with a wide range of moral theories, and not just utilitarianism. My argument so far has vindicated this position, and as such should concern people who are turned against utilitarianism by the Integrity Objection, but want (possibly unlike Williams) to hold onto morality. But the argument could also be read as a *reductio* against morality – or at least, against moralities that include Rescue, Non-Instrumentalisation, Responsibility or utilitarianism, by those want to affirm commitment.

[41] Han van Wietmarschen points out that there might be a further important distinction between utilitarianism, Rescue and Non-Instrumentalisation on the one hand, and Responsibility on the other: the latter implies that we ought to regard our projects as dispensable only in *unjust* worlds, while the others do not. (Though in a more just world it would probably be the case that the other three principles less frequently required us to abandon our projects.) One might think that it is most implausible that commitment should be precluded in a just world, so that Responsibility is a more palatable principle. I am unconvinced about this, largely because a just world is a remote possibility, and extremely difficult for me to imagine.

One compelling reason to reject such moralities would be that to regard all of one's projects as dispensable, as these principles require, alienates one from them. Williams himself uses the language of alienation in his critique of utilitarianism (1973, 131), as does Railton in his response (1984, 134–35). In this section I use socialist insights to argue that commitment too is a source of alienation: alienation from other people, which, given mutual interdependence, prevents proper engagement with our projects.

Why would regarding one's projects as dispensable alienate us from them, as Williams suggests? The thought, as I understand it, is this. Alienation is a problematic separation between a self and other that properly belong together (Leopold 2018). Our projects should be close to us, especially those projects so important to us that they could become commitments. To regard these projects as dispensable, to make holding them conditional on the moral demands of utilitarianism (or Rescue, or Non-Instrumentalisation, or Responsibility), is to separate them from oneself. It is to add a psychological distance – an extra thought ('ought I maintain this project?') – between recognising the project as yours and pursuing it.[42]

I do not wish to straightforwardly reject this argument. But I will argue that *failure* to regard a project as dispensable, as one must with a commitment, is also alienating in an important way. This weakens the position of those who would cling to commitment, in the face of my arguments above, due to concerns about alienation.

The alienation involved in commitment is from our fellow humans. Just as our projects ought to be close to us, we ought to be close to other people. Utilitarianism, Rescue, Non-Instrumentalisation and Responsibility each, in their own way, ask us to entertain as alternatives outcomes in which we abandon our projects for other people. Commitment to those projects, which would preclude such an attitude, may bring us closer to our projects, but drive us further from those others. They do not figure in our thoughts when commitment is involved. Commitment therefore alienates us from them. As Railton puts it:

> 'because of his very willingness to question his life morally, [the agent who regards his projects as dispensable] avoids a sort of alienation not sufficiently discussed - alienation from others, beyond one's intimate ties. Individuals who will not or cannot allow questions to arise about what they are doing from a broader perspective are in an important way cut off from their society and the larger world.' (1984, 151)

[42] Note the similarity with Williams's 'one thought too many' argument. In the case he uses there, a man is alienated from his drowning wife by considering whether saving her could be morally justified (1981a, 17–19).

We thus appear to have two kinds of alienation to choose between. On the one hand, regarding our projects as dispensable alienates us from them, by making us think twice before acting on them. On the other, steadfast commitment to our projects alienates us from others, by pushing them from our thoughts altogether when commitments are involved.

One might think that it is more important to avoid the first form of alienation than the second. It is through pursuing our projects, one might say, that we realise ourselves; it is our projects – in particular those projects which could be commitments – that make us who we are. It may be good, too, to be part of society, but in distancing ourselves from our projects in order to be closer to others we subsume ourselves into the collective. We are primarily, on this view, active individuals, most fulfilled when we are authors of our own lives. If the opposing view is stated as the inverse of this – that we are primarily mere parts of society and therefore must avoid alienation from others at the cost of our individuality – it is not particularly appealing. Indeed, as Williams put it in later work, 'unless I am propelled forward by the conatus of desire, project and interest, it is unclear why I should go on at all.' (1981a, 12)

But there is more to be said for the importance of avoiding alienation from others. There is not a simple trade-off between one's projects and one's relations with others. I have shown that at least one kind of alienation from one's projects – a conditionality and therefore additional distance in thought – falls out of regarding one's projects as dispensable, whilst a kind of alienation from others – absence from thought altogether, when commitments are involved – is implied by commitment. But there are other ways in which one can fail to engage properly with one's projects. Firstly, one could misunderstand them. Secondly, one could fail to find them meaningful. These failures may be called other forms of alienation, but whatever they are called, they are problematic. I will now argue that given our mutual interdependence, avoiding both of these problems together requires non-alienation from other people. Therefore, it is not that we are faced with a choice between proper engagement with our projects, or with other people. It is rather that alienation from other people prevents proper engagement with our projects – so that the ideal of an active self-authoring individual whose projects provide the foundation for their life is at least as threatened by alienation from others as by the alienation from projects that regarding them as dispensable implies.

As I have already argued, our projects are only possible, and some only conceivable and meaningful, because of the actions and attitudes of other people. Therefore, to consider them in isolation from others, as one's own private creation, is to misunderstand them. This is not to say that projects come to us as finished social products that individuals simply choose, or have thrust upon them, adding nothing of their own. It is rather that each person's projects (unique as they might be) are forged (with differing degrees of creativity) from elements produced socially. If we want to understand our own projects, and thus our own lives, we have to understand them in this social context – that is, we have to consider their connections with, and reliance upon, others.

Now, some may experience this realisation as deflating. Some prefer to think of themselves as 'self-made men'[43] and their projects as an expression of their private will rather than social forces. 'Why,' they may ask, 'should I build my life around projects if those projects are so contingent on other people?' The answer is that other people matter. Recall Marx's dictum that 'that which I make of myself, I make of myself for society', and Scheffler's reflections on the meaning afforded to our lives by the supposition that future generations will enjoy what we produce. If other people did not matter to us, such motivation and meaning would be mysterious. Insofar as we value others, regarding our projects as in part their creation should not make those projects less meaningful; on the contrary, we would find meaning in our projects belonging to a vast social endeavour. There is something powerful in regarding our projects as products of millions or billions of valuable beings, as the products of centuries or millennia of history – and as aimed towards having effects upon further valuable beings in a future that is the next chapter of an unfolding human tale. We feel this power only when we have the value of others before our minds.

In this way, we can acknowledge our mutual interdependence without destroying the motivational importance of our projects. They move us because we care about others. Our care for others would be what gives us reason to 'go on'. It is no accident that socialists such as Marx and Bakunin, who emphasise our mutual interdependence, propose that we hold others in high regard. By doing so, they can embrace the ideal of the active project-oriented self without turning to the individualism of self-made men. Avoiding alienation from others becomes a precondition for realising ourselves. Bakunin writes:

> 'The liberty of every individual is only the reflection of his own humanity, or his human right through the conscience of all free men, his brothers and equals…
>
> I am not myself free or human until or unless I recognise the freedom and humanity of all my fellowmen…
>
> Only by respecting their human character do I respect my own. A cannibal who devours his prisoner… is not a man but a beast. A slave owner is not a man but a master. By denying the humanity of his slaves he abrogates his own humanity…' (1973, 137)

[43] I use 'men' advisedly. See Walker (2008, 137–60) for an exploration of the connections between the Integrity Objection and gender.

Marx makes much the same point in his 'Comments on James Mill', a rare glimpse of his vision of communism. Characteristically, Marx focuses on production rather than freedom, but the thought, like Bakunin's, is that our individual value is affirmed by recognising the value of others.

> 'Supposing that we had produced in a human manner; each of us would in his production have doubly affirmed himself and his fellow men… In your enjoyment or use of my product I would have had the direct enjoyment of realising that I had both satisfied a human need by my work and also objectified the human essence and therefore fashioned for another human being the object that met his need… In my expression of my life I would have fashioned your expression of your life, and thus in my own activity have realised my own essence, my human, my communal essence.
>
> In that case our products would be like so many mirrors, out of which our essence shone.' (2000b, 132)

The metaphor of reflection in both passages suggests a positive correlation between how one values others and how one values oneself. The brighter and more vivid an object reflected in a mirror, the brighter and more vivid is the image produced. The closer one is to other people, the closer – according to these socialists – one is to one's own projects.[44]

To recapitulate: if one is to properly understand one's projects, one must recognise their dependence on other people; if one is to find that this enhances, rather than undermines, the meaning of one's projects one must recognise the value of others. But how is such recognition possible when one refuses to allow considerations of other people to enter into practical deliberation where projects are concerned, as must be the case with commitment (as Williams defines it)? Thus, the alienation from others involved in commitment undermines not only our relations with others but our relations with our own projects. Moreover, if one does recognise these things when considering one's projects, it is a small step to asking whether one's project really ought to be pursued – or be given up, for the sake of those valued others on whom it depends. Thus, regarding our projects as dispensable might be a symptom of understanding them and finding them meaningful, as much as it is a threat to our close, unalienated relationship with them.

[44] As Marx and Bakunin emphasise, such closeness might not be achievable simply by a change of mindset, but might necessitate a change of social conditions. Ideologies and socio-economic structures which obscure our mutual interdependence, and which imply that some people's lives do not matter, are significant barriers to each of us properly engaging with our projects.

This section began with the thought that regarding our projects as dispensable alienates us from our projects, and that this is such a cost that we ought to reject otherwise attractive moral principles that imply it. I have argued that commitment also demands a kind of alienation – alienation from others. But this latter form of alienation is also a threat to our relationship with our projects, since to both understand them and find meaning in them requires recognition of others' role in them, and others' value. That it is crucially important to engage properly with one's projects, therefore, is not grounds to favour commitment. In fact, it may be further reason to regard one's projects as dispensable.

5. Conclusion

Williams condemns utilitarianism on the grounds that it requires us to have attitudes that make commitment impossible: namely, it requires us to regard our projects impartially and as dispensable. In this chapter I have focused on the requirement to regard our projects as dispensable. I showed that, given the conditions of mutual interdependence in which we live, three other principles, more widely assented to than utilitarianism, also make such a requirement. Abiding by such a requirement may alienate us somewhat from our projects. But those same conditions of mutual interdependence suggest that there is another form of alienation that it is important to avoid – alienation from our fellow people – and it is commitment, rather than utilitarianism and these other principles, which leads to it. Given mutual interdependence, avoiding such alienation is a precondition of properly engaging with the social nature of our projects.

As such, this chapter has argued against the adequacy condition for moral theories that lies behind Williams's Integrity Objection: that theories must permit agents to have commitments. Commitment, as Williams defines it, because it means regarding some projects as indispensable, is impermissible according to a range of moral principles applied in an interconnected world. It also alienates us from others on whom we depend for our projects. Thus, utilitarians have a good response to Williams: our theory is incompatible with commitment, but this is no mark against it, as commitment is ethically unattractive.

A more complete response, however, must tie up some loose ends. Utilitarianism precludes commitment not only because it requires us to regard our projects as dispensable, but also because it requires impartiality. That we ought to regard our projects as dispensable, as I have argued in this chapter, is enough to undermine commitment. But impartiality might still be objectionable, for reasons other than its incompatibility with commitment. In particular, it is alleged that impartial regard for our projects alienates us from ourselves. In Chapter Five, I will attempt to rebut this allegation.

Chapter Five: Impartiality and Alienation from the Self

0. Abstract

Even if we ought to regard our projects as dispensable, and thus not be committed to them, it may still be claimed that utilitarianism is defective in virtue of its requirement to regard them impartially. Williams and others suggest that adopting the impartiality of utilitarianism (and of other modern moral theories) alienates agents from themselves. In this chapter I respond to this objection, showing that impartial attitudes can be rationally motivated without eliminating the special relationships between individuals and their projects, without the detachment from the self that 'a view from nowhere' seems to involve, and without reducing one's valuation of oneself. I draw once again on socialist ideas about the connections between people, as well as on Buddhist and Afro-communitarian thought.

1. Introduction

In this chapter I complete my response to Williams's Integrity Objection. Recall that the objection runs like this:

1. Having a project as a commitment is incompatible with regarding that project impartially or as dispensable.
2. If one accepts utilitarianism, one regards all projects impartially and as dispensable.
3. By 1 and 2, those who accept utilitarianism cannot have commitments.

I accept that this argument is sound. I also argued in Chapter Three (against the self-effacing response) that utilitarianism requires agents to accept utilitarianism, and thus that it follows from the objection that utilitarianism requires agents not to have commitments, at least in Williams's sense. In Chapter Four I argued that utilitarianism is justified in asking us to regard our projects as dispensable, and thus that commitment in Williams's sense is morally problematic. I also argued that commitment alienates us from other people, and thereby undermines our relationships with our projects. I thereby made the case that utilitarianism's incompatibility with commitment does not condemn it as a moral theory.

However, it may still be claimed that utilitarianism's requirement that one regard one's projects impartially is objectionable, because it alienates one in another serious way: namely, from oneself. This objection is suggested by Williams's charge that asking an agent 'to regard as one satisfaction among others… a project or attitude around which he has built his life' is 'in the most literal sense, an attack on his integrity.' (1973, 116–17) In this chapter I respond to this claim.

To repeat some definitions from Chapter Two: to regard projects impartially is to refrain from valuing one project more than another simply because of whose it is. Most importantly, someone who is impartial does not value their own projects more than those of others simply because they are their own. This captures Williams's description that the projects of a utilitarian agent are, to them, 'one satisfaction among others'.

Why does utilitarianism require such impartiality? As Scanlon puts it (1998, 95–100), to value X is to take oneself to have reasons for certain attitudes and actions towards X. In Scanlon's view these attitudes and actions may be several, and may vary depending on what X is. With projects, very generally, we can say that a crucial part of valuing them is taking oneself to have reasons to pursue them (when they are one's own), and to assist in them (when they belong to others). To value project Y over project Z, then, would be to take oneself to have stronger reasons to pursue/assist in Y than Z. For a utilitarian (as far as morality goes) one has stronger reasons to do one thing than another if and only if it would result in greater overall well-being. I should do what leads to more well-being, whether that involves acting on my projects or those of a stranger. But if I value my own projects more, simply because they are mine, I take myself to have more reason to pursue them than to assist in others, even when the latter would result in greater well-being. Such an action would be wrong, according to utilitarianism.

Insofar as one employs utilitarianism as a decision-procedure, then, one regards one's projects impartially. Whether one takes oneself to have stronger reason to pursue/assist in one project rather than another does not depend on whose projects they are, but simply on the well-being that would result from one's pursuing or assisting actions.

Utilitarianism's impartiality with respect to projects is just one species of a more general impartiality. For any thing that might attach to an individual – not just projects, but also interests, needs, well-being, preferences, and so on – utilitarianism will deny that one has stronger reasons to promote that thing simply because of the identity of the person who has it. The utilitarian agent does not take themselves to have stronger reasons to satisfy their preferences, promote their well-being, guarantee their needs, and so on, simply because they are theirs. They recognise stronger moral reasons to promote one thing over another only if it is more conducive to well-being. We can express this very general impartiality by saying that utilitarianism values each person equally, and thus that agents do so, insofar as they employ utilitarianism as a decision-procedure.

The objection is that such impartial attitudes alienate us from ourselves. In the previous chapter, I focussed on a specific kind of alienation – a distance in thought – between an agent and their projects. In this chapter I do not specify my notion of alienation so tightly, but rather use it as a general term for defective engagement, or the absence of proper engagement, with the self. I show how such attitudes need not preclude our having projects, need not involve taking up a position detached from the self and need not lead us to reduce our self-valuation. Each of these is a denial that impartiality is alienating, on some interpretation of 'alienation'.

The aim of this chapter, then, is to deny that utilitarianism's impartiality need alienate the utilitarian agent from themselves. It is worth clarifying what I am not aiming to do. I am not arguing that partial attitudes are impermissible. Neither am I taking a position on a related debate which is often characterised in terms of impartiality: whether agents have special prerogatives or obligations to favour themselves or those close to them, such as their children[45] - my focus is on attitudes rather than actions. I am trying to explain, simply, how one might adopt utilitarianism's impartial attitudes whilst avoiding alienation from the self.

2. Can a utilitarian agent have projects?

The 'insofar' clauses in the previous subsection are important. I argued in Chapter Three that utilitarianism is not self-effacing. That is, it requires us to sometimes employ utilitarianism itself as a decision-procedure. However, it does not require us to do so to the complete exclusion of others. Sometimes employing other decision-procedures would bring about more well-being. Sometimes the optimific decision-procedure will involve valuing one's projects more than others simply because they are one's own – and more generally, valuing oneself more than others. In these cases utilitarianism permits, and even requires, partial attitudes. However, since utilitarianism is sometimes itself to be the decision-procedure, it requires impartiality too.[46]

[45] For a defence of this kind of impartiality, see Crisp (2018). For attempts to make room for partiality of this kind within broadly utilitarian moral theories, see Jackson (1991) and Goodin (1988).

[46] One might worry that utilitarian impartiality is not psychologically possible. If so, the objection to utilitarianism that it requires impartiality falls, however alienating impartiality is. Utilitarianism does not require agents to do impossible things, but to do the thing, of all things possible for them, that maximises well-being. So if impartial attitudes are impossible, utilitarianism does not recommend we adopt them. In any case, this is not the objection I am concerned with here. I am concerned with the objection that a utilitarian agent's impartiality, successfully achieved, would alienate them from themselves.

This caveat helps us to reject one concern about the alienating nature of utilitarian impartiality: that it prevents the pursuit of one's own projects.

Consider first Virginia Woolf's description of the supposed ideal of late Victorian womanhood, 'the Angel in the House':

> 'She was intensely sympathetic. She was immensely charming. She was utterly unselfish. She excelled in the difficult arts of family life. She sacrificed herself daily. If there was chicken, she took the leg; if there was a draught she sat in it – in short she was so constituted that she never had a mind or a wish of her own, but preferred to sympathize always with the minds and wishes of others.' (Woolf 1943, 150; for discussion see Driver 2005)

Woolf claims that women who are like this (or who aspire to be like this) will not be able to pursue creative projects, which require their authors to have minds and wishes of their own. This would plausibly constitute a kind of alienation from oneself.[47]

Utilitarianism need not require us to be Angels-in-the-House. Very often, developing and pursuing one's own projects and passions maximises well-being. One usually has a better idea of one's own needs than those of others. One is usually better able to fulfil one's own needs than those of others, too. And developing one's distinctive character and skills and productive capacities opens up many more opportunities for future beneficence. Self-sacrifice can expend our reserves of do-gooding before we get the chance to do the most good that we can. For all these reasons (and more) utilitarianism will likely recommend that much of the time we *are* partial towards our own well-being, needs and projects (see also Railton 1984; Driver 2005).

When we are partial, we are not employing utilitarianism as a decision-procedure, but instead other principles and motives. As I argued in Chapter Three, utilitarianism will *sometimes* require us to employ utilitarianism as a decision-procedure, to look at the world through the theory's eyes, which involves impartiality. As Williams defines them, this precludes commitments – but, as I argued in the previous chapter, commitments as Williams defines them are ethically unattractive. But it does not preclude us from – unlike

[47] Note two things about the Angel's alienation. Firstly, it may not be felt as a problem by its subject. A woman raised in the kind of society Woolf describes may well strongly identify with and prefer her angelic role. Thus, if there is alienation here, it is 'objective' rather than 'subjective'. Secondly, as Woolf describes her it is not quite true to say that the Angel-in-the-House is impartial: she seems to value her own interests, projects and well-being *less than*, rather than equally to, those of others.

Woolf's Angel – having our own projects, and showing partiality to them, and to ourselves more generally, some of the time.

Another way in which utilitarian impartiality might make our projects disappear is more conceptual. It might be said that impartiality expunges the special reasons we have to pursue our projects, and if A has no special reason to pursue project P, P just isn't a project of A's. Scheffler writes:

> 'to value a project of one's own is, among other things, to see it as giving one reasons for action in a way that other people's projects do not, and in a way that other comparably valuable activities in which one might engage do not... If I do not see myself as having any more reason to attend to my own projects and goals than I do to engage in other activities or to attend to the projects and goals of other people, then it no longer makes sense to think of them as my projects and goals at all, still less to think that I value them non-instrumentally.' (2010, 105)

One response is that if Scheffler only means it is necessary for agents to *see* their projects as giving them reasons, utilitarianism may have no problem with this, for the considerations mentioned above – at least some of the time. If it is necessary for them to live valuable lives, it may maximise well-being for agents to sometimes be partial towards themselves and their projects, that is, to take themselves to have reasons to pursue them simply because they are their own – even though they really have such reasons due to an impartial duty to maximise well-being.

Utilitarians can in fact go further. They can affirm that agent actually *have* special reasons regarding their own projects. Recall Scanlon's definition of valuing: to value X is to take oneself to have reasons for certain attitudes and actions towards X. I stated above that utilitarianism denies that we have reasons to value some project *more* than another simply because it is our own. But as Scanlon emphasises, valuing as he defines it is not simply a matter of degree but also of kind: the attitudes and actions constitutive of valuing vary depending on what is being valued. It may well be that different attitudes are involved in valuing my own projects and valuing other people's, even when is value them equally. This is why, for instance, it felt natural to say that I 'pursue' my own projects but 'assist in' those of others.

Utilitarianism can also recognise non-moral reasons for each person to value their own projects more, simply because they are their own. As we saw in Chapter 2, Section 4 utilitarians tend to assert that it is rational for individuals to maximise their own well-being and use this claim in arguments for utilitarianism, making morality a generalised version of prudential rationality (Harsanyi 1977). As Scanlon notes (1998, 126–33), prudent individuals rarely make decisions by working out which option would maximise their well-being; rather, they pursue their projects, which usually brings well-being with it. Utilitarians can affirm the rationality of this, if not its morality. They may also endorse reasons that are neither moral nor prudential, such as reasons stemming from roles, identities and projects themselves. The utilitarian need not deny that a footballer has a reason to shoot at goal, even if doing so would not maximise well-being (either overall or hers). They simply deny that this is a moral reason. Similarly, those involved in a particular project might have non-moral, non-prudential reason to pursue it. But insofar as utilitarians think morality is important, such reasons ought to be revisable in the light of their impact upon overall well-being.

So, the requirement of impartiality need not preclude utilitarian agents from having their own projects, either by making them constantly self-sacrificing or by dismissing special reasons for projects. This is one way in which utilitarian impartiality is not alienating.

My argument has been founded on the premise that utilitarianism does not make it impermissible to show partiality. It merely requires that sometimes – in Hare's 'cool hour', when one considers one's life from a moral perspective – one ought to regard them impartially (1981, 52). This is important as a wider point about the dialectic of this chapter, and the debate about impartiality and alienation more generally. The charge against utilitarianism is not that showing partiality sometimes is necessary for proper engagement with the self (utilitarianism can permit this), but that taking up impartial attitudes sometimes make such engagement impossible.

3. Does impartiality require detachment from the self?

One allegedly alienating aspect of utilitarian impartiality is that it is impersonal. The utilitarian agent cares about persons, it is said, only as receptacles of an impersonal value (well-being). On the one hand this charge is odd, since well-being just is what is good for persons. On the other hand, if we imagine a utilitarian agent thinking along the following lines, their attitudes do seem objectionably impersonal:

> 'Bob is in agony. My goal is to maximize utility, i.e., the balance of pleasure over pain. There is some agony (namely, Bob's) that I am in a position to relieve. Doing so would serve my goal. So I will act to relieve Bob's suffering.' (Chappell 2015, 324)

This seems like a kind of alienation from Bob, and, if the agent in question were themselves Bob, they would be alienated from themselves. To my mind, this objection has been well-answered by Richard Yetter Chappell (2015) and Elizabeth Ashford (2005; 2021). Utilitarians need not think in the way described above. Rather, utilitarians may affirm that well-being matters only because it matters for persons, and deny that there is 'just one thing, the global happiness, that is good. Instead, there is my happiness, your happiness, Bob's, and Sally's, which are all equally weighty but nonetheless distinct intrinsic goods.' (Chappell 2015, 328) There is no inconsistency between this view and the claim that the morally right option is that which maximises overall well-being, interpreted as the sum of these distinct goods. When we conceptualise utilitarianism as asking us to maximise not some abstract good, but rather 'the inclusive combination of each person's lifetime well-being', it seems, as Ashford puts it, not impersonal but 'omnipersonal' (2021, 173).[48]

Critics might still think that coming to 'omnipersonal' impartiality involves taking up a perspective that is detached from the particularities of one's own life. This perspective is variously described as a 'view from nowhere' (Nagel 1989), 'the point of view of the universe' (Sidgwick 1884) or that of the 'impartial spectator' (Smith 2002). Sometimes it is held that the view from this perspective is necessarily identical with the view of true moral theory. It is not this that I criticise here. Rather, it is the thought that in order to reason impartially, agents must try to occupy this perspective. This is how impartiality is often introduced – for instance, Parfit writes:

> 'We have an impartial point of view when we are considering possible events that that would affect or involve people who are all strangers to us. When our actual point of view is not impartial, we can think about possible events from an imagined impartial point of view. We can do that by imagining possible events that are relevantly similar, except that the people involved are all strangers to us.' (2011, 41)

From such a point of view, my project (or well-being, etc.) would appear not as mine but as some stranger's, and thus I would value it no more than anyone else's.

Such a detached perspective has been criticised as alienating, impossible and ideological. As Young puts it:

[48] Ashford believes that utilitarianism so described implies – or at least makes room for – different conclusions to the 'classical utilitarianism' this thesis is about. As far as I can tell this depends on a non-aggregative, or partially aggregative, method of 'inclusive combination'. But such views are difficult to specify and sustain. We should, like Chappell, consider aggregation as at least one permissible way of doing such 'inclusive combination' of each person's well-being. In that case, we arrive at my utilitarian principle without neglecting persons in favour of some abstract overall well-being, but rather as a way of respecting persons, through valuing their well-being.

> 'the stances of detachment and dispassion that supposedly produce impartiality are attained only by abstracting from the particularities of situation, feeling, affiliation and point of view.' (1990, 97)

Drawing on Williams, Young sees such detachment as inappropriate for practical deliberation, which is inevitably first-personal – it is you, not some impartial spectator, nor every individual in inclusive combination, who must act (1990, 104–5). Because Young assumes that impartiality requires this kind of detachment, she concludes that impartiality tends to be myth, covering for what is, consciously or not, the partiality of dominant groups towards themselves (1990, 107–21).

But is there a way to motivate impartial attitudes without the kind of detachment from the self that Parfit describes and Young criticises? We find an attempt to do just this in the *Bodhicaryāvatāra* of the eighth-century Buddhist monk Śāntideva:

> VIII.90 At first, one should meditate intently on the equality of oneself and others as follows: 'All equally experience suffering and happiness. I should look after them as I do myself.'
>
> VIII.91 Just as the body, with its many parts from division into hands and other limbs, should be protected as a single entity, so too should this entire world which is divided, but undivided in its nature to suffer and be happy.
>
> VIII.92 Even though suffering in me does not cause distress in the bodies of others, I should nevertheless find their suffering intolerable because of the affection I have for myself.
>
> VIII.93 In the same way that, though I cannot experience another's suffering in myself, his suffering is hard for him to bear because of his affection for himself.
>
> VIII.94 I should dispel the suffering of others because it is suffering like my own suffering. I should help others too because of their nature as beings, which is like my own being.
>
> VIII.95 When happiness is liked by me and others equally, what is so special about me that I strive after happiness only for myself?
>
> VIII.96 When fear and suffering are disliked by me and others equally, what is so special about me that I protect myself and not the other?
>
> VIII.97 If I give them no protection because their suffering does not afflict me, why do I protect my body against future suffering when it does not afflict me?

VIII.98 The notion ‘it is the same me even then’ is a false construction, since it is one person who dies, quite another who is born.

VIII.99 If you think that it is for the person who has the pain to guard against it, a pain in the foot is not of the hand, so why is one protected by the other?

VIII.100 If you argue that, even though this conduct is inappropriate, it proceeds from the sense of self-identity, [our response is that] one should avoid what is inappropriate in respect of self and others as far as one can.

VIII.101 The continuum of consciousness, like a queue, and the continuation of constituents, like an army, are not real. The person who experiences suffering does not exist. To whom will that suffering belong?

VIII.102 Without exception, no sufferings belong to anyone. They must be warded off simply because they are suffering. Why is any limitation put on this?

VIII.103 If one asks why suffering should be prevented, no one disputes that! If it must be prevented, then all of it must be. If not, then this goes for oneself as for anyone else. (Śāntideva 1995, 96–97)[49]

Here, Śāntideva seems to be attempting to encourage a kind of impartial attitude similar to that of the utilitarian: it is morally irrelevant to whom suffering or happiness accrues; it is simply that suffering ought to be eliminated and happiness promoted. The notable thing about this passage is that Śāntideva does not propose some external perspective (a view from nowhere, from behind a veil of ignorance, from the point of view of the universe or from an impartial spectator) from which to adopt impartial attitudes. Instead, he tries to motivate them by proper reflection on the self: recognising that the self is similar to others, part of a social whole, or non-existent.

If Śāntideva is right, this is good grounds for rejecting the claim that impartial attitudes alienate one from oneself. To be clear, I am not endorsing this passage as an argument for impartial attitudes.[50] We already have such arguments in the arguments for utilitarianism,

[49] I use the Crosby and Skilton translation (Śāntideva 1995; compare Garfield, Jenkins, and Priest 2015, 59–60). Note also that although I attribute the passage to one author (Śāntideva), parts may have been added by later editors.

[50] For one thing, this might not have been Śāntideva’s intention. His passage might be read as a tool for meditation rather than a positive philosophical argument. For discussion about the exegesis of this passage, see Garfield, Jenkins, and Priest 2015. For a canonical commentary by Prajñākaramati, see the appendix to Cowherds 2015. For a broader survey of the connections between Śāntideva and utilitarianism see Goodman 2017.

and for other impartial moral theories. Rather, I take from Śāntideva that there are rationales, based not on impersonal detachment but on engagement with oneself, which can motivate agents to adopt utilitarianism's impartial attitudes. This gives us reason to think that such attitudes are, or at least can be, a form of proper engagement with the self, and therefore of disarming the alienation charge against them.

We can extract from this passage three such rationales. The first is that each individual is similar: we are all sentient beings, subject to the same kinds of suffering and happiness (90, 92-95, 103). The second is that each individual is part of a single whole (91, 99). The third is that there is no self at all, merely conventionally defined collections, so that there is no deep distinction between suffering that accrues to one person and to another (97-98, 100-102).

The third rationale is based on controversial metaphysics. If it avoids alienation from the self, it does so by eliminating selves from the world altogether, and if impartiality relied upon such metaphysics it would, as critics allege of utilitarianism, neglect the separateness of persons and regard all individuals as mere receptacles for value, not as individuals with their own lives to lead. (It is also somewhat in tension with the other two rationales.) That is not to say that such a view should be ruled out: Parfit's *Reasons and Persons* (1984, pt. III) suggests (though does not officially endorse) a position that has been called 'complete utilitarianism' (Broome 2004, 110) which has much in common with Śāntideva's view thus interpreted. But I will leave this view here. The other two rationales, parthood and similarity, have more to recommend them, and I explore these in the next two subsections.

To reiterate, the point of the arguments in this section is not to provide compelling arguments for some requirement to adopt impartial attitudes. It is to show that there are ways of thinking that can motivate and maintain impartiality that do not involve flight to a detached, impersonal point of view. Thus, to return to Williams, the utilitarian agent need not be characterised as 'stepping aside from his own projects' (1973, 116), when she considers them impartially but rather (as he recommends) 'trying to understand them' (1973, 118) in certain ways that I will now describe in more detail.

a. Parthood and community

Śāntideva uses the metaphor of the individual as a part of the body, and others as other parts of that same body. There is no alienation in a hand protecting a foot in the same body from a painful experience, he says, so why (we might continue) should there be alienation involved in my taking another individual's pain to be as important as my own?

It's not immediately clear what Śāntideva's line of thinking is, but here's one possibility. A hand is part of the body. A pain in the foot of that same body, though not a pain in the hand, is a pain in the body of which the hand is a part. From the perspective of that body, a pain is a pain, whether it occurs in the hand or the foot: the body is impartial between its hands and feet (things might be different if we consider vital organs). This impartiality is demonstrated by the fact that the body will risk a lesser pain in the hand to save the foot from a greater one.

Now, could individual persons be like this – each a part of a wider whole, from whose perspective their projects, well-being, etc. matter equally? As with the 'no self' justification, it may be objected that this line of reasoning depends on controversial metaphysics. Just as denying the existence of individuals is highly controversial, so is claiming that individuals are parts of a wider whole, as hands and feet are part of the body, such that their projects belong to some collective. Moreover, it seems to diminish the individual to reduce them to parts of a whole. It has been alleged that this is precisely the view utilitarianism takes. We may be reminded of the charge from Rawls that utilitarianism 'conflates all persons' (1971, 26–27), from Gauthier that utilitarians 'suppose that mankind is a super-person' (1963, 126), of Nagel's 'mass person' (1970, 134) or Nozick's assertion that, when some person's interests are traded off against another's:

> 'There is no social entity with a good that undergoes some sacrifice for its own good. There are only individual people, different individual people, with their own individual lives.' (2013, 32–33)

But there may be something more to be said for these parthood claims, given the social nature of persons. We are not Hobbes's mushrooms, 'emerged from the earth… and grown up without any obligation to each other' (1997, 102)[51]. Rather, as the Akan proverb has it, when a person is born, they alight in a town. This motivates what Kwame Gyekye calls 'communitarianism':

> 'Communitarianism immediately sees the human person as an inherently (intrinsically) communal being, embedded in a context of social relationships and interdependence, never as an isolated, atomic individual… The fact that a person is born into an existing community must suggest a conception of the person as a

[51] In fact, Hobbes's metaphor is misplaced: mushrooms often engage in mutualistic relationships!

> communitarian being by nature, even though some people insist on the individuality of the person.' (2010, 104–5)[52]

The communitarian nature of individuals could underwrite an analogy with the body. As a hand is part of the body, so I am part of the community. Thus, though a pain (or project, or need, or joy) of my neighbour may not be mine, it does belong to something of which I am a part, namely the community. From the perspective of the community, some pain/project/ etc. does not matter more than some other simply because of who has it. Insofar as I can take up my community's perspective – and the communitarian nature of persons suggests that I can – I can thereby motivate and maintain impartial attitudes.

Do communities have perspectives? And if they do, are these perspectives impartial between people? Nozick, as we have seen, would deny the former. And even if it were accepted, one might associate the good of a community with the good of its king, or its military, or its customs and traditions, whatever biases they involved. But remember here that I am not making an argument for impartiality. Rather, I am showing how a non-alienating line of thought might motivate impartial attitudes. Thus, I do not need to rule out alternative lines of thought that take us to different places. I just need to show how this one can take us to impartiality.

Here's one way in which it might, based on natural ways of interpreting a community's perspective. A community's good (we might think) is a function of the good of each of its members. This function obeys the constraint known as 'anonymity'. Anonymity can be expressed most easily with respect to a comparison of options. In option A Colin has a good of 3, Curtis a good of 2 and Kamil a good of 1. In option B Colin has a good of 2, Curtis a good of 1 and Kamil a good of 3. In option C Colin has a good of 1, Curtis a good of 3 and Kamil a good of 2. Since each option involves the same amount of good, in the same distributive pattern, the only difference being where individuals are placed in that pattern, anonymity holds that the good of the three-person community to which Colin, Curtis and Kamil belong is equivalent in each option. In general: distributions of goods across members that are

[52] Gyekye notes communitarianism's ties with the African socialism of Nyerere and Nkrumah (Gyekye 2010, 104). For a European social democrat (as he was then) tying our communitarian nature to impartiality (and similarity), consider this definition of socialism from Tony Blair: 'It is a moral purpose to life, a set of values, a belief in society, in co-operation, in achieving together what we cannot achieve alone. It is how I try to live my life, how you try to live yours - the simple truths - I am worth no more than anyone else, I am my brother's keeper, I will not walk by on the other side. We are not simply people set in isolation from one another, face to face with eternity, but members of the same family, same community, same human race.' (Blair 1995)

permutations of one another – that is, that can be reached from one another simply by shuffling names – are equivalent, from the perspective of the community. Thus, if a community's good is given by a function of this type, the community's perspective is impartial between people.

Although utilitarianism is one such function it is far from the only one. The idea of a 'social welfare function', from individual to communal good is a leading way of considering the latter, and anonymity is one of the least controversial constraints on it (Adler 2019, 97). In any case, I am not arguing that a community's good is given by this kind of function. I am merely arguing that one may think along these lines, and in so doing reach impartial attitudes without alienation from oneself, by taking the perspective of a whole of which one is a part.

Now, unlike the hand, I can occupy other perspectives too. It might well be alienating to regard oneself (and others) all of the time from the impartial perspective of the community. This is because, as Gyekye recognises:

> 'besides being a communitarian being by nature, the human person is, also by nature, other things as well. By "other things", I have in mind such essential attributes of the person as rationality, having a capacity for virtue and for evaluating and making moral judgments and, hence, being capable of choice. It is not the community that creates these attributes; it discovers and nurtures them. So that if these attributes play any seminal roles in the execution of the individual person's lifestyle and projects, as indeed they do, then it cannot be persuasively argued that personhood is fully defined by the communal structure or social relationships.' (Gyekye 2010, 112)

To value our own projects simply in terms of the value they contribute to the community is, therefore, an impoverished form of engagement with the self. (Valuing others in the same manner may be an impoverished way of engaging with them, too.) Rather than Śāntideva's picture of humanity as one body incorporating (like the frontispiece of Hobbes's *Leviathan*), each individual as a part, the Akan motif of a crocodile with one stomach but two heads may be a more apt illustration of our situation. We do share a common good, but we are individuals too. My projects belong to the community but also to me, and valuing myself does not simply mean valuing my community. This means that to regard ourselves merely as communitarian beings would be a form of alienation.

But, as stated in the previous section and in Chapter 2, utilitarianism does not ask us to adopt impartial attitudes all the time, and will likely recommend partiality towards our own projects in many instances. This division between partial and impartial attitudes is a reflection on the fact that we are both communal and individual beings. It is not clear that impartiality motivated by such a consideration would be alienating – or at least, if it is, it is not clear why it would be more alienating than *not* adopting impartial attitudes, and being partial to one's own projects, which reflects only the individual but not the communal side of oneself. Maybe the best way of engaging with our natures, as both social and individual creatures, is to employ, at different times, impartial and partial attitudes. This is exactly what utilitarianism asks of us.

b. Similarity

Śāntideva's other motivation for impartiality is our similarity with others. He seems to invoke the following pattern of reasoning:

1. I have reason to relieve my own suffering, and promote my own happiness.
2. I am like other individuals.

Therefore,

3. I have equal reason to relieve other people's suffering, and promote their happiness.

We could substitute other reason-giving properties for happiness and suffering, such as projects, well-being and so on. Like all arguments from analogy, this argument is of course defeasible. The fact that I am like other individuals does not prove that others have all the same properties that I have.[53] However, arguments from analogy can still be informative – they can have inductive rather than deductive validity. Moreover, remember that we are not looking for arguments justifying impartiality here, but rather rationales that can motivate impartial attitudes. It is enough that someone can be rationally led to believe the conclusion, without alienation, even if it is not proven.

Crucially for my purposes, this argument is grounded not in detachment from the self, but engagement with it. The premises are about oneself. The first is an affirmation of one's own value. The second is a comparison of oneself with others which, I will argue below, can enhance our self-understanding. There is little to suggest that someone who reasons in this way is alienated from themselves.

[53] In particular, such an argument is no evidence for others sharing my indexical properties – that being me, here, associated with this body, this personal history and so on. It can be evidence for others having their own corresponding indexical properties – having a body, a self, a place, a personal history – but not mine. My assumption is that one's indexical properties are no more reason-giving than anyone else's.

Consider an actual instance of this kind of reasoning: a speech of the socialist politician Tony Benn, speaking against the bombing of Iraq:

> ‘War is an easy thing to talk about; there are not many people [in the current Parliament] of the generation that remember it… But I was in London in the blitz in 1940, living in the Millbank Tower, where I was born… And every night, I went down to the shelter in Thames House. Every morning, I saw dockland burning. Five hundred people were killed in Westminster one night by a land mine. It was terrifying. Aren't Arabs terrified? Aren't Iraqis terrified? Don't Arab and Iraqi women weep when their children die? Does bombing strengthen their determination? What fools we are to live in a generation for which war is a computer game for our children and just an interesting little Channel Four news item.’ (Benn 1998)

Benn starts with some properties of his own: *he* found being bombed terrifying, *he* saw people being killed. This grounds the disvalue of that bombing, the reasons there were not to do it. Then the argument turns to others: since Iraqis are like him, we can take it as evidence that they will be terrified by bombing too, that they will see death. This, therefore, is an equally strong reason not to bomb them.

What is striking about Benn’s speech is that he begins with a portrayal of events in his life, and how he felt, and leverages them – through an implied premise that he and Iraqis are alike – to motivate sympathy with the situation of others. Nobody could claim, I think, that Benn misunderstood, or was not properly engaged with, his youthful memories of war when making this speech. (In his last sentence, he indicates that what such alienation might look like: a reduction of wartime experience to a game or news item. This might be possible for his children – but for him, having experienced war, it would be alienating.) He is not adopting a view from nowhere nor rejecting the self nor subsuming himself to a social whole. It is his very self-understanding, in conjunction with the premise of similarity with others, that leads him to impartiality.

Moreover, this premise can itself enhance self-understanding. There are some aspects of individuals that we can only recognise in others, or at least, that we see in others more clearly than we do in ourselves. Thus, reflecting on our similarity to others can help us understand ourselves better. Moreover, we can draw a further conclusion from this reasoning: not only am I y, but being y is a property I share with others. Understanding this fact may be in important part of understanding oneself.

A parable from the Buddhist tradition makes this point. A young mother is beset with grief after her son passes away. In fact, she cannot accept that he is dead. Looking for a healer, she turns to the Buddha. The Buddha tells her he will need a handful of mustard seeds – and they must be collected from a household in which nobody has ever been bereaved. The mother goes from house to house to find mustard seeds that will meet this description, but she cannot. In every house, there is someone who has lost someone close to them, and in every house they tell her of their grief. She realises how like them she is. She gains neither mustard seeds nor a resurrection of her son but the knowledge that life is fragile, bereavement inevitable, and grief both appropriate and universally shared.

Now, the woman in the story is not alienated from herself by her recognition of similarity with others. Rather, she better understands what her own position is: one that is common, shared, inevitable, though also fitting to be upset about, and sympathised with by others.

Thus, by reflecting on our similarity with others we can motivate impartial attitudes; far from detaching us from ourselves in favour of a 'view from nowhere', this is an engagement with and method for better understanding who we are. Of course, it would be odd, and perhaps alienating, to consider oneself maximally similar with others, with no uniqueness – just as it would be to consider oneself a mere part in the social whole. But utilitarianism, as we've seen, doesn't demand that one always treat oneself impartially, just that one sometimes does. And when one does, I suggest, this can be achieved by reflection on one's own value-conferring properties and one's similarity with others. If one never engaged in such reflection, one would undermine the self in a different way, missing out on the chance to further one's understanding of oneself.

4. Does impartiality require lower self-valuation?

One thing aspect of impartiality that we might worry about is that valuing others as much as one does oneself necessitates valuing oneself less than our usual, partial attitudes lead us to – and this low self-valuation might be a kind of alienation. This would be so if valuing people is a zero-sum game: the more we value others, the less we value ourselves. In this section I want to push back against this thought.

The zero-sum thought is not conceptually necessary. As I stated above, to value something more is to take oneself to have stronger reasons for certain attitudes and actions towards it. The partiality that most of us have (and that utilitarianism is likely to permit some but not all of the time) means that we take ourselves to have stronger reasons to pursue our own projects and promote our own well-being than we have to assist others in their projects, promote their well-being. Now we could reach impartiality by 'levelling down' our judgment of the strength of these reasons: taking ourselves to have reasons regarding ourselves that are as weak as the reasons we tend to take ourselves to have regarding others. But we could also 'level up' to impartiality (Chappell 2022): taking ourselves to have reasons regarding others that were as strong as the reasons we tend to take ourselves to have regarding ourselves, and leaving the latter untouched. We could even reach impartiality at a level where we take both reasons regarding ourselves and those regarding others more strongly than we currently tend to do.

In fact, there are good reasons to think that as we value others more, our self-valuation will grow rather than shrink, in at least some ways. Firstly, if one brings oneself to adopt impartial attitudes through either of the rationales described above, one takes other people to be importantly connected with oneself. Therefore, one might take reasons to do certain things with respect to them as also reasons to do certain things with respect to oneself. As one takes those reasons to be stronger, one is valuing both others and oneself more. This accords with Railton's observation:

> 'When one studies relationships of deep commitment – of parent to child, or wife to husband – at close range, it becomes artificial to impose a dichotomy between what is done for the self and what is done for the other. We cannot decompose such relationships into a vector of self-concern and a vector of other-concern, even though concern for the self and the other are both present. The other has come to figure in the self in a fundamental way or, perhaps a better way of putting it, the other has become a reference point of the self... and that which affects these reference points may affect the self in an unmediated way. These reference points do not all fall within the circle of intimate relationships, either.' (1984, 166–67)

More generally, it may be a mode of valuing someone that one takes oneself to have reasons to value others in their community. And it may be a mode of valuing someone that one takes oneself to have reasons to value others who are similar to them. Once one starts to think of oneself as part of a community, or as similar to others, seeing reasons to do things for others, then, may be partly constitutive of valuing oneself. As these reasons grow in strength, so does one's own self-valuation.

Nor does this point simply turn on the degree to which we deny our individuality by emphasising our fellow parthood with and similarity to others, as might be feared. Note that a large part of what makes us the unique, active individuals we are – our pursuit of projects – is also bound up with others. This was the lesson of the fact of mutual interdependence that I stressed in Chapter Four. As Marx summed it up:

> 'Not only is the material of my activity given to me as a social product (as is even the language in which the thinker is active): my own existence is social activity, and therefore that which I make of myself, I make of myself for society and with the consciousness of myself as a social being.' (Marx 1988, 105)

How does this help correlate self-valuation with the valuation of others? Part of what it is to value oneself is to take oneself to have reasons to pursue one's projects. Now those projects depend on others, as we saw in the previous chapter. And the strength of at least some of the reasons we have to pursue them depends on the strength of our reasons for certain actions and attitudes towards others. The more important we think it is to affect others, the more important it is to pursue projects that will affect them. The more we respect the work of others, the more we will respect our own projects, which have in reality been co-created (both materially and conceptually) with others. Thus, the more we value others (in certain ways), the more we value ourselves (in certain ways). Recall, then, the passages from Marx and Bakunin quoted in the previous chapter:

> 'Supposing that we had produced in a human manner; each of us would in his production have doubly affirmed himself and his fellow men… In your enjoyment or use of my product I would have had the direct enjoyment of realising that I had both satisfied a human need by my work and also objectified the human essence and therefore fashioned for another human being the object that met his need… In my expression of my life I would have fashioned your expression of your life, and thus in my own activity have realised my own essence, my human, my communal essence.
>
> In that case our products would be like so many mirrors, out of which our essence shone.' (Marx 2000b, 132)
>
> 'The liberty of every individual is only the reflection of his own humanity, or his human right through the conscience of all free men, his brothers and equals…
>
> I am not myself free or human until or unless I recognise the freedom and humanity of all my fellowmen…
>
> Only by respecting their human character do I respect my own. A cannibal who devours his prisoner… is not a man but a beast. A slave owner is not a man but a master. By denying the humanity of his slaves he abrogates his own humanity…' (Bakunin 1973, 137)

Their common theme of reflection suggests that these socialists believe in a positive correlation between our valuation of others and valuation of the self.

Now, this positive correlation only goes so far. There will inevitably be some reasons to do things with respect to ourselves that weaken as we increase the strength we find in reasons to do things for others. Promoting our own interests, for instance (at least with regards to purely selfish interests like consumption). But another lesson we might learn from foregoing is that recognising these reasons is not all there is to valuing oneself. One can value oneself in a different – perhaps we should say more 'solidaristic' – way, by recognising the value of others, who are similar to you, part of your community, and on whom you depend for the meaning and necessities of your life. This will provide stronger reasons to pursue certain projects of your own, even as other more selfish reasons wane. Coming to see things this way need not be an alienation from the self even as it is a reduction in alienation from others.

It might be objected that there is a 'budget constraint' on how much valuing we can do. The reason that the value of self and other is thought to negatively correlate is because there's only so much valuing we can do, and the more of it you give away the less you can keep for yourself. I'm not sure that this is true. Firstly, why couldn't one just come to see more and stronger reasons in the world? We can imagine a nihilist, who takes nothing to give him a reason. Now he is persuaded that there is a reason for him to do something. Then another one. Then another one – then he is persuaded that one of the reasons he already recognises is in fact stronger than he previously thought. At each stage, it seems, the (former) nihilist has increased the degree to which the world includes reasons for him: he never had to give up one reason to gain another. Now, perhaps the partial person is in a similar position, though further along the spectrum than the nihilist. They see more reason to promote/pursue their own well-being/projects and so on than that of others. Why can't they, like the nihilist, come to be persuaded of new and stronger reasons involving others without their previous judgements about other reasons being affected?

There is, of course, one kind of inevitable trade-off in this neighbourhood: the trade-off with respect to resources. We have a limited amount of time, money, energy, and attention, and the more we spend it on others the less we can spend on ourselves – no matter how much we value both. This is an actual budget constraint. It is mitigated somewhat by the thought captured in the passage from Railton above: many of the things we do for others are to a great degree done for ourselves too. But this cannot eliminate the problem. It is often the case that of two options, one does more for oneself and one does more for others – even within the intimate relationships which best make Railton's case, and even if, as he notes, it is hard to disentangle the vectors and both options necessarily do something good for each party. In these cases, it is true that if we adopt impartial attitudes, we will likely choose options that do less for ourselves and more for others, compared with the partial agent. This would be so even if one valued oneself more than the partial agent valued themselves.

One might worry that the necessity of such trade-offs makes talk of the absolute strength of reasons idle. What does my claim that one could value both oneself and others more amount to, if one's choices when the interests of oneself and others conflict are identical with what they would be if one 'levelled down' to impartiality? For an answer consider what Chappell has to say about choices that benefit different individuals, but to the same degree:

> '[I]t's morally perverse for an agent to be *indifferent* between options that equally benefit distinct people, for that is to disrespect the individuals by treating them as fungible means to the aggregate welfare. But of course we do not want to favour either person over the other, since such bias would constitute disrespect for the person whose equal benefit we counted for less. Instead, I propose, the fitting response to a trade-off between two distinct but equally weighty values is to feel *ambivalent* about the choice. There are distinct reasons pulling you in either direction, corresponding to the distinct values served by either choice. But these reasons are equally weighty, so the agent is *torn* rather than pulled without resistance towards one choice over the other.' (2015, 326–27, his italics)

The ambivalent agent and the indifferent agent both judge the relative reasons for one option rather than the other to be balanced. But they are importantly different: the ambivalent agent values each of the individual patients more. Thus their experience of the choice is quite different. Presumably, Chappell's distinction can be extended to cases in which reasons are not balanced. The agent who is ambivalent in the balanced case is never 'pulled without resistance' to an option that benefits some more than one that benefits others even if the total benefits are greater. They choose it, but with some resistance, some tearing, some feeling of loss or regret.

Return to our choices between options that benefit oneself more and options that benefit others more. Which option one chooses will depend on one's judgment of relative reasons. The utilitarian impartial agent will choose the latter insofar as the benefit to others will be greater. But if their judgment of their own value is high – if they have levelled up rather than levelled down to impartiality – this choice will be experienced as one with costs, with accompanying resistance and regret. This may make the choice more painful than to one who has reduced their self-valuation. But the pain would indicate that one is not alienated from oneself.

Lastly, note that even if there is a constraint on how much valuing we can do overall, it would still be possible to increase both one's valuation of oneself and of others. This would be possible if we were to reduce our valuation of other things. Many value not only people, but institutions, traditions, nations, gods, laws, rights, money and so on. It is a striking similarity between socialists and utilitarians that they tend to place little value on these impersonal concerns. For socialists, particularly Marxists, they are superstructural distractions from where the real action is – the material lives of actual people. For utilitarians, their value is incorporated in the value of people: we have reasons with respect to such things only insofar as they affect people's well-being. We may reach impartial attitudes without undervaluing the self, then, by generally lifting up human concerns over impersonal ones.

5. Are there limits to impartiality?

I have suggested that the impartial attitudes of utilitarianism can be motivated by, and reflect, our nature as parts of a community, similar to others. We might worry that this justifies not impartiality but partiality to (a) our own community, (b) people who are like us.

I respond to these worries in turn. Begin with (a). In the modern, globalised world we stand in social relations with almost every other human alive. We participate in many of the same economic and political structures. Our individual interests are to a large extent connected, and we all stand to gain or lose depending on which collective actions we take as a human community in response to problems such as climate change. It is our parthood in this global social whole that can motivate impartiality.

But aren't we in stronger social relations, or more social wholes, with some people than others? And so, wouldn't this line of thinking lead to partiality based on degree of such connectedness? This objection mistakes the argument above. The view isn't that the degree to which we value people ought to be based upon our sharing parthood in a whole. It is rather that reflection upon our parthood in the global social whole can motivate impartial attitudes in a non-alienating manner. Similar kinds of reflection might motivate, for instance, partiality to nation, tribe or family. But this is no problem for my view – in fact, it might make it more plausible. One might find it more plausible that I should be indifferent between my projects and those of my family members (or fellow tribespeople, or compatriots) than that I should be impartial between everyone's projects. I have suggested the same kind of thinking that motivates the former can motivate the latter – insofar as talk of 'the global village', 'the human race', or 'the party of humankind' (Hume 1965, 115) is appropriate. Note that an important part of the socialist project has been extending the moral bounds of family and nation beyond their traditional places: thus, 'the brotherhood of man' becomes both a motivating identity and an aspiration.

Now consider (b). It is of course true that we are more similar to some people than to others. In the pattern of reasoning I outlined above, then, our evidence for their having equivalent value-conferring properties to ourselves is relatively stronger – as is our capacity to learn about ourselves from observing them. However, this need not issue in partial attitudes towards them over those with whom we share less. The pattern of reasoning issues the same conclusion (albeit with different levels of certainty) insofar as we count others as similar to us at all.

Both parthood and similarity rationales can lead to fully impartial attitudes, then, even amongst being to whom our social relations and similarity may vary. Both rationales, though, may have limits. There may be beings to whom we have no social relations, or such attenuated ones that we cannot consider ourselves as part of a whole that includes them. There may be beings so different to us that we cannot treat our value as any kind of evidence for theirs. Treating ourselves and our projects impartially with these beings could not result from these rationales, and so may well demand detachment from the self (a view from nowhere) and thus be alienating.

Which beings might this include? Much contemporary utilitarian discussion centres on non-human animals, and future generations (who may themselves be 'transhuman' or artificial intelligences). We do of course bear some similarities with such beings: we feel pleasure and pain, for instance (which Śāntideva seemed to think was sufficient to motivate impartiality through similarity). But it may neglect what makes us 'us' too much – to an alienating extent – to judge our own value according to such things, and thus to secure an impartial valuation of those beings through the similarity rationale.

We also could conceive of ourselves as standing in some part/whole relationships with them. I have focused thus far on the socio-economic notion of 'community' as the most important whole to which we belong, but one might regard oneself as a part of God's creation (as are all animals), a part of human history (as are all human and transhuman generations), and so on. According to the parthood rationale I described, if one did regard oneself in such a way, it might be possible to treat oneself impartially with respect to these beings without alienation. But whether one can regard oneself in such a way, with enough force, is a different matter. If one cannot, and utilitarianism demands such impartiality, utilitarianism is indeed likely to be alienating.

The question, in such circumstances, is whether we adjust morality to fit our self-conceptions, or vice versa. In some cases the latter seems justified. Imagine some chauvinist who could not see the similarities or shared community they had with people of other ethnicities. Utilitarian impartiality would be alienating for them; the rationales of parthood and similarity would be unavailable. But this would be no reason to dismiss utilitarianism. Indeed, insofar we find utilitarianism plausible, it would be reason to encourage this person to adjust their self-conception in a more cosmopolitan direction (along with other reasons for them to do so, which utilitarianism and common-sense moralities will likely endorse). In other cases we find the maxim that 'morality is made for man, not man for morality'

(Frankena 1973) more appealing. When Bostrom talks of the importance of bringing about 'value-structures, such as sentient beings living worthwhile lives' (2003, 308), I must admit that considering the value-structures that are *not* sentient beings living worthwhile lives (I think Bostrom has in mind here some kind of artificial intelligence), thus not like me, leaves me utterly cold. This is so even if we stipulate that these structures enjoy a kind of well-being. I find also a waning sympathy for those sentient beings living millions of years, or millions of light-years away, and those of little or no social relation to me, especially if they are supposed to be different species, or artificial, or have a well-being that consists in something quite different to mine. I find 'longtermist' arguments (e.g. Greaves and MacAskill 2021) more plausible when I consider that future generations will either have lives, hopes and dreams, pains and projects quite like mine, and that they will be part of a collective human story to which I also belong (see also Scheffler 2018).

If utilitarianism requires impartial attitudes across strange and distant beings, are we to reject it or demand, as we did with the chauvinist, that it our self-conceptions need to change such that we can appreciate our similarity and fellow-parthood with them? The latter may be too much to swallow. But there may also be a third way. We could accept utilitarianism (or at least not reject it on these grounds) whilst holding that it applies only within a limited realm. Within that realm, there would be impartiality, but for beings beyond it, some other theory may be preferred. Whether that other theory be called a moral theory, or whether morality itself only deals with those with whom we share a sufficient amount, is a further question. These are big questions, which run far beyond Williams's Integrity Objection (though he was undoubtedly interested in them (1985b)). I will leave them here.

6. Summary of my response to Williams

I can now summarise this thesis so far. Williams's Integrity Objection holds that utilitarianism is to be rejected because it is incompatible with commitment. This incompatibility is the result of utilitarianism's insistence that we regard our projects as dispensable, and impartially. One response to Williams is that utilitarianism does not require such attitudes, because it is self-effacing. I rejected the claim that utilitarianism is self-effacing for individuals – though it may permit us to use other decision-procedures alongside it, and it may be for groups, and this may mitigate some of the problems associated with the Integrity Objection. Next, I showed that utilitarianism's requirement to regard projects as dispensable is shared with a number of weaker and more appealing moral principles, given our mutual interdependence. I also showed how commitment may alienate us from others, and why avoiding such alienation is important. Thus, making room for commitment, as Williams defines it, is not a good adequacy condition for moral theories. In this chapter, I argued that utilitarianism's requirement of impartiality need not alienate us from ourselves. It need not preclude us pursuing our own projects, and even showing partiality sometimes; it need not require a flight to a detached, impersonal point of view but rather a recognition of our communal nature and similarity with others. It also need not imply a reduction in our valuation of ourselves: in fact, drawing on socialist insights and once again on mutual interdependence, I argued that the value one places on others and on oneself may be positively correlated to some degree. These arguments may have limits – that is, they support impartiality amongst a group that is less than universal. We might take this to be a welcome or unwelcome implication.

Chapter Six: Pre-Emption and Collectivist Utilitarianism

0. Abstract

Though it was not the primary point of the Integrity Objection, Williams's cases of George and Jim raise a further problem for utilitarianism: the problem of pre-emption. George and Jim could be the proximate cause of a non-optimific outcome; however, if they do not play this role, someone else will, with at least as bad consequences for well-being. In such cases, utilitarianism will provide reasons for agents to cause non-optimific outcomes themselves, even when, if everybody refrained from such actions, the outcome would be more conducive to well-being. I argue that pre-emption is widespread and important, and includes the case of capitalist exploitation with which socialists concern themselves.

Such cases bring into tension two guiding lights of utilitarianism: that well-being matters, and that an action's moral status is determined by the difference it makes to the world. Existing attempts to revise utilitarianism fail to do justice to both of these in pre-emption cases. I present my favoured solution, collectivist utilitarianism, which is the application of utilitarianism to groups as well as to individuals. This view captures the two guiding lights, and has the resources to explain other intuitions we have about such cases. It requires a view about group agency which I explain and defend, drawing on Marx's work on class. I then distinguish collectivist utilitarianism from a similar view (the 'joint ought' view) and show its distinctive implications for practices such as blame, praise, guilt and punishment.

I believe that collectivist utilitarianism, by acknowledging the importance of our membership in group agents, invites us to look at moral agency in a more appealing way than utilitarians, focusing on individual agency alone, typically have.

1. Introduction

Williams's cases of George and Jim are primarily meant to show how utilitarianism is incompatible with commitment. But they also raise another problem for utilitarianism (one which is also discussed by Williams (1973, 131–34), though with less depth and far less influence than his discussion of integrity). This is the problem of pre-emption. In this chapter I will show why pre-emption is important, why it poses a problem for standard versions of utilitarianism and other theories, and then suggest a version of utilitarianism – 'collectivist utilitarianism' – that can handle it well.

To grasp the shape of pre-emption cases, recall Williams's case of George (1973, 97–98):

> ‘George, who has just taken his PhD in chemistry, finds it extremely difficult to get a job... An older chemist, who knows about this situation, says that he can get George a decently paid job in a certain laboratory, which pursues research into chemical and biological warfare. George says that he cannot accept this, since he is opposed to chemical and biological warfare. The older man replies that he is not too keen on it himself, come to that, but after all George's refusal is not going to make the job or the laboratory go away; what is more, he happens to know that if George refuses the job, it will certainly go to a contemporary of George’s who is not inhibited by any such scruples and is likely if appointed to push along the research with greater zeal than George would. Indeed, it is not merely concern for George and his family, but (to speak frankly and in confidence) some alarm about this other man's excess of zeal, which has led the older man to offer to use his influence to get George the job... What should he do?’

It appears that utilitarianism, given simply the considerations in this passage, requires George to take the job. There will be less overall well-being if he refuses it than if he accepts. As we have seen, utilitarianism’s treatment of such cases makes commitment incompatible with a utilitarian mindset. I suggested that we reject commitment. But there is another aspect of this case that ought to worry utilitarians. Research into chemical and biological warfare (CBW) is bad with respect to overall well-being.[54] Utilitarianism, therefore, ought to have something to say for opposing it. But how can it, if it requires people in George’s position to take such jobs?

This chapter investigates such cases and having evaluated other possible responses proposes as a solution ‘collectivist utilitarianism’, or the application of the utilitarian principle to group agents. Group agency is an important part of the socialist tradition. Socialists have often emphasised that production is paradigmatically a collective action, and this forms the basis for socialist claims about the permissibility of redistribution and the practicability of economic planning. In the Marxian theory of history, it is classes, not individuals, who are the agents of change. The importance of group agency to socialists is reflected in songs, slogans and creeds from across the tradition: the Industrial Workers of the World sang ‘it was we who built the railroads and the cities where they trade’; the UK Labour Party has its members carry cards that claim ‘by the strength of our common endeavour we achieve more than we achieve alone’; the German social democrats entered a recent election with the slogan ‘das Wir entscheidet’ [‘the we decides’]; Bernie Sanders

[54] I will simply assume this. If you doubt it, substitute CBW for some other occupation whose consequences you believe to be bad.

fought a presidential campaign on the basis of 'not me, us'. According to the argument of this chapter, this is no accident. As I will show in the next section, a crucial concern of socialists – the exploitation of workers by capitalists – is a case of pre-emption, structurally similar to that of George. I will subsequently argue that an attractive utilitarian interpretation of this phenomenon (which itself draws on Marx's work on class) emphasises the importance of collective action.

2. Pre-emption cases

George's situation is similar to the career choices faced by many people. Should I go into this immoral line of work? Well, if I don't, somebody else will, and they will probably be less scrupulous then me. So maybe I should. Similar remarks apply to consumer choices. If I don't buy this chicken, one might think, someone else will. So my purchase has no effect on the number of chickens raised and slaughtered – and maybe I'll enjoy it more, or use it to feed more people. Now, this kind of thinking assumes that the supply of jobs and chickens is not responsive to additional demand, which isn't always true.[55] Your willingness to take a job or buy a chicken might have some effect on supply, rather than just replacing someone else's action – so your actions are not completely pre-emptive. But those effects will often be very small,[56] so that the harm your action does might still be outweighed by the fact that if you don't do it, somebody else will, and worse.

Once one takes a job and considers one's actions within it, pre-emption is often more complete. Your boss orders you to do something bad. You can say no, but the work will just be handed to someone else, and you may be fired. So what's the harm, you might wonder, in following orders? Nor is this simply a problem for subordinates. Consider the story of Chuck Prince, former CEO of Citigroup, as told by John Cassidy (2010). Citigroup had refrained from investing in the innovative financial instruments backed by US mortgage debt which boomed in the early 2000s, with a result that its profits and share prices lagged behind its rivals. Prince's position was under increasing pressure. So from 2005 he 'authorised a rapid expansion of Citi's securitization businesses, especially those dealing with subprime mortgages' (Cassidy 2010, 296). As Cassidy points out, Prince had compelling reasons to do this even if he had reservations about the risk or ethics of such business. If he had not taken this course, either Citigroup would have continued losing market-share to rivals who did, or

[55] It also assumes that there are no alternative better uses of your money or labour power. The value of these can also be balanced by other people's use of the product or performance in the job, being worse.

[56] For recent discussions of such effects and their moral significance, see Budolfson (2019) and Wilkinson (2022).

he would have been replaced by a CEO who would. Prince was merely pre-empting the actions of others. (And presumably he could tell himself that it was better that he kept his job and his bonuses than some other banker took them: if, for example, he was donating more of his income to good causes, or he was more pleasant to work for.) So Prince had some utilitarian justification, stemming from pre-emption, for his approach, even though it ended up being a large part of the cause of the financial crisis of 2007-8. As he himself put it: 'as long as the music is playing, you got to get up and dance.' (*Financial Times* 2007) Pre-emption is likely to be very common in market economies, because, as Pinkert observes agents typically 'know that were they to stop participating, others would step in and compensate by taking over their transactions' (2015, 978), as long as those transactions are profitable.

The general structure of pre-emption cases is this. There is some individual action available to me that would be the proximate cause of a bad outcome. However, were I not to take such action, some other individual would probably take a similar action which would be even less conducive to well-being. The likelihood with which they would take such an action, and the worseness of the outcomes they would produce are such that the expected overall well-being of my taking the action is greater than or equal to the expected overall well-being of my not doing so. This is so even though it would be optimific if nobody took the action.[57]

Why are these cases problematic for utilitarianism? Utilitarians should want their theory to encourage the most well-being to be realised. One guiding light of utilitarianism is that the amount of overall well-being is of fundamental moral importance. At the same time, a second guiding light of utilitarianism is that the difference an action makes to outcomes is crucially important to its moral status.[58] And indeed, this is part of utilitarianism's appeal: 'what difference would it make?' is acknowledged by many people to often be a very good question to ask when deliberating about what one should do. In pre-emption, the actions that would collectively bring about the most well-being seem to be forbidden on the grounds that they, individually, would make no or negative difference. Pre-emption cases, then, put two of utilitarianism's most characteristic and appealing features into conflict with one another.

[57] Compare Williams's general description of his cases: 'The situations have in common that if the agent does not do a certain disagreeable thing, someone else will, and in Jim's situation at least the result, the state of affairs after the other man has acted, if he does, will be worse than after Jim has acted, if Jim does. The same, on a smaller scale, is true of George's case.' (1973, 108)

[58] It is important to recall here that in this thesis by 'utilitarianism' I mean something akin to act-utilitarianism and not rule-utilitarianism (which is discussed in the next section).

It may be helpful at this point to distinguish pre-emption cases from other cases in which what my act makes no positive difference to the outcome. They are often discussed alongside – or conflated with – cases of overdetermination (e.g. Tadros 2021). In overdetermination cases, the same outcome will occur if I do not act, because of other people's acts. However, in these cases others will act whatever I do – by contrast, in pre-emption cases, my not acting results in their acting in my place. In overdetermination cases, if I do not act and all else is held equal, the same outcome occurs; in pre-emptive cases, if I do not act all else will not be held equal, because someone else will act differently to bring about the same outcome. More controversially, there may be cases of what has been called 'causal impotence', where some set of actions causes an outcome, but at least some individual actions in that set make such a small contribution that the exact same outcome would have occurred without them. In these cases it is not that the actions of others are sufficient for the outcome, as with overdetermination (though they may be) nor that others will respond to my not acting by acting to bring about the at least as bad an outcome, as in pre-emption cases. Rather, whatever others do, my action is too small to make a morally relevant difference – just as a drop in the ocean makes no difference to its currents. Causal impotence is controversial because some deny it is logically possible, though it is intuitively appealing in some cases (see Nefsky 2019 for discussion). Overdetermination and causal impotence cases pose some similar problems for the utilitarian, and I think my collectivist utilitarianism will help to account for them too. But they are different from pre-emption cases in various ways; generally, I think that pre-emptive cases are both more widespread and hardest for utilitarians to deal with. It may also be that many of the cases that have been discussed under the heading of overdetermination or causal impotence are in fact pre-emption cases (especially if causal impotence cases are logically impossible), so my analysis should be useful for dealing with these.

I have shown how pre-emption cases arise in several mundane contexts: career choice, consumption decisions, workplace obligations. This is enough to make them of interest to moral theory: we want our moral theories to be able to address everyday moral questions. But pre-emption is also important because it crops up in some decision-contexts of the greatest social significance.

One is the exploitation of the global working class. Take Iris Marion Young's analysis of the global clothing industry (2006) – which I take it is representative of many supply chains under capitalism. (We already met this case in Chapter Four.) Many of the workers who manufacture clothes face serious harms, including overwork, precarity, low pay, unsafe conditions and restrictions on their rights to organise. The proximate causes of most of these harms are the actions of those who employ or manage them – usually small enterprises in poor countries, dependent on larger exporters, who are in turn dependent on large multinationals who sell to consumers in rich countries. However, as Young says:

> 'In this system, each of the links in the chain believes itself to be operating close to the margin in a highly competitive environment, and usually is under heavy pressure to meet orders at low cost by firms higher up the chain.' (2006, 110)

This is a case of pre-emption. The sweatshop employer takes actions that are the proximate cause of ill-being for their employees. But if they were not to do so, another employer would step in to perform similar actions. Given the competitive nature of the market, employers can be almost certain that such a rival employer exists, and given the pressure of the multinationals, it is unlikely they could afford to improve working conditions very much. Indeed, the original employer might reasonably think that their business being outcompeted by rivals would be worse for their workers, given the costs of being thrown from one employer to another, and that if a rival outcompetes them it is probably because they have lower costs, and therefore probably sweat their workers even harder. Similar remarks go for the multinational buyer: if they paid manufacturers more, allowing them to improve working conditions, they too would be outcompeted.[59] It seems that even though it would be optimific for employment to be less exploitative, each individual exploiter brings about better consequences by being involved in exploitation than by not being.

Capitalist exploitation of workers is, of course, a central concern of socialists. That it has the structure of pre-emption is reason enough for socialists to be interested in the ethics of such cases. It also explains several claims socialists – in particular, Marxist socialists – tend to make. The first is that moral appeals to capitalists to treat their workers better are ineffectual and misguided. Under Young's description of exploitation as pre-emption, a capitalist who did treat their workers better risks being replaced by a less scrupulous employer. Thus: (1) even if a capitalist does care about the welfare of their workers, they may reason that their improving conditions would not improve this welfare; (2) if they did

[59] What of the consumers to whom they ultimately sell, and whose demand for cheap clothing puts downward pressure on costs throughout the chain? Well, our buying fewer clothes may make things even worse for the workers by putting them out of work, or driving prices down further. Signalling that we would be willing to pay more for their clothes may simply increase the profits of multinationals, and cost us money we could better use elsewhere. This is not exactly a case of pre-emption, for consumers, but it has similar results: participating in an exploitative industry turns out to have better consequences than not doing so, given the behaviour of others. Thus the effective altruist William MacAskill, who generally commits to the utilitarianism or something like it, recommends buying clothes made under such conditions (2015, 158–63).

become convinced of a moral duty to improve their workers' conditions, their compliance with that duty might not in fact help workers in the long run. The pre-emptive structure of exploitation also explains the Marxian claim that capitalists, too, are unfree – subject to what Marx calls 'the coercive laws of competition' (1990, 433). Each individual capitalist is severely constrained in the impact they can have upon the world because they know that deviation from profit-maximisation risks being outcompeted, and therefore they have little choice but to follow the market. Lastly, if pre-emption cases are to be solved, as I will suggest they ought to be, through group agency, this explains the common socialist thought that it is the capitalist *class*, not individual capitalists, who are the proper target of critique.

Climate change might also involve pre-emption. Let's assume that it would be optimific if all of us who could do so without drastically reducing our quality of life reduced our greenhouse gas (GHG) emissions. One might think that one's individual emissions make no difference to climate change because they have such a tiny impact on the levels of GHG in the atmosphere, too tiny to cause significant climate-associated harms (Kingston and Sinnott-Armstrong 2018). Broome (2019) counters that even a tiny increase might push GHG levels over a threshold at which such a harm is caused. Therefore, emitting any amount of GHG, however small, makes a negative expected difference to outcomes – and one ought, other things equal, to reduce one's emissions.

But perhaps our personal carbon emissions are to a significant degree pre-emptive.[60] In that case, your individual emissions would not add to the quantity of GHG emitted – they just mean that more is emitted by you and less by others. Here are two ways in which emissions could be pre-emptive. One is if all carbon that can be emitted will be emitted. There's a finite supply of fossil fuels. Unless some is left in the ground, the peak of GHG in the atmosphere is wholly determined by the size of this supply. If it is to be burnt anyway, my emitting will not incentivise any more burning. The second mechanism is if carbon emissions are regulated by a cap-and-trade scheme. Under such a scheme, firms buy permits to emit a given amount of GHG. If I reduce my emissions, the firms that would have supplied me with carbon-intensive products will buy fewer permits. This will make those permits available to other firms, who will buy them and emit equivalent amounts of GHG.[61]

It's unlikely that pre-emption will render individual emissions wholly inefficacious. Not all emissions require permits, and there is a chance that some fossil fuels will be left in the

[60] The closest I can find to this argument in the literature is Maltais (2013, 591–97) and Cripps (2013, 123–25). They do not closely distinguish between the causal impotence, pre-emption and overdetermination, which I think leads Broome (2019) to mischaracterise and dismiss their argument.

[61] Some environmentalists buy permits and retire them, showing awareness of this problem.

ground. Slowing the pace of emissions may reduce harm. But probably, due to pre-emption, reducing your emissions by one tonne will reduce total emissions by less than a tonne, and thus reduce the harm done by climate change by a smaller amount than Broome and others who ignore pre-emption estimate. Broome estimates that going for a drive in an SUV does expected harm worth $1 (2019, 111). If we account for pre-emption, that figure will be even less – and emitting GHG as a means to a small amount of personal pleasure would maximise expected overall well-being.

So pre-emption cases are present in most people's lives, and also in very high stakes social problems. This is ample reason to investigate them, and to want our moral theories to handle them adequately. As intimated above, utilitarianism seems to struggle to do so. However, other views find these cases problematic too. Moreover, there is a version of utilitarianism that handles them very well, for which I will argue below.

3. Existing responses

I have presented pre-emption cases as a problem for utilitarianism. The problem is that utilitarianism seems destined to require actions that are not only proximate causes of non-optimific outcomes, but are incompatible with the collective action that would bring about the most well-being. As we saw, this places two of utilitarianism's most characteristic and appealing features – its emphasis on outcomes, and its emphasis on difference-making – into conflict with one another.

These cases also pose problems for non-utilitarians. It is easy to see how non-utilitarian act-consequentialists face it: the consequences of chemical warfare, and of George's research job, could be measured by values other than well-being, with similar results. And the problem extends beyond act-consequentialism, too. Many non-consequentialist theories have an act-consequentialist component, bounded by deontological side-constraints. (For instance, those that endorse Otsuka's 'Restricted Principle' (Otsuka 1991).) If act-consequentialism fails to condone and condemn correctly in these cases, these theories would need to include appropriate deontological side-constraints to deal with them, increasing their complexity. And a theory like Scheffler's (Scheffler 1994), which is act-consequentialism plus agent-centred prerogatives, faces arguably even greater problems here than traditional consequentialism, since it would permit (for example) George to take the job even if he would push the research in a slightly *more* dangerous direction than his rival, given its importance to his family and if we subtract his commitment to opposing CBW from his case, whilst utilitarians and act-consequentialists would require rejection in these circumstances (which is why Williams includes the fact that the rival would be more dangerous than George). But for all these non-utilitarian theories, the problem of pre-emption cases arises insofar as they share with utilitarianism a concern for good outcomes and for the moral significance of difference-making. So I will focus primarily on how utilitarianism might be revised to accommodate these concerns in pre-emption cases, which may then suggest amendments to these other theories.

I suggest that there are two desiderata for a revised utilitarian theory in its handling of pre-emption cases.

1. The theory discourages outcomes that are very bad for well-being: catastrophic climate change, exploitation, the development of CBW, and the like. These things would not happen if all agents complied with the theory.
2. The theory acknowledges the importance of difference-making to an action's moral status.

Firstly let's assess standard act-utilitarianism – what I have so far in this thesis referred to simply as 'utilitarianism' – with respect to each of these desiderata. Standard act-utilitarianism fulfils (2): the only thing it takes to determine the status of an action is the difference it makes to outcomes. But it struggles with (1). As we have seen, it requires George to take the job. Now, it could be that enough scientists are in George's position – able to predict that if they don't take the job, someone else will, with worse consequences – that if they all follow the standard act-utilitarian recommendation and take up CBW jobs, CBW will be developed.

There is a stark contrast with rule-utilitarianism. This is the view that an action is morally required if and only if it complies with a rule, widespread compliance with which would result in optimific outcomes (for discussion see Hooker 2000). In general, rule-utilitarianism will require individuals not to be the proximate cause of bad outcomes, because widespread compliance with such a requirement would be optimific. It would be best for well-being if everyone followed the rule 'do not work in chemical and biological warfare', for instance, because if they did, CBW would be eliminated. Therefore, if everyone complied with rule-utilitarianism, CBW would be eliminated. Rule-utilitarianism fulfils (1).

What can rule-utilitarianism say for (2)? At first, it may appear, very little. It is not what will happen if I *x* in which the rule-utilitarian is interested, but in what would happen if everyone (or at least most people) were to *x*. It makes no difference to the rule-utilitarian what would happen if George refused the job given the actual situation – that is, that someone else will take it. But the rule-utilitarianism could adjust their theory such that it did. They could argue that the optimific rules include reference to difference-making. It might be optimific if everyone followed the rule 'do not work in CBW, except when doing so would make a positive difference to the outcome', for instance. But if the rule-utilitarian endorsed this, their theory would cease to fulfil (1). It may be that they can include difference-making caveats to more limited extents ('do not work in CBW, except when doing so would avoid disaster'), but then, by altering the stakes we can always create pre-emption cases in which everyone's following rule-utilitarianism does not bring about the optimific outcome. In general, the more they acknowledge the importance of difference-making, thereby meeting (2), the rule-utilitarian does worse with respect to (1). It would be nice to have a theory according to which both could be secured.

Theories Julia Nefsky (2015) labels 'participation views' pursue a middle way between act- and rule-utilitarianism. They combine both 'act-based' and 'pattern-based' reasons. Parfit suggests such a view. His 'C7' principle says:

> 'Even if an act harms no one, this act may be wrong because it is one of a set of acts that together harm other people. Similarly, even if some act benefits no one, it can be what someone ought to do, because it is one of a set of acts that together benefit other people.' (1984, 70)

As the 'even if' formulation suggests, C7 belongs alongside other principles which take the fact that an act does harm or benefit as morally significant. More recently, Woodard has endorsed a similar pluralism, writing that:

> 'There is a pattern-based reason for agent S to do action X if and only if and because it is possible for S to do X and doing X is S's part in an eligible pattern of action P that would be good.' (2019, 92)

Woodard also recognises that:

> 'There is an act-based reason for agent S to do action X if and only if and because it is possible for S to do X and S's doing X would be good.' (2019, 90)

Since participation views, like rule-utilitarianism, recognise the differences made by collections of acts and, like act-utilitarianism, those made by individual acts, and acknowledge that these may be in tension, they appear to be good candidates for fulfilling all three desiderata. However, I will argue that they fail to do so.[62]

[62] A subtly different theory in this vicinity is Donald Regan's 'co-operative utilitarianism' (1980). This holds that what each agent ought to do is to co-operate with whoever else is co-operating, in the production of the best consequences possible given the behaviour of the non-co-operators. Co-

Insofar as such views incorporate standard act-utilitarian reasons, they will fulfil (2). But participation views fail, I think, to meet desideratum (1). Let's see how they work in the case of George. Consider three relevant patterns of action:

a) All scientists refuse jobs in CBW.

b) All scientists accept jobs in CBW when they know that the alternative is a less scrupulous scientist taking up the role.

c) All scientists refuse jobs in CBW, except George.

Now, according to Woodard, (b) might be the only eligible pattern of action here. In response to an objection from Dietz (2016), Woodard has come to endorse 'the willingness requirement' on pattern-based reasons: we only have such reasons for actions that would be part of patterns in which others are willing to participate (2019, 99). In George's case, (a) and (c) are ruled out by this requirement since it appears that his rival is unwilling to refuse CBW work. In that case, it looks like Woodard's view cannot fulfil (1). But not all participation views need endorse the willingness requirement – and maybe we can avoid it in George's case by changing 'all scientists' to 'all scientists who would be willing to refuse CBW work'. Assuming that the latter covers enough scientists that were they all to refuse, CBW would be eliminated, the problem retains the same structure.

The problem is that all three of these patterns would be good, though they give George incompatible reasons. Moreover, (c) appears to be the optimific pattern, though it provides a reason for George to accept the job. So, in general, participation views don't seem to discourage the production of the bad outcome, the development of CBW.

Pattern (a) would be conducive to well-being, let's assume, because it would bring about a world free of CBW. George's part in this pattern would be refusing the job, so it seems that he has a pattern-based reason to refuse, as do other scientists in similar positions. If the story stopped there, participation views would fulfil (1). But it does not. For pattern (b) would also be somewhat conducive to well-being, because it would bring about a less dangerous development of CBW. And participating in (b), for George, would mean accepting

operation, for Regan, involves a certain kind of motive and a certain kind of action. With respect to action, co-operation requires 'doing one's part in the pattern' that will bring about the most well-being. Thus Regan's view is a kind of participation view. Unlike Parfit's or Woodard's proposals it is not pluralistic; however, it can do justice to our act-consequentialist intuition as in case there is nobody to co-operate with, it gives the same recommendation as act-consequentialism. I believe that it suffers from a similar problem to that I raise for participation views and Nefsky's view.

the job. So it looks like he has reasons on both sides. Given that there is act-based reason for George to accept the job, if pattern-based reasons are undecided it seems that there is nothing encouraging refusal, and thus participation views fail to meet (1).

Now, it might be thought that George's pattern-based reasons to derived from (a) are stronger than his pattern-based reasons derived from (b), since a world without CBW is likely to contain more well-being than a world with relatively safe CBW. But if we are to weight pattern-based reasons in accordance with the amount of well-being the pattern in question would lead to, (c) appears the best of all. This pattern would also lead to a world free of CBW (with as much certainty as all scientists including George refusing, since one scientist is not sufficient for a CBW programme)[63] but would also result in additional well-being for George and his family. Thus, the strongest pattern-based reasons here seem to suggest that George should accept the job.

It is probably true of all scientists that their accepting CBW work whilst the vast majority of others refuse would be better than everyone rejecting CBW work. In both cases CBW is ended, but in the former they receive some additional income. The point is not that it would be better *for them* if they accepted such work, and so they will put their selfish interests ahead of moral duty. It is that it would be better overall, since it would be good for them and no additional harm would come of their accepting the work. So all scientists appear to have strongest pattern-based reasons to accept such jobs. And, of course, they often have act-based reasons, when, like George, they are pre-empting a less scrupulous scientist. But then, participation views do not encourage the best outcomes. Following such views, scientists will keep accepting these jobs and CBW will go on being developed. They fail to meet desideratum (1).[64]

How about Nefsky's own view? She claims that one has reason to help bring about good outcomes, and that one can help to bring about an outcome even if one's action doesn't make any difference to whether it comes about or not. One helps to bring an outcome about, for Nefsky, if at the time one acts it is possible that the outcome will fail to occur for lack of acts of the type you perform (Nefsky 2017).

[63] Thus, it is not the case that George 'is a member of the smallest group of whom it is true that, had these people all acted differently, the other people would not have been harmed.' This is Parfit's condition for George's act being covered by C7 – so Parfit's view, strictly speaking, would not help deliver (1) in cases like George's (Parfit 1984, 71–72).

[64] A similar problem is raised for participation views, in the context of alleged causal impotence cases, by Otsuka (1991).

Does Nefsky therefore not meet desideratum (2)? Strictly speaking, her view does not give difference-making any significance. However, it is possible (as Nefsky herself argues, 2017, 2766–67) that our attraction to difference-making is in large part an attraction to the idea that it matters what an action helps to bring about. Nefsky's analysis of 'help' distinguishes it from difference-making, but speaks to the same motivation. So let's grant that she gets close enough to (2), or at least to the intuition behind (2).

With respect to (1), however, I think Nefsky's view will suffer from a similar problem to that that afflicts participation views. Let's apply it to George's case. Consider three relevant outcomes, corresponding to the three patterns discussed above:

a) A world free of CBW.

b) A world with relatively safer CBW.

c) A world free of CBW, and George is employed.

The problem is that all three of these outcomes would be good, though they give George incompatible reasons. Moreover, (c) appears to be the optimific outcome, though it provides a reason for George to accept the job. So, in general, Nefsky's view don't seem to discourage the bad outcome that is the production of CBW.

I take it that the proposal is meant to work like this: at the time George acts, it is possible that (a) will fail to occur for lack of scientists refusing CBW jobs. Therefore, George's refusing the job could help bring about a good outcome, and therefore he has reason to refuse. All scientists have a similar reason, and thus Nefsky's proposal seems to encourage the best outcome, a world free of CBW. However, there are some problems.

The first is whether (a) is possible (for it to be possible that (a) could fail to occur for lack of scientists refusing jobs, it must be possible that (a) could occur (Nefsky 2017, 2753)). George has been told that if he doesn't take the job 'it will certainly go' to his rival. The implication is that such jobs will generally find willing takers. But then, is there any chance that enough other scientists will refuse to work in CBW that CBW is eliminated?

If not, then Nefsky's view will not provide any reason for George to refuse the job, and she has no claim to fulfil (1). As she says, 'there being other people who will or might act in the relevant way is necessary for being able to satisfy the conditions for helping, and thus necessary for having reason for action coming from how doing so could help.' (2017, 2755–56 - this could be read as her version of Woodard's willingness requirement.) However, she also says that 'the relevant notion of possibility cannot be one that, in general, holds fixed what agents will choose to do in the future' since, in practical reasoning 'we think of agents (both ourselves and others) as typically being able to choose between several different courses of action, where different outcomes can result depending on what they choose to do.' (2017, 2762) So there would seem to be a low bar for Nefsky to consider a set of actions by others possible. Let's grant that on the relevant notion of possibility it is the case that other scientists might refuse such jobs.

Even so, Nefsky's view will struggle to meet (1). As well as (a), (b) is a good outcome – at least, it is much better than many other salient possibilities. And at the time George acts, it is possible that (b) will fail to occur for lack of scientists with his scruples accepting CBW jobs (thus pre-empting more dangerous research). Therefore, George's accepting the job could help bring about a good outcome, and therefore Nefsky's view gives him reason to accept, as well as reason to refuse.

Now, it might be thought that the reasons to help bring about a world free of CBW are stronger than the reasons to help bring about relatively safer CBW, since the former would be a better outcome. Nefsky does indeed say that the strength of a reason to help bring about some outcome will correlate positively with how 'serious' that outcome will be. But if we are to weight help-based reasons in accordance with the value of the outcome in question, (c) appears the best of all. The outcome with respect to CBW would be the same as in (a), but there would be an additional good, George being able to provide for his family. And at the time George acts, it is possible that (c) will fail to occur for lack of George accepting the job. So is the strongest help-based reason a reason to accept the job, on Nefsky's view?

Nefsky mentions two other factors determining the strength of one's reasons to help bring about an outcome: the size of the causal role one would play and the closeness of the chances of the outcome to 50% (2017, 2764). Now, George would surely play a bigger causal role in ensuring (c) than (a). And (a) cannot, from the description of the case, have a significantly different probability than (c): George's refusal to take the job would not make a difference to the likelihood of CBW existing, and all he has to do to have the job is accept an offer that has already been made. So, by Nefsky's account, George's reasons to help bring about (c) – and thus to take the job – are stronger than his reasons to help bring about (a).

It is probably true of all scientists that their reasons for helping to bring about (c)-like outcomes – with their name substituted for George's – are stronger, according to Nefsky's view, than their reasons for helping to bring about (a). In both outcomes CBW is ended, but in (c)-like outcomes they are employed. The point is not that (c)-like outcomes would be better *for them* but that they would be better overall, since they would be better for them and no more harmful with respect to any other value. Moreover, each scientist plays a bigger causal role in delivering a (c)-like outcome – no CBW and they are employed – than in delivering (a), since one's role in one's own employment is bigger than one's role in a large collective action. And (c)-like outcomes can fail to occur, in each case, for lack of the scientist in question taking a job in CBW. So each scientist appears to have, according to Nefsky, strongest help-based reasons to accept such work. But then Nefsky's view does not encourage the best outcomes. Following such views, scientists will keep accepting these jobs and CBW will go on being developed.

Now, Nefsky might respond here that (c)-like outcomes could also fail to occur for lack of scientists refusing CBW jobs. Return to George. The second part of (c), his employment, might fail to occur for lack of him accepting jobs, but the first part of (c) might fail to occur for lack of scientists refusing CBW jobs. So George could help to bring (c) about either by accepting or by refusing the job. But the best this could achieve is indeterminacy with respect to what George has most reason to do. And note that George's causal role in bringing about the second part of (c) by accepting would be far bigger than his role in bringing about the first part by refusing: thus, by Nefsky's own logic it seems that if one reason is stronger here, it is the reason to refuse.

The general problem for participation views, and Nefsky's, is this. Firstly, in cases where other individuals are very unwilling to participate in the optimal pattern, or bring about the best outcome, these theories may have nothing to say. But even if they overcome this, there are a variety of good patterns and good outcomes. And the optimific ones appear to be (c)-like patterns and outcomes, which license agents to do things such as take CBW jobs, burn fossil fuels, exploit workers, and so on. Now, Woodard and Nefsky might want to rule (c)-like patterns and outcomes from playing a reason-giving role. But it is not obvious how this could be done in a way that is justified rather than *ad hoc*, and will not narrow eligible patterns and outcomes too far. (Woodard has a brief discussion of something like this problem himself and comes to the same conclusion.)[65]

Another view is Felix Pinkert's 'modally robust act consequentialism' (2015). Pinkert says that agents ought not only do whatever brings about the best outcome, given the actions of others, but also to be such that they would do the whatever brings about the best in possible worlds where others acted differently. Let's apply this to George's case. The first part of Pinkert's view is shared with standard act-utilitarianism. George should act so as to bring

[65] Does rule-utilitarianism not suffer from the same problem? I think not. The notion of a rule plausibly precludes (c)-like exceptions (is 'don't work in CBW unless you are George' a rule?), whilst the notion of a pattern or (especially) an outcome does not. Moreover, rule-utilitarians will take the costs and difficulties of imposing a rule into account in working out what the optimific set of rules is. As we will see in the next section it is easier to impose a general rule than one with exceptions, so (c)-like rules, even if conceptually admissible, are unlikely to be optimific.

about the bets outcome he can, and thus should take the job, given the greater threat to overall well-being posed by the alternative. This also means Pinkert's view meets desideratum (2) – what George ought to do is determined by the difference his action would make to the outcome. Pinkert's innovation is the second part of his view. This implies that George ought to be such that, if it were not the case that another more dangerous scientist would take the job if he didn't, then he would refuse it. This is an intuitive claim. But is it sufficient to fulfil desideratum (1)?

If CBW research goes ahead, on Pinkert's view, someone is at fault. If George conducts the research, then either: George is not such that he would refuse the job if another more dangerous scientist would not take it – so George is at fault, or: there is another scientist who is such that they would take the job and push the research in a more dangerous direction – so that scientist is at fault. If George refuses the job and another more dangerous scientist takes it, then George is at fault according to the first, standard act-utilitarian part of Pinkert's view, as is the other scientist unless there is a third more dangerous scientist who would take the job if the second scientist didn't, in which case he is in the same situation as George taking the job.

So someone must be at fault if the bad outcome is realised. However, does modally robust act-consequentialism *discourage* such outcomes? It is the characters, rather than the actions, of agents that Pinkert's view would condemn. If George takes the job, nobody has *done* wrong: the fault lies in the bad dispositions of either George or his potential replacements. In fact, as Williams describes the case, it seems that George is modally robust in the way Pinkert requires: he would certainly refuse if it weren't for the threat of a more dangerous alternative. If George is like that, and he takes the job on act-utilitarian grounds, then the parties at fault for CBW are these scientists who serve as potential replacements for George. This is a somewhat strange result. They have not done anything wrong; they need not have done anything to further research in CBW. But because of what they *would* have done, they are the guilty parties. Thus it is debatable whether Pinkert can fulfil (1) in an appealing way.

4. Collectivist utilitarianism

a. Introducing collectivist utilitarianism

My preferred response to pre-emption cases is a version of utilitarianism I will call 'collectivist utilitarianism'. Like Pinkert's, it does not withdraw the standard act-utilitarian judgment that in pre-emption cases, individual actions are required solely on the basis of the (expected) difference they make. But it applies the same judgment to groups: groups ought to act such as to maximise overall well-being. The factory owners *are* required to stop their polluting actions, albeit as a class of factory owners, rather than as individuals.

Although I present this as a revision to the standard act-utilitarian view, it is not a concessive one, but rather an extension of utilitarianism to group agents. Thus, Postow (1977) calls such a proposal 'generalised utilitarianism'. Furthermore, applying utilitarianism to groups is not alien to the utilitarian tradition: Bentham focussed mainly on encouraging governments to maximise overall well-being, assuming that individuals were to a large extent interested in only their own; contemporary effective altruists first aimed their arguments at individuals rather than groups[66] but are increasingly – following criticism on this score (Srinivasan 2015; Dietz 2019; Collins 2019a) – concerned with social decisions (Greaves et al. 2020; Todd 2018).

Collectivist utilitarianism meets both of our desiderata.

1. It discourages outcomes that are very bad for well-being; for instance, it will require that scientists as a group eliminate CBW.
2. It makes the same judgment as standard act-utilitarianism as far as requirements for individuals go. Thus, its requirements for individual action depend entirely on the difference that those actions (can be expected to) make.

Furthermore, it has the virtue of explaining why, in many cases of pre-emption, individuals feel torn between opposing courses of action and justification – why these appear as a kind of moral dilemma. In pre-emption cases, what an individual ought to do stands at odds with what their group ought to do. Scientists, collectively, ought to refuse to work in CBW – but George, a scientist, ought to accept. It is impossible for George to both play his part in his group doing what it ought to do, and do what he ought to do. And George, as we have seen already in this thesis, is not solely an atomised individual, but a social being, whose identity depends in part on the communities to which he belongs. So he faces a choice between two aspects of his identity.

It may be thought that this proposal is vulnerable to a similar objection to that which I levelled at participation and Nefskyan views. It seems that an (maybe even *the*) optimal way

[66] As Peter Singer said in a recent interview: '*Animal Liberation* is essentially saying, "Stop eating the products of cruelly treated animals. You don't need a social revolution—or the social revolution is too difficult to achieve, and rather we should focus on getting individuals to change their practices." Maybe that's had some influence on my thinking, and why I think about individual change.' (Gross 2021)

for the group to discharge its duty is to bring about a pattern of actions that sees George accept the job, with all its benefits, but CBW nevertheless eliminated by sufficient other scientists refusing similar work. This outcome would be better for overall well-being than general refusal. However, it is counterintuitive that the group ought to aim for this outcome, and if it did George would not face the dilemma described above. My response is that large groups in the real world do not typically have it in their power to bring about such patterns of action. Consider the tools that groups have at their disposal. They cannot ensure that each of their members performs the action that is part of the set of actions that brings about the best; they have no mind-control device. Instead, their main tools are rules backed by sanctions, material or moral. In our case, the group of scientists probably has as its main tool the instilling of a general norm to refuse such work. So even if a mixed pattern of actions such as 'everyone apart from George refuses CBW jobs' would be optimal, there is often reason for a group not to aim at it. If groups aimed for such outcomes, they would likely fail: norms with arbitrary exceptions are very difficult to enforce. The group might expect better results by aiming for a general refusal of scientists to take such work by aiming at than selective, minimally sufficient refusals.

Therefore, even if it is possible and optimific for George to accept the job whist CBW is eliminated, scientists as a group ought not aim for this outcome, according to utilitarianism. They would maximise expected overall well-being by aiming at outcomes in which all their members refuse such jobs, including George. When the group does its duty, then, it is highly probable that George refuses the job – since it successfully does its duty when it instils a norm of general refusal sufficiently well that most of its members refuse. At the same time, George's own duty might be to take it.

b. Group agency

Now, the foregoing relies on a picture of group agency according to which what groups can do differs from what their members can do, and, more fundamentally which holds that groups can be treated as moral agents. In technical terms, groups must be subject to non-distributive moral oughts. To grasp non-distributive predicates, consider the following case: the multiple dots in Seurat's pointillist painting *Un dimanche après-midi à l'Île de la Grande Jatte* are beautiful, but this property does not distribute – it is not the case that each dot is beautiful (T. H. Smith 2009, 32; for broader discussion see Oliver and Smiley 2013). 'Beautiful' may therefore apply non-distributively. In the case of George, I claim that scientists as a group ought to refuse jobs in CBW, but that it is not the case that each scientist ought to refuse jobs in CBW.

Some will deny that groups should be treated as moral agents at all. General scepticism of group agency – a kind of methodological individualism, whereby the only agents in the world are individual human beings – is attractively parsimonious but at odds with common-sense. This is evidenced by the ubiquity of group agency in ordinary language. On the *Guardian* website frontpage today I can see that *the House of Lords* has voted through some amendments; that *the Israeli police* has targeted dissidents with spyware; that *a crowd* has paid respects to a murder victim; that *Poland* expects 60 000 daily Covid cases by mid-February; that *we* must stop imprisoning pregnant women and that *we* must tackle gambling addiction. The individualist will try to analyse these attributions of group agency away, and may sometimes be successful: but if there is no need to do violence to a very common feature of language, why should we?[67]

Moreover, individualist paraphrasing will come up against the problem of multiple realisability. The House of Lords voting through an amendment can occur due to any of a very large number of actions of individual lords. It happens if any combination of more than half of the lords who vote vote for the amendment. This may be reducible to a disjunction of possible sets of actions of individual lords. But more problematically for the individualist, the House of Lords can vote through an amendment in a world where the composition of the House of Lords is entirely different. If every lord resigned and was replaced tomorrow, the statement that the House of Lords voted through some amendment the next day would have the same meaning as it did yesterday, even though the individuals involved are different. What makes it the House of Lords's vote is not that certain individuals are involved, but that those individuals are members of a certain group.

If there are group agents, are moral obligations applicable to them? Some of the ordinary language examples mentioned above suggest so ('we must stop imprisoning pregnant women', for example) – and it makes sense, *prima facie*, to say that the House of Lords ought not have passed the amendment. Some groups, such as corporations, are held legally liable for their actions, which also suggests we take them to be moral agents. Exactly which groups have moral agency will be discussed at greater length below, but we have good reason to think that some do.

Suppose then that we are permitted to include group agents in our ontology, and that such agents are in some cases morally responsible. Which sets of individuals are group

[67] There are more theoretical considerations to permit group agents into one's ontology: they feature in some of our best social scientific explanations (Tollefsen 2002), and on at least one plausible recent account of agency, group agency is the paradigmatic form of agency: this is Hyman's account (2015), according to which agency is a matter of various parts of a system acting in an integrated manner.

agents? If the collectivist utilitarian solution is to work in the case of George, or of capitalist exploitation, scientists and capitalists must count as group agents. One account on which they would is Jackson's. He claimed that 'any old aggregation or mereological sum of individual actions counts as a group action' (1987, 93) which suggests that any random collection of individuals can count as a group agent. Most philosophers reject this claim, which Estlund (2020, 217–22) characterises as the 'easy agency' view, and Smith (2009, 36) as 'dubious metaphysics'. I also reject it. It is far too ontologically fecund, generating more group agents than any respectable social science could contemplate.

A less fecund alternative is what we might call the 'corporate agency' view. Such accounts model group agency on individual agency,[68] claiming that a group is an agent only if it shares certain features common to individual agents. For instance, Collins (2019b) holds that a group agent must have a decision-making procedure, and List and Pettit (2011) that it must have motivational and representational states of its own, and the capacity to act on the basis of these. On such accounts only formally structured groups, such as trade unions, political parties and governments, count as agents. Call such accounts 'corporate accounts' and the kinds of group agents they admit 'corporate agents'.

Corporate accounts are compatible with collectivist utilitarianism. It could be that corporate agents ought to maximise well-being, as should their individual members, and these two oughts can come apart. But the corporate agency view will not cover very many of the pre-emption cases in which we are interested.[69] In George's case, is there a corporate agent to which George and sufficient other scientists belong? It is not necessitated by Williams's description of the case, but it is also not ruled out. Perhaps George belongs to a trade union of scientists, with decision-making procedures, stated aims and policies, an outlook on the world. This union ought to see to it that its members refuse CBW jobs, if it's sufficiently powerful that its doing so would end CBW. But we might want our response to George's case to apply even if there were no such union. And there is no corporate agent with respect to other pre-emption cases such as climate change, or capitalist exploitation.

Moreover, if there is no union, we might want to say that scientists ought to establish one. 'Scientists', in such a claim, could not refer to a group agent according to the corporate agency view, since we are prior to the establishment of a corporation. But such establishment is, plausibly, a collective action – not something that someone can do on their own, and

[68] This is not the only way to go: as mentioned in the previous footnote, Hyman's account of agency models the individual on the group.

[69] Or, for that matter, some of the cases we meet in ordinary language – the House of Lords may be a corporate agent, but is Poland? Is the 'we' who must tackle gambling addiction?

perhaps something that each individual scientist ought not to do (for instance, if the costs to them were high and they knew that insufficient numbers of others would join for such a union to be effective). Corporate agency views cannot judge that there is a group who ought to incorporate.

We need, then, an account of group agents narrower than Jackson's but wider than the corporate agency view. One prominent option in the literature is that of Margaret Gilbert. Gilbert (1989; 2013; 2006) claims that many groups constitute 'plural subjects': single bodies of whom we can say they do and believe things beyond what their members do and believe – that is, non-distributively. On her account, such group agents are formed through the 'joint commitment' of the members. The difference between two people walking on the same stretch of road and two people walking *together, as a group*, is that the latter have (perhaps implicitly) undertaken certain commitments with respect not just to each other, but to the group they form.[70]

There is much to like about Gilbert's account, especially for my purposes. It means that non-corporate groups – two friends walking along, a crowd, a whole society – can be agents in their own right, and thus bear non-distributive oughts. However, it suffers from a similar problem to the corporate agency view, with respect to the cases with which we're concerned. In George's case, part of the problem is that scientists have *not* jointly committed to opposing CBW, thus, on Gilbert's view, they have not established a plural subject which can (or ought) do anything.[71] We might want to say that scientists ought to make such a commitment, but again, this seems like it should be a non-distributive ought. It need not be the case that each scientist, if the costs of doing so were high and the likelihood of others joining low, ought to express their readiness to enter into such a joint venture. (This observation is not meant as a criticism of Gilbert's account of joint action, which is not supposed to solve the kind of moral problem with which I'm concerned, but rather to characterise a set of social phenomena.)

So: we need an account of group agents that will include less than the easy and more than the corporate agency view, or Gilbert's. In particular, it must permit group agents that are not corporate agents, or not jointly committed, but could be. For such an account I turn again to Marx. Analysing French society in 1852, he wrote:

[70] For a somewhat similar account, stressing collective acceptance of an 'ethos' rather than commitment, see Tuomela (2013).

[71] It would be different, again, if scientists had established a union – this would involve a joint commitment to abide by union rules, and then it might be the case that the union ought to institute rules preventing its members from working in CBW.

> 'The small-holding peasants form a vast mass, the members of which live in similar conditions but without entering into manifold relations with one another… The isolation is increased by France's bad means of communications… Each individual peasant family is almost self-sufficient; it itself directly produces the major part of its consumption and thus acquires its means of life more through exchange with nature than in intercourse with society. A small holding, a peasant, and his family; alongside them another small holding, a peasant and another family. A few score of these make up a village and a few score villages make up a Department. In this way, the great mass of the French nation is formed by simple addition of homologous magnitudes, much as potatoes in a sack form a sack of potatoes. In so far as millions of families live under economic conditions of existence that separate their mode of life, their interests, and their culture from other classes, and put them in hostile opposition to the latter, they form a class. In so far as there is merely a local interconnection among these small-holding peasants, and the identity of their interests begets no community, no national bond, and no political organisation among them, they do not form a class. They are consequently incapable of enforcing their class interests in their own name… They cannot represent themselves and must be represented… The political influence of the small-holding peasants, therefore, finds its final expression in the executive power subordinating society to itself.' (2000c, 346–47)

This is the source of Marxism's famous distinction between a 'class in itself' and a 'class for itself'.[72] In Marx's view, French peasants constituted the former but not the latter (unlike the proletariat, which constituted both). A class in itself may have a lot in common: they live in similar conditions, with similar culture and interests as Marx says. But only a class for itself has agency – the peasantry can have things done to it ('be represented') but not do things. What distinguishes a class for itself? Such a class has common interests too. But it also has the capacity to struggle for them: by, in the examples Marx gives, establishing community, bonds, and political organisation. That is, we might say, by establishing joint commitments and corporate agents.

Inspired by Marx's conditions for a class to be 'for itself', I propose that a set of individuals constitutes a group agent when it is possible for them to establish a corporate agent to represent them or joint commitments between them. This possibility will often depend on material circumstances: do they have the means to communicate to make corporate decisions; are they in such social relations that they can hold one another to the

[72] For a similar distinction in later Marx-influenced literature, see Jean-Paul Sartre's (2004) distinction between series and group – discussed, with emphasis on gender, by Young (1994).

norms emanating from joint commitments? It may also depend on their mentality: do they identify with one another sufficiently to establish such representatives and commitments; do they have any inclination to do so? Typically, we will find group agents where individuals live in similar conditions and have similar goals, goals that can only be achieved co-operatively, so that they are inclined to work together.[73]

On this view not only are corporate agents moral agents, but so too are less organised groups which corporate agents may represent. (For instance, not only are trade unions moral agents, but so are workers in a particular industry.) The view is not that a potential agent is an agent – but it is that a group with the potential to form a corporate agent, or plural subject, is thereby a member of the broader class of agents itself. Therefore, the set of moral agents is larger than Pettit and List or Collins countenance. This allows us to solve cases like George's, and also to make the kinds of claims narrower accounts of group agency could not: that a set of individuals ought to establish a corporate agent, or to jointly commit to something.

The Marx-inspired account makes scientists and capitalists group agents, and so allows us to ascribe the moral duty to eliminate CBW to the former, and to reduce exploitation to the latter. Scientists have a lot of interests in common: in the ongoing fruitfulness of their work, in securing funding for it, in their rights as workers, perhaps in progress towards truth or at least understanding. Often, these interests can be fulfilled most effectively by working together: establishing unions, research teams, committees and journals; jointly committing to norms such as sharing results, and so on. They also tend to identify as scientists and recognise other scientists as peers. They can and do communicate with one another (we will change this aspect of the case for a less realistic one in the next subsection). Thus, it seems likely that they have the capacity to establish representative corporate agents and joint commitments. Moreover, through the establishment of such things they could eliminate CBW, as we saw with the imagined trade union case above, and doing so would be optimific. Therefore, the collectivist utilitarian can say that scientists, as a group, ought to eliminate CBW – even though individual scientists, due to pre-emption, cannot, and in fact may have a utilitarian duty to accept jobs in it.

Capitalists – take the factory-owners in Young's case – meet these criteria for group agency too. Although they compete with one another, they share certain interests, most

[73] Thus, Elizabeth Cripps's view is roughly equivalent to the Marx-inspired one: 'A set of individuals constitutes a collectivity [i.e. a bearer of moral obligations] if and only if those individuals are mutually dependent for the achievement or satisfaction of some common or shared purpose, goal or fundamental interest, whether or not they acknowledge it themselves.' (Cripps 2013, 28–29)

pertinently in profit and property ownership. Individual capitalists cannot defend these interests alone: they are a minority in society and thus at risk of expropriation by the majority; property rights only persist insofar as they are respected by others; the clothing manufacturers of Bangladesh may only be able to attract orders from American multinationals by guaranteeing low production costs, and this may only be possible if they work together. They also have a degree of class consciousness; identifying with one another if not as capitalists as 'businessmen', or at least 'people like us'. Marx and Engels overstated the case when they wrote that the capitalist state is 'but a committee for managing the common affairs of the bourgeoisie' (2000a, 247), but it is certainly true that capitalists do act together to create corporate agents (chambers of commerce, centre-right parties, industry bodies, and so on) to defend their interests. Thus, capitalists (like scientists) have the capacity to form representative corporate agents – and perhaps joint commitments too (think of 'gentlemen's agreements' around prices in a cartel). They form, according to my Marx-inspired account, a group agent. Capitalists as a group agent have the capacity, which no individual capitalist has (due to pre-emption), to improve workers' conditions in a significant and lasting manner. Given that this would be optimific, collectivist utilitarianism will say that they ought to do so. As long as they do not, collectivist utilitarianism will find fault with capitalists as a class, even if pre-emption means each individual capitalist has an act-utilitarian justification for their actions.

In expanding the set of group agents to include not just corporate agents but also groups with the potential to establish them (or other forms of joint action), I am vulnerable to a common objection: to be a moral agent an entity must have the capacity to consider reasons and perform actions in the light of them (List and Pettit 2011, chaps 1, 7; Tollefsen 2015, chap. 6; Collins 2019b, chaps 2–3). Only agents with decision-making procedures – that is, corporate agents such as unions and parties – seem to have this; groups such as 'scientists' or 'capitalists' do not.

I do not agree, however, that group agents must be able to consider and act on reasons in order to be moral agents. Consider how Pettit and List support this criterion (List and Pettit 2011, chap. 7).[74] They analogise from the individual case, in which – inspired by Christian catechisms and intuition – they hold that morally responsible agents must face morally significant choices, be able to understand and assess those choices, and be able to control which option they choose. The second and third conditions together are put by other authors

[74] Stephanie Collins has a different defence of this condition, stemming from Kant-inspired views about moral worth. I leave a full response to this for another occasion, though I am confident that one can be provided. The List and Pettit argument seems to be more influential and widespread in the literature.

in terms of reasons, as the capacity to consider reasons and act in the light of them. Pettit and List point out that only group agents with decision-making procedures can be properly described as understanding and assessing options (or reasons). As they put it:

> 'A group forms a judgment or other attitude over a certain proposition when the proposition is presented for consideration – it is included in the agenda – and the group takes whatever steps are prescribed in its organizational structure for endorsing it. As we have discussed, these steps may involve a vote in the committee-of-the-whole, a vote in an authorized subgroup, or the determination of an appointed official.' (2011, 159)

Groups without organisational structure, then, cannot form judgments about options, and cannot consider reasons. Non-corporate groups can therefore not be moral agents, on the Pettit and List view. They admit that this implication of their account has 'a distressing aspect', that in cases such as pre-emption, 'although the individuals do something bad together, no one is fully fit to be held responsible.' (2011, 166) I think we can avoid this distress.

Consider the motivation for the condition on moral agents that they are responsive to reasons. Pettit and List give little beyond analogy from our intuitions and Christian doctrine about the case of individual agents. Tollefsen (2015, chap. 6) gives a more profound motivation. The crucial question, she says, is whether addressing an agent in moral terms is futile (2015, 126). The reason that individuals must be responsive to reasons in order to be moral agents is that there is no point in offering moral reasons to an individual who lacks the capacity to recognise reasons and control their actions in the light of them. Telling such an agent – consider a pet dog – that they've done wrong has no force, at least, when 'wrong' has a moral aspect and is not simply registered by the dog as a word that often precedes punishment. The same does not go for groups that are made up of reasons-responsive members. Telling a group that it has moral reasons to do this or that need not be futile, so long as some of its members are responsive to such reasons. And in many cases rational individuals are. Economists have shown experimentally, and justified theoretically, that rational agents often 'team reason'; that is, they act in the light of considering what the group of which they are a member ought to do (Bacharach 2018; Karpus and Gold 2017; Sugden and Gold 2007). Addressing moral arguments to a group, then, need not be futile even if that group itself is not able to respond to reasons. Such address is likely to change its members' behaviour. This doesn't mean that it is reducible to giving reasons to its members as individuals. As Tollefsen puts it:

> 'Normative competency involves the appreciation of norms. This appreciation [in the case of group agents] is best understood in terms of the capacities of individual group members. But group members will appreciate norms qua group members. Outside of one's group, these norms may not have any force for the individual qua individual.' (2015, 129–30)

In the case of George and the scientists, we might say that the scientists ought to eliminate CBW research, and that individual members like George are capable of responding to that moral injunction – including by doing things such as enforcing a norm against taking such jobs – since they regard it, in the manner of team reasoning, as applying to them as members of the group. At the same time, George as an individual ought to (according to collectivist utilitarianism) maximise expected overall well-being, which might mean taking the job.

Tollefsen's official focus is on corporate agents, but we can see how this argument may include non-corporate groups as morally responsible agents. Non-corporate groups such as 'scientists' or 'capitalists' have members who are reasons-responsive (or 'normatively competent'), and who at least sometimes are affected by the reasons they take their group to have. Pettit and List seem to appreciate something similar. Having officially denied the moral agency of non-corporate groups, they then concede that 'something close to holding a group responsible seems more appropriate, even when the group is not an agent in the ordinary [corporate] mould.' (2011, 167) They focus on non-corporate groups that are in some way represented by a corporate agent as a 'spokesbody' – as a nation might be by a government, members of a profession by a union, or a class by a political party. They note that ascribing moral responsibility to non-corporate groups has an important point:

> 'To refuse to ascribe responsibility to the group as a whole, on the grounds that the evil done was done entirely by the spokesbody, would be to miss the opportunity to put in place an incentive for members of the group to challenge what the spokesbody does, transforming the organizational structure under which they operate: making it into a structure under which similar misdeeds are less likely. By finding the group responsible, we make clear to members that unless they develop routines for keeping their government or episcopacy in check, they will share in member responsibility for what is done by the group…' (2011, 168–69)

I do not see, firstly, why the same reasoning cannot be applied to non-corporate groups that lack a corporate spokesbody. Ascribing responsibility to such a group could put in place an incentive for its members to *produce* a spokesbody that would alter the structures under which each member acts so as to produce better outcomes; it could also make clear to members that they should develop routines for keeping *one another* in check.[75] Members will respond to such ascriptions of group responsibilities insofar as they engage in team reasoning with respect to the group. Secondly, it is unclear why, having made this point, Pettit and List still officially deny that non-corporate groups can be morally responsible agents. They admit that they can sometimes be appropriately held responsible. They say that this simply means that they are 'responsibilisable' rather than 'fully responsible agents' (2011, 169). But this dispute may be merely verbal. Once we are ascribing moral reasons and duties to something, what is to be gained by denying it is a moral agent? For collective utilitarianism to make sense, in any case, I think it sufficient that we have a rationale for ascribing moral duties to such groups.

The view is, then, that we can treat an entity as a moral agent even if it is unable to recognise and act in light of the reasons that it has, as long as ascribing duties to it is not futile. It will not be futile so long as the entity includes members who are able to recognise and act in light of the reasons that it has. Non-corporate groups meet this condition by having members who engage in team reasoning, or something like it. Thus, non-corporate groups may bear moral duties, as the collective utilitarian solutions to the cases of George, and of capitalist exploitation, require.

c. Cases without group agents

In this section so far, I have argued that the most plausible position is that group agency exists and that groups such as scientists and capitalists have it. This means that collectivist utilitarianism meets our three desiderata in a range of pre-emption cases. But since I reject 'easy agency', I have to concede that it will not do so in all pre-emption cases. If the agents involved do not belong to some group agent that maximises overall well-being by discouraging the actions in question, collectivist utilitarianism looks just like standard act-utilitarianism.

In a case of Jackson's (1987), drivers on a motorway each drive at 80 miles per hour. It would be safer (to an extent that outweighs any costs of slower travel) if all drove at 60 miles per hour. However, because it is safer if all cars travel at similar speeds, given that other

[75] Thus, by treating the group as a moral agent we help it turn into a corporate agent or plural subject. Margaret Little, crediting Mark Lance with the point, notes that we do something similar with children, which she calls 'proleptic engagement': 'it is by treating the young as if they already were in the next stage that we help usher them into it. It is by treating the infant as a creature capable of learning that we help it become more robustly a creature that can learn; it is by treating a young child as a fledgling agent that we help it turn into one.' (2007, 342)

drivers will drive at 80 (which can be reasonably expected), any individual driver's driving at any different speed (including 60) would be less safe. This is not strictly speaking a pre-emption case, but it is close to it. Jackson's response is along collectivist lines: each driver ought to drive at 80, but the group ought to drive at 60. The problem is, it is hard to see how Jackson's drivers constitute a group agent. They all have interests in safe and efficient travel. However, each in their own cars, without communication, and thus no way of agreeing or enforcing any collective decision, it is hard to see how they could organise: they are more like Marx's peasants than his proletarians.

Here, an alternative to collectivist utilitarianism suggests itself. Rather than a group ought, we might posit a joint ought. The basic idea of the joint ought solution is that the predicate 'ought to X' can be borne by *multiple agents non-distributively* (T. H. Smith 2009; Estlund 2020, chap. 12; Mellor, n.d.; Schwenkenbecher 2020; Pinkert 2014). Collectivist utilitarianism also endorses non-distributive oughts: it can be true that a group ought to do something, but false that any member of the group ought to do it. This is key to its approach to pre-emption cases. But according to collective utilitarianism there is still a single agent, the group, who bears the ought. The joint ought view has the ought borne by individual agents, though non-distributively. To return to Seurat: the art critic who is the analogue of the collectivist says that the painting – a single whole made up of dots – is beautiful, though each dot is not; the one who is the analogue of the joint-ought theorist says that the dots – as distinct but multiple entities – are beautiful, though each one is not.

Since joint oughts do not distribute, the joint ought view provides a very similar response to pre-emption cases as collectivist utilitarianism. It could be the case that scientists jointly ought to refuse CBW jobs, but George singly ought to accept his offer. But in cases like Jackson's, the joint ought view can issue an ought that collectivist utilitarianism cannot: it can hold that the drivers – not as a group, but jointly – ought to reduce their speed.

One reason to prefer the joint ought theory over collectivism is that it can get by with a sparser ontology; it has no need to posit groups as further entities (perhaps worse, as further agents) over and above individuals. But it pays for this, I think, with a more revisionary *ideology* (in Quine's sense, not Marx's). Even proponents of joint oughts find them 'odd' (Mellor, n.d.) or 'obscure' (Estlund 2020, 236). Most oddly, I think, they imply the possibility of wrongdoing with no wrongdoer. If the drivers do not reduce their speed, who has done wrong? No individual amongst them. No group agent either. We could say that they have jointly done wrong. But what practical import does this have? There is no agent responsible for righting it. Once everyone learns that they are transgressing a joint obligation, are they motivated to act differently at all? (Admittedly, most proponents of the joint ought view hold that individual obligations derive from joint ones (e.g. Schwenkenbecher 2020, chap. 5). But this view is not available to the utilitarian, as pre-emption cases show.) In contrast, on the collectivist view, once everyone learns that they are part of a group that is doing wrong, the group can be held responsible, and groups can act to change their ways; furthermore, individuals insofar as they identify with the group will be moved to assist in the group doing the right thing (even where this is in tension with what they, as individuals, ought to do). A joint ought, in short, risks being practically impotent.

However, if the joint ought view gives a better account of cases where no group agent plausibly exists, this might tilt the scales in its favour, vis-à-vis collectivist utilitarianism. As stated above, the joint ought theorist can say what the collectivist utilitarian (at least, one who rejects 'easy agency') cannot – that in such cases there is something morality can say for encouraging the best outcome. It thus looks like it meets our desideratum (1) in a wider range of cases.

But perhaps it is intuitive that where there is no group agent, desideratum (1) is less important to meet. Is it really the case that Jackson's drivers ought to reduce their speed? (We can answer 'no' to this question and still maintain that the government ought to enforce a reduced speed limit.) It is telling that in making this case for the joint ought view, Smith has to posit the possibility of communication and mutual commitment between the drivers (2009, 43). Thus, they become more like a group, before they can bear an obligation. But in this case, why insist that it is the drivers separately and a joint ought, rather than the drivers as a group and the standard, singular ought?

Return to George's case. As argued above, according to the Marx-inspired criteria scientists form a group agent. But imagine a possible world in which CBW research is conducted by atomised individuals unable to communicate with one another. George, like each of them, is simply told to conduct some particular experiments, alone; and he is told that success in these experiments will contribute to CBW. He is not told anything about any other people doing similar experiments – how many of them there are (if any), how alike they are in background and interests, whether they are working on different parts of the same research programme or each trying different approaches. The only thing he is told about other people is this: there is a very unscrupulous researcher who will be employed in his place if he refuses the job.

Scientists in this case would have no prospect of organising a corporate agent or joint commitments. They would thus not form a group agent, on my Marx-inspired view. The case is so far-fetched that it is difficult to rely on intuition (demonstrating how much social connection is involved in the original case), but my inclination is to think that George's moral obligations in this non-social case are more straightforward: he should simply take the job.

This suggests that where no group agent exists, meeting desideratum (1) is less important, so amounts to a defence of the collectivist utilitarian response.

This claim is also strengthened by comparing George's case with that of Jim. Williams himself suggests that the standard act-utilitarian recommendation is 'probably right' in Jim's case (1973, 117), though he is more sceptical about its requirement for George. One explanation for this is that George belongs to a group agent, whilst Jim does not. Jim, recall, is a foreigner who knows nothing of the people involved in his situation. He and Pedro are not mutually dependent in any way, so it seems more reasonable for him to treat Pedro as a force of nature, and adjust his actions to maximise well-being given what Pedro will do. George on the other hand knows the man who is making him the offer; he shares a discipline and profession with him and with his rival; the act in question can only be done co-operatively by members of the group (scientists) to which he belongs. According to collectivist utilitarianism, this means there is an obligation for some agent to aim at George refusing the job (though the obligation belongs to the group, not to George), no counterpart of which attaches to Jim's refusing Pedro's offer. If it is more intuitive that Jim ought to accept than George, collectivist utilitarianism therefore has a good explanation of this; moreover, we have more evidence for the claim that meeting desideratum (1) is less important where no group agent exists.

Collectivist utilitarianism still has some things to say about cases where there is no group agent which reduce any advantages which the joint ought view might possess. There might not be a group of scientists with sufficient capacity to organise who ought to eliminate CBW and so impose norms to refuse such work, but there might be other agents in this position: for instance, the anti-war movement, or the state. And it might be that some individual agents, possibly including George, have a duty to (a) pressure agents like the state to impose such norms (what Cripps (2013) calls 'promotional duties') and/or (b) bring into being a group agent of scientists who can eliminate CBW into being (the kind of process Marxists might refer to as 'class formation', or feminists as 'consciousness raising'). This would not be because individuals have such duties as a matter of course, but rather because they are plausible means to bringing about well-being. All we cannot say, given that such an agent does not presently exist, is that there is an agent, of which George is a member, who ought to see to it that George refuses the job. The joint ought view, on the other hand, can say that there are agents, including George, who jointly ought to refuse such jobs. Given the obscurity of the joint ought, is this judgment, in these cases, advantage enough to prefer the view over collectivist utilitarianism? I do not think so.[76]

5. Compunction

But what, to reiterate Williams's question, should George do? It seems that the collectivist utilitarian's answer is the same as the standard act-utilitarian's: he should take the job. What importance does the group obligation to refuse really have for George, or any actual scientist? Collectivist utilitarianism issues non-distributive group obligations: thus, George is not obligated to play his part in scientists collectively refusing to work in CBW research (in fact, he is obligated to do something that would undermine this group action, namely, to take the job). Dietz, who argues for group duties on the grounds of other collective action problems, writes that as far as utilitarians and consequentialists are concerned, the duties of groups 'have no implications for what individuals ought to do' (2016, 980). But I now want to argue that, on the collectivist utilitarian picture, there is another sense of 'ought' according to which what his group ought to do does affect what George ought to do, and that this may account for the intuition that he ought to refuse the job.

Call the sense of 'ought' used so far in this paper 'ethical ought' (Sidgwick 1884, 33–34): the one familiar from contemporary moral philosophy, the one present in conclusive answers

[76] Ashford (2007) considers a similar set of cases, based on variations of Parfit's 'harmless torturers' example (1984, 80–82). In her first case, there is a Torturers' Union set up to carry out torture in such a way that no individual torturer makes a (significant) difference. She judges that 'the behaviour of these agents should still be classified as torture and as a human rights violation' (2007, 195). Then she imagines the group agent, the Torturers' Union, does not exist, and the torturers work in atomised isolation, with the same effects. She writes that 'it is still plausible to hold each of them responsible for their share of the overall harm they are together causing' (2007, 196). To what extent does my view diverge from Ashford's? In the first case, collectivist utilitarianism will hold that the Torturers' Union transgresses the ethical ought not to torture, though the individual torturers may not; in the second, if the individual torturers have no more optimific options (e.g. because they are pre-empting worse torturing by others), then there is no ethical ought transgressed. However, in both cases, I may agree with Ashford that it would be better if our legal and moral norms *classified* what the torturers do as torture, and that they *hold them responsible* – for punishment, blame and so on. These are matters of political and compunctive oughts. Below I discuss the distinction between political, compunctive and ethical oughts. The difference, if there is any, between my view and Ashford's is about what the individual torturers ethically ought to do – I say, 'if there is any', because although Ashford is clear about how their actions should be classified she appears to leave open the question of their ethical oughts.

to the question ‘What should I do?’ (Williams 1981b) – also called ‘the prescriptive ought’. According to this ought, George ought to take the job but scientists as a group ought to refuse such work. There is another sense of ‘ought’ in the utilitarian tradition, that which Sidgwick called ‘political ought’ (also called ‘evaluative ought’). It is familiar to us from everyday speech. Using this sense, we might say something like ‘Children ought not be dying in factories’ (Humberstone 1971, 8). The meaning of this sentence is something like ‘it would be better if children did not die in factories’.

In the political sense of ‘ought’, it may be true that George ought to refuse the job – in a better world (for instance, one in which scientists as a group established a norm of refusing such work) he would. But one might object that this is mere wordplay. The intuition that Williams’s case brings out is not just that there is an English sentence ‘George ought to refuse the job’ which is true under some interpretation, but rather that people in George’s situation bear many of the marks of being under an ethical obligation to refuse: they feel motivated to do so and guilty if they do not; they are liable to criticism, blame and perhaps punishment if they accept such jobs. Political ought doesn’t seem to explain these things.

Collectivist utilitarianism can, I think, account for these marks of obligation, even whilst it sticks to the line that ethically, George ought to accept the job. Consider a third sense of ought, ‘compunctive ought’ (this name inspired by Hare 1981). When someone fails to do something they compunctively ought to do, they appropriately feel remorse and guilt, and are liable to blame, criticism and punishment by others. These things being unwelcome, people are generally motivated to follow compunctive oughts.

Often, compunctive oughts will line up with ethical oughts. Pre-emption cases are interesting, I think, for belonging to a small set of cases in which they do not. George compunctively ought to refuse the job – and this is well explained by collectivist utilitarianism.

With regard to guilt, blame, criticism and punishment, consider this from Mill (2008c, 184):

> ‘We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, then by the reproaches of his own conscience.’

Mill focuses on cases in which some action would be conducive to utility (thus ethically ought to be done), but punishing its absence would not be (thus it is not ‘wrong’, or it is not the case that it compunctively ought to be done). But in pre-emption cases, we have actions that we ethically ought to do, but compunctively ought not to. Recall that groups may have an ethical obligation to make their members perform some actions, as a means to the optimal collective action. For instance, the group 'scientists' ethically ought to see to it that all of its members refuse jobs in CBW. One tool it has to do this is criticism, blame and punishment of those who accept (and praise of those who refuse). It could be justified in using this tool, according to collectivist utilitarianism, even if the individuals in question have themselves, as individuals, acted as they ethically ought. For instance, George might be liable to ostracism from the scientific community, as a warning to others not to go down the same route.[77]

Moreover, George might be justified in internalising many of these norms and thus blaming himself – that is, feeling guilty – for accepting the job. Firstly, it is likely that he, as well as his group, ought to blame, criticise and punish other scientists who take such jobs. Such blame encourages scientists to refuse. Whilst in individual cases like George's it would not be optimific to encourage refusal, it may well be that as a general policy George would maximise well-being by encouraging his colleagues to refuse. (This will of course depend on how many of his colleagues George can influence, and by how much.) Moreover, it is very likely that he politically-ought to internalise this norm in this way: if the group discharges its duty of eliminating CBW, it will likely be by bringing it about that people like George would feel guilt for taking these jobs, so in the salient better states of the world, George internalises the norm such that he would feel guilt.

At the limit, these group-imposed norms may solve pre-emption cases entirely – that is, make them no longer pre-emption cases by aligning what the individual and the group ethically ought to do. Guilt and blame and punishment adversely affect well-being. So where these norms are imposed George's choice is no longer between taking the job and the money, and refusing the job to see his rival pursue it more zealously. Rather, it is between taking the job and the money minus the costs of punishment, blame and guilt, and refusing the job to see his rival pursue it more zealously. The psychic and punitive costs of taking the job could become so high that it no longer maximises well-being for George to do so.

This resolution is not particularly plausible in George's case, given the high stakes: the costs of a rival scientist making CBW more dangerous are likely higher than the costs of any guilt George could feel, or punishment he could suffer. But other cases could be resolved in

[77] There is an interesting parallel here with Gilbert's view that parties to a joint commitment have obligations (to one another and the group) to abide by this commitment, and that breaches make them fit for 'rebuke' (i.e. blame, criticism, sanction) by fellow parties, and guilt. On Gilbert's view these obligations are not moral, just as in my Millian account they belong to a compunctive rather than ethical ought.

this way; for instance, the guilt I feel – or maybe the punitive tax I am made to pay – for going for a drive in my petrol car might outweigh the joy I get from it even if the environmental costs are to be discounted because I am merely pre-empting others. But note that the group may be justified in blaming, criticising and punishing – in imposing a compunctive ought – even if these things cannot fully resolve the problem in this way. That is because most of us, most of the time, do not perform the optimific action, but are motivated disproportionately by the norms we internalise and our own well-being. So a group imposing these things on its members could secure the optimific collective action even if the individual optimific actions remained counter to it.

The picture we now get is this. Groups ought to perform optimific actions, and so ought individuals. In pre-emption cases, they come apart. This explains the dilemmatic feel of such cases. Furthermore, we can explain the attraction of the thought that individuals like George really ought to play their parts in optimific collective actions: this is true in two senses of ought other than the primary one used in ethics; the latter of these also justifies many of the practices and attitudes which often attach to wrongdoers being attached to George, even if he correctly follows the utilitarian recommendation.[78]

6. Conclusion

In this chapter I have shown how pre-emption cases are important for a moral theory to adequately respond to, and shown how collectivist utilitarianism can make a better response to such cases than standard act-utilitarianism and several other possible revisions of it. This constitutes a point in favour of collectivist utilitarianism. Since capitalist exploitation is one of the pre-emption cases collectivist utilitarianism solves, and it does so in accordance with the socialist emphasis on group agency, we also have another point of contact between utilitarian and socialist thought.

Collectivist utilitarianism, by acknowledging the importance of our membership in group agents, invites us to look at moral agency in an appealing way. The standard act-utilitarian agent is a lone wolf, tasked with making the world as good as possible whilst taking the actions of others as given, treating them, as Broome has it 'as a force of nature, not as a group of strategic agents interacting with you.' (2019, 121) Such an attitude is especially

[78] My proposal here is consonant with the project Peter Railton calls 'valoric utilitarianism' (1988), which also emphasises the different kinds of moral assessment a utilitarian might give, and the possible detachment of judgments of rightness from judgments of blameworthiness. Railton, though, does not put his proposal in terms of different 'oughts', or the interaction between groups and individuals.

unsettling given the demands that utilitarianism places on agents to treat the well-being of others as seriously as their own:[79] compared with common-sense morality, standard act-utilitarianism reduces of the status of others as agents, but increases their status as patients. There is something to be said for this picture: when I ask 'what should I do?', I want an answer that reflects the way the world actually is, including how others actually are, and what I can actually do about it. But there is a degree of dissociation from others involved, and with it a narrowing of possibilities for joint action. Collectivist utilitarianism, through its group agency aspect, recognises our bonds with others.

Morality, moreover, is at least in part a collective enterprise. As Regan puts it, 'one of the basic facts about the world is that there is a multiplicity of moral agents. I am a moral agent, and I have moral decisions to make, but I am not alone. I share that condition, both liberating and burdensome, with many other persons.' (1980, 207) Emphasising collective duties and the oughts that spring from them brings others back into the picture as agents, and allows individuals to cleave to identities not simply as lone wolves, but also pack animals. This, I think, is a fuller understanding of the human moral condition.

[79] It is also profoundly un-socialist: recently, Gabriel Wollner has argued that a central motivation for socialism is that it 'make[s] it possible for people to really act together' (2020, 286), regarding one another as co-operating agents, and thus avoiding alienation not just from them, but from the actions we perform with them.

Chapter Seven: Utilitarianism and Egalitarianism

0. Abstract

This chapter responds to the objection that utilitarianism is insufficiently egalitarian. In fact, I argue, utilitarianism will generally favour the kinds of equality that real-world egalitarians care about: equality of wealth and of social relations. There are a variety of reasons why it should do so, drawn both from recent social scientific work on the determinants of health and subjective well-being, and recent philosophical work on relational egalitarianism and hierarchy.

It may be objected that merely favouring egalitarian policies is insufficient for a theory to be egalitarian, for instance, if that favouring is contingent. I argue that utilitarianism is every bit as egalitarian, in every way that matters, as contemporary egalitarian theories. I also show that its attitude towards distributive justice is similar to the one Marx propounds in his *Critique of the Gotha Programme*. Utilitarianism, therefore, should be regarded as an egalitarian theory.

1. Introduction

a. Preliminaries

Williams writes that utilitarianism is ‘indifferent to issues of justice or equity’ (1973, 137), and criticism of the theory from the egalitarian side is widespread (Rakowski 1993, chap. 2; Rawls 1971, 22–33; Dworkin 2000, 328–30).[80] It is said by egalitarian critics that utilitarianism cannot accommodate their concerns: the utilitarian does not care how things are distributed, or how we stand in relation to one another, so long as overall well-being is maximised. This criticism is one reason for utilitarianism’s low stock in philosophy, and amongst socialists. In recent times, it has pushed several philosophers who are relatively sympathetic to utilitarianism to revise the theory to accommodate distributive concerns – for instance, we have the priority views of Parfit (1991) and Buchak (2017); the sufficiency views of Crisp (2003) and Frankfurt (1987); and the sensitivity to distribution that motivates Ashford’s non-aggregative individualist utilitarianism (2005). In this chapter I want to investigate how far an unrevised utilitarianism can accommodate egalitarian concerns. My

[80] More frequently than explicitly rejecting utilitarianism on these grounds, egalitarian philosophers neglect it: as Woodard points out, most contemporary discussions of distributive justice often overlook or dismiss utilitarianism (Woodard 2019, chap. 7 note 3; he cites G. A. Cohen 1989, and Anderson 1999 as examples).

conclusion is that it can do so very far indeed, far enough that it may be properly called a form of egalitarianism. I draw this conclusion based largely on four things:

1. A recognition that the demands of active egalitarian political movements tend to centre on the redistribution of wealth and the elimination of hierarchies.
2. Recent social scientific work on the importance of wealth distribution and hierarchical social relations to health and well-being.
3. A recent account of hierarchy that, suitably amended, implies that hierarchy is always, other things equal, non-optimific.
4. A comparison with a leading contemporary egalitarian theory, which shows utilitarianism to have equivalent egalitarian credentials.

So far in this thesis I have discussed utilitarianism as a requirement on individual (or group) actions. But it is not actions that may be more or less equal, but social arrangements. As a preliminary, we should clarify exactly how utilitarianism applies to these. As a general thesis, it holds:

U_s: Society ought to be arranged such that overall well-being is maximised.

There are practical problems with using U_s as a decision-procedure. For one thing, it is radically uncertain which, out of all possible social arrangements, are the ones which will maximise overall well-being. For another, it may be impossible for any agent – even the large group agents discussed in the previous chapter – to execute a complete reform of society. Lastly, even if some agent could do so, its aiming for the social arrangements that maximise well-being might increase the risk of bringing about arrangements that were very suboptimal, such that it would be better to aim for slightly suboptimal social arrangements.

Acknowledging these difficulties, utilitarians will find the following principle more useful for social decision-making:

> *U_t: Out of two social arrangements that one could bring about, one ought to bring about the one that leads to greater overall well-being.*

In this chapter I will argue that U_t often favours social arrangements that are preferred by egalitarians. This is no accident; it is because the features of social arrangements that egalitarians value, such as flat distributions of wealth and non-hierarchical social relations, are features that tend to lead to more overall well-being. They provide utilitarians with *pro tanto* reasons: other things equal, a more egalitarian social arrangement is more conducive to utility. In cases where other things are not equal, such reasons are overridable and sometimes overridden, according to U_t, but this is true for all plausible egalitarian theories.

b. The distribution of well-being

It may seem that there is an obvious response available: utilitarianism respects equality because overall well-being is calculated by summing each individual's well-being, *equally weighted*. This means that a unit of well-being (whatever that is) matters equally, for the utilitarian, whether it accrues to you or to me or to anyone else. Utilitarianism gives no priority to the powerful, the high-born, the fortunate, or the owners of property.

In its early days, this aspect of utilitarianism certainly put it on the progressive side of politics. But it is this very equal weighting that leads utilitarianism – seemingly – to be quieter than contemporary egalitarians on distributive questions. Because the theory puts equal weight on each individual's well-being, it cannot hold that it matters if some get more than others. A unit of well-being has equal value whether it accrues to Bill Gates or a Bangladeshi factory worker. Utilitarianism gives no priority to the weak, the hard-working, the unfortunate, or the worst-off.

Is it true that utilitarians do not care how well-being is distributed? The answer is a qualified 'yes'. Qualified, because it is too quick to say that a unit of well-being matters equally, for the utilitarian, whomever it accrues to. Firstly, this is because well-being can beget well-being, and does so to varying extents. The mother of five children becoming happier has positive downstream effects on the well-being of her children; if a hermit becomes happier to the same degree, such effects are less likely. Secondly, there may be negative downstream effects on well-being stemming from inequalities in well-being. If those around you are much happier than you are, this may become a source of pain for you that would not exist if your lives were going equally well. So if a utilitarian somehow had a parcel of well-being to hand out, they would not necessarily be indifferent as to who should receive it.[81]

However, this is only because the parcel could affect overall well-being to a different degree than was contained within the parcel itself. For any *fixed amount* of overall well-being, utilitarianism is indifferent to how it is distributed. Another way of putting this:

[81] There is some evidence that unequal distributions of well-being lead to less well-being overall, partly through lower levels of trust in societies with less equal well-being levels (Helliwell et al. 2020, chap. 2). If this is true, utilitarianism cares about the distribution of well-being now because it might affect the sum of well-being in the future. But it would remain true that holding the sum of well-being fixed, utilitarianism is indifferent between distributions of it.

utilitarianism will never prefer a state of affairs which instantiates less overall well-being than some alternative, no matter what distributional differences there are between the two.

Here is one sense, then, in which utilitarianism is silent on distributive questions. For fixed amounts of overall well-being, it does not care how that well-being is distributed. It puts no weight on the distribution of well-being, except insofar as it affects its overall sum. This is the form of the most common philosophical criticisms of utilitarianism from equality. Philosophers calling themselves egalitarians often make claims such as 'it is clear that there is unfairness when some are better off than others and that this unfairness is absent when people are equally well-off' (Otsuka and Voorhoeve 2018, 68), with 'better off' and 'well-off' conceived in terms of well-being.[82] Utilitarianism denies this. As Williams characterises the utilitarian view, it holds that:

> 'there is nothing to choose between any two states of society which involve the same number of people sharing in the same aggregate amount of utility, even if in one of them it is relatively evenly distributed, while in the other a very small number have a very great deal of it.' (1973, 142–43)

Williams takes this position to be obviously false and counter to the egalitarian spirit. But when we consider egalitarianism not as some philosophical principle but an aspect of progressive political movements this denial does not seem to be a flaw. The distribution of well-being is not something that most egalitarians – at least, those active in streets and legislatures rather than seminar rooms – care about. Consider the following case:

> WEALTHY MAN: A wealthy man, respected by all and with great power over his employees, suffers from regular migraines. This severely affects his well-being: aside from the pain itself, he becomes unable to focus on his work and his bad mood strains his close relationships. His life is now worse (in terms of well-being) than that of the woman who cleans his office: although she has one-thousandth of his wealth and her opportunities are restricted by racial and gender discrimination, she is in good health, has a happy and loving family and spends her small amount of free time pursuing an inexpensive hobby which gives her a lot of joy.

If we care about the equal distribution of well-being, then we should think that it would be in some way good to redistribute well-being from her to him. Since money and leisure time make a positive contribution to well-being for both of them, it should be an improvement,

[82] Temkin (2003) and Arneson (1989) both endorse (roughly) this welfare egalitarian view, with the caveat that inequalities in well-being that result from the fault or choice of the worse off are unproblematic.

from the point of view of equality, for her wages to be reduced and for him to keep the difference, or for her to work unpaid overtime doing more chores for him. Now, I take it that most of us find such a policy unattractive. That we find it unattractive does not mean that we do not care about equality – in fact, it seems to be our concern for equality, in part, that such a policy offends. This shows that insofar as we care about equality, we care about something other than the equal distribution of well-being. In this we agree with egalitarian movements in real-world politics. Feminists, socialists and anti-racists take themselves to be engaged in a struggle for equality. But they do not endorse redistribution towards the boss. A boss's headache, however painful, is not a pressing matter for egalitarians, and greater exploitation of the poor by the rich is not something they should favour.[83]

If egalitarians are not concerned with the equal distribution of well-being, what are they concerned about? One thing all egalitarian theories are concerned about is the distribution of wealth. We see this in egalitarianism outside the seminar room. The French Revolution's call for 'égalité' was a demand for – amongst other things – a redistribution of the wealth of the aristocracy and church.[84] Socialists characteristically favour redistributive taxation and the collectivisation of property. Movements for racial and gender equality draw attention to wealth gaps between whites and people of colour, and between men and women. Egalitarian philosophical theories also imply that the distribution of wealth is important. Many such theories claim that there is a 'currency of egalitarian justice': some good such that (as far as egalitarianism goes) we ought to all have the same amount of it. Some take the currency of egalitarian justice to be wealth itself (M. O'Neill 2008; 2017), or bundles (such as resources or primary goods) of which wealth is part (Dworkin 1981; Rawls 1971). Those who take well-being, freedom (van Parijs 1997), capabilities (Sen 1987), or 'access to advantage' (G. A. Cohen 1989) to be the currency of egalitarian justice should also be concerned with the distribution of wealth, since inequalities in wealth cause inequalities in these things.

[83] On the other hand, we ought not be indifferent to the boss's suffering, and of course utilitarianism, with its concern for well-being, holds that his migraines should be relieved (if it is possible without significant costs to others). Denying this, I take it, would be extremely counterintuitive. Now it is the egalitarian's turn to be silent. For, given that egalitarians are not interested in equalising levels of well-being, it is difficult to see how a concern for equality justifies curing a privileged man's migraines. A plausible egalitarian view, then, will combine concern for equality – which does not mean an equal distribution of well-being – with some other concerns, concerns which can justify curing wealthy men's migraines. The point – that egalitarians must be pluralists – will be important below.

[84] Not that it was successful: as Piketty finds, 'After a slight decrease during the French Revolution, the concentration of wealth [i.e. share of wealth held by the top one per cent] increased in France (and even more in Paris) during the nineteenth-century to the eve of World War I.' (Piketty 2020, 128)

Some egalitarians deny that there is anything that justice requires us to have equal amounts of. 'Relational egalitarians' (Anderson 1999; Wolff 1998; Scheffler 2003) point out that many of the things condemned and advocated on egalitarian grounds are not matters of distribution: consider prejudice, exclusion and disrespect. These do not concern what we have, but how we relate to one another. As Elizabeth Anderson puts it: egalitarianism's 'proper positive aim is not to ensure that everyone gets what they morally deserve, but to create a community in which people stand in relations of equality to others.' (1999, 288–89) But even relational egalitarians will also be concerned with the distribution of wealth. How wealth is distributed affects the relations in which people stand to one another – consider how relations between workers and bosses (or men and women, or white and black people) would change if they were equally wealthy.

Political movements for equality seem to agree with relational egalitarians that equality means more than can be captured by questions of distribution. The movements for gender and racial equality, for instance, ask for respect, representation and inclusion as much as redistribution. In a word, these movements, and relational egalitarians, oppose *hierarchy*. And 'distributive egalitarians' (those who define equality in terms of some currency) should be concerned about hierarchy too. Hierarchies, just like inequalities of wealth, lead to unequal distributions of opportunity, access to advantage, freedom, capabilities, resources, primary goods and so on. They also tend to lead to inequalities of wealth (causal connections between the two run in both directions).

Thus, the two core concerns uniting the range of contemporary philosophical theories of equality and egalitarian movements in real politics are (1) the distribution of wealth, and (2) hierarchical social relations. Unlike the distribution of well-being, there is no obvious reason why utilitarianism should be silent on such questions. In what follows, I will argue that it is not. It will consistently find reasons to favour more equal social arrangements: those with flatter distributions[85] and less hierarchical relations.

[85] There are, of course, many different ways to measure the equality of a wealth distribution: Gini coefficient, share of income going to the top 10%, etc. Usually, these point in the same direction (that South Africa is less equal than the USA which is less equal than Sweden, for instance; and that almost all European countries were more equal in the middle of the twentieth-century than they are today), and it is intuitive which distributions are 'flatter', so I do not provide a precise definition of what this means (for discussion see Temkin 2001, pt. II; R. Wilkinson and Pickett 2009, 17–18).

2. The distribution of wealth

We might think that utilitarianism is unconcerned with the distribution of wealth – as with well-being, utilitarianism cares about the total rather than the distribution. However, this would be so only if total well-being were directly proportional to total wealth, however it is distributed. Utilitarianism would then hold that the best distribution of wealth is one in which the most wealth exists, and be indifferent between different distributions with the same total wealth levels. However, such proportionality does not hold. Different distributions of the same sum of wealth generate different sums of well-being. Utilitarianism therefore does care about which individuals wealth accrues to – and will prefer more equal distributions of it, other things equal.

It may be true that, typically, the wealthier one is, the happier one is. This relationship seems likely on most plausible theories of well-being. Take a hedonistic theory of well-being, that defines well-being in terms of pleasurable experiences minus painful experiences. Wealth typically increases the prevalence of the former and decreases the latter: certain pleasures need to be bought, including not only consumption goods but also leisure time (the poorer have to work longer hours merely to survive); furthermore, financial security reduces stress. This is one reason that wealthier people tend to have better health in terms of longevity and susceptibility to chronic mental and physical illness; on any plausible theory of well-being, better health makes one better off. Other theories of well-being place weight on the quality of relationships. Compared with pleasurable experiences, this perhaps correlates less strongly with wealth, but poverty certainly makes it harder to have good relationships (as does ill health which we have already linked to lack of wealth). If one cannot afford to go out or take leisure time, or reciprocate in gift-giving, or is ashamed of being worse off than others, one cannot establish and maintain good relationships with others. Other theories of well-being stress the (successful) pursuit of goals and projects. The extent to which wealth aids such pursuits depends on their nature, but it is easy to see how money it would typically do so, and hard to think of cases in which wealth would inhibit the pursuit of goals.

So wealth can be expected to have a positive effect on well-being. But this does not mean that the relationship between wealth and well-being is directly proportional. There are several reasons to think that it is not. Below, I explore two of these reasons: the diminishing marginal utility of wealth and the effects of relative poverty on individuals.

The main finding of this section is that utilitarianism consistently finds reasons to favour more equal distributions of wealth to less equal ones. My claim is not that utilitarians must *always* prefer more equal distributions of wealth to less equal ones, come what may. (As we will see, no plausible egalitarianism will do this.) It is rather that they typically will, other things equal – and often when other things are not equal – because a distribution's being flatter will tend to make it more conducive to well-being in certain ways.

a. Diminishing marginal utility

Something has diminishing marginal utility (DMU) iff an additional unit of it would produce more well-being in the hands of someone with fewer units of it to start with, than it would in the hands of someone with more. If wealth has DMU, then utilitarians will care to whom wealth accrues: they will prefer it to accrue to those who have less of it. Such a preference is shared by all egalitarians worthy of the name.

In fact, for any given sum of total wealth, if wealth has DMU, utilitarianism will prefer a perfectly equal distribution of wealth. Of course, we don't have a given sum of total wealth. The production and distribution of wealth are not independent. It could be that permitting a degree of wealth inequality increases the total wealth of a society, for example, by providing incentives to work hard or take risks.[86] In this case, given that overall well-being increases with total wealth, DMU of wealth would not mean that utilitarianism mandates a perfectly equal distribution. It would, however, mandate a more equal distribution that it would were well-being directly proportional to wealth. DMU of wealth implies that a society in which the rich are less rich and the poor less poor could have greater overall utility, even if it had less overall wealth.

Does wealth have DMU? Sidgwick thought so (1998, 110), and Hume (1965, 44). It is intuitively highly plausible, especially at the extremities of the global wealth distribution. Imagine that the total value of Jeff Bezos's Amazon shares increases by $500. His well-being would hardly, if at all, be altered; he very probably would not even notice. Imagine that a professor receives a $500 bonus. She might fly first class rather than economy to her next conference or stay in a fancier hotel when she arrives. These things would probably make her life a little better. Imagine that a Bangladeshi factory worker receives $500. This would roughly double their annual income. They would be able to put another child through school,

[86] This does not mean that there is, for actually existing societies, a trade-off between reducing inequality and increasing growth, at the margin. Reviewing recent empirical studies, which have shown effects in both directions, Schmidt and Juijn suggest 'that it is rational to assign a higher credence to believing income inequality reduces growth in developed countries than that it boosts it.' (2021, 9) But it may nevertheless be that hypothetical very equal distributions would inhibit growth for the reasons mentioned here. For further discussion see subsection 4e below.

eat more nutritious food, pay off the medical debt that has been haunting them or fix the leak in their roof that has been making them ill. Their life will be made far better.

Harry Frankfurt is sceptical of the DMU of wealth (1987). He writes that:

> 'The argument [from DMU to equal distributions of wealth] presupposes: (a) for each individual the utility of money invariably diminishes at the margin and (b) with respect to money, or with respect to the things money can buy, the utility functions of all individuals are the same… In fact, however, both (a) and (b) are false.' (1987, 25)

Taking (b) first, he notes that different individuals have different needs and capacities, meaning that they derive different amounts of well-being from the same amount of money. This is undoubtedly true, and not something a utilitarian can ignore. Utilitarianism prefers wealth to go to people who will enjoy it more. Thus utilitarianism is unlikely to mandate a perfectly equal distribution. We already saw that it does not, even given DMU, if inequality has some positive effect on wealth production. However, unless the rich systematically enjoy money more than the poor, Frankfurt's observation in no way undermines the notion, that wealth tends to have DMU, and thus that utilitarianism will tend (other things equal) to favour more equal distributions.

Thus (a) is the real crux of Frankfurt's disagreement with the argument from DMU to more equal distributions. Frankfurt hypothesises that there are 'utility thresholds', such that at certain points in the scale of wealth, an extra dollar pushes one over the threshold and thereby increases one's well-being to a greater extent than it would were one lower down the scale (1987, 27). This directly challenges the DMU of wealth, which holds that the lower down the scale one is, the greater the impact on one's well-being an extra dollar has. Frankfurt also notes that the fact that most goods have DMU does not entail that wealth does: goods have DMU, typically, because one gets bored of them or simply has enough of them; money, though, can be used to buy many kinds of good – once one gets bored of one kind, one will buy another kind of which one is not bored.

The notion of a utility threshold is coherent, and it is not obvious that they do not exist. Frankfurt is right, moreover, that the DMU of wealth is not entailed by the DMU of most goods. However, this merely shows that we should not *assume* that wealth has DMU, not that wealth does not have DMU. Frankfurt offers not a falsification, but a challenge: can we show, rather than assume, that wealth has DMU?

Empirical studies typically confirm that wealth has DMU within countries, and throughout the wealth distribution. Wealth (or, as is more typically measured in such studies, income) has a positive but diminishing relationship with 'subjective well-being' –

that is, people's self-reports of their well-being (Layard, Mayraz, and Nickell 2008, 1856; Clark et al. 2018, chap. 2) – and with health outcomes (Mackenbach et al. 2005, 287; Bueno 2011; Mortensen et al. 2016; Ecob and Davey Smith 1999).

Now, a defender of Frankfurt might be sceptical that these empirical studies are to be trusted. Are surveys of reported well-being good guides to well-being? There are good reasons to think that self-reported well-being is at least suggestive of actual well-being. The strongest of these is that well-being (experience machines notwithstanding) should be something noticeable to the agent, and something that typically feels good and strikes us as worth wanting. These give reasonable constraints on accounts of well-being: one could criticise a proposed account on the grounds that it makes well-being something inaccessible, or painful, or unattractive. As Railton puts it:

> 'it does seem to me to capture an important feature of the concept of intrinsic value to say that what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him.' (Railton 2003, 47)

Given this thought, it is likely that people can assess their own well-being with some degree of reliability, if imperfectly. If I notice that I am getting more good feelings in my life, and more things worth wanting, I am (a) likely to be getting an actual increase in well-being, whatever account of well-being is correct, and (b) likely to report an increase in my well-being. Secondly, studies have shown that one's self-reports of well-being correlate strongly with how other people in one's life assess one's well-being.[87] This undermines certain sceptical doubts about the value of self-reported well-being, for instance that it is dependent on our mood at the moment we are asked, or on idiosyncratic definitions of well-being.

But perhaps some lives are such that both those living them and others are systematically mistaken about their well-being. Take the fact that women typically report their well-being at least as positively as men (Blanchflower and Oswald 2004, 1366; Bangham 2019, 13), in spite of the well-documented additional obstacles, reduced opportunities and increased pains that women suffer. One explanation of this is the phenomenon of adaptation. Our judgments about how well our lives are going depend in part on comparison to our judgments about how well we can reasonably expect our lives to go. People who cannot reasonably expect their lives to be much better might judge their lives to be very good, whilst they are no better than

[87] Self-reported well-being is found to be similar to the reports of friends and family members (Sandvik, Diener, and Seidlitz 1993), reports of spouses (Costa and McCrae 1988), reports from clinical experts (Goldings 1954).

lives of people with high expectations who judge their lives more negatively. If men suddenly started facing workplace and street harassment, sexual and domestic violence, erasure of their contributions and responsibility for domestic work at the rates that women do, can we imagine them not reporting this as adversely affecting their well-being? But women have come to expect these things and adapted their expectations accordingly. There be a similar third-party adaptation effect. Others (men and women) come to expect certain things in a woman's life and judge her well-being against those expectations. That women self-report high well-being, and that these reports are corroborated by others, then, need not mean that their lives are better than those of men: it could be rather that they are systematically mistaken to be so (see the literature on 'adaptive preferences', esp. Sen 1995, 259–84; Elster 1983; Nussbaum 2001; Mitchell 2018).

Survey data are not a perfect guide to the relationship between wealth and well-being, then, but do provide some relevant evidence, and give some support to DMU. And the studies showing similarly shaped correlations between wealth and health may be better evidence. Any plausible account of well-being, I assume, implies a positive correlation between well-being and health. Poor health is a source of pain; it endangers relationships; it prevents us from accomplishing goals; it reduces our lifespan. Again, health will not perfectly track well-being – there is more to a good life than good health – but health will be suggestive of well-being. There is good evidence that wealth has diminishing marginal returns to health outcomes, especially at higher levels of wealth. As one study acknowledges, this finding provides 'a powerful argument for income redistribution as a measure to improve average population health… one would expect that at the same level of aggregate income larger income inequalities are related to lower levels of average population health.' (Mackenbach et al. 2005, 287; see also Bueno 2011; Mortensen et al. 2016; Ecob and Davey Smith 1999) Redistribution from rich to poor, if wealth has diminishing returns to health, would improve the health of the poor by more than it would damage the health of the rich. If well-being, as is highly plausible, tracks health, then wealth has DMU and such a redistribution would be a net well-being gain, so utilitarianism would mandate it.

There is good reason, then, to believe that wealth has DMU. Wealth tends to generate more well-being in the hands of the poor. Therefore, for any given sum of societal wealth, utilitarianism will mandate a perfectly equal distribution of wealth, other things being equal. Insofar as inequalities might increase the sum of societal wealth, utilitarianism will recommend deviation from perfect equality in distribution, but it will still mandate more equal distributions than if wealth were directly proportional to well-being.[88]

b. Relative wealth and poverty

In the previous subsection we saw that empirical studies on subjective well-being tend to confirm a positive but diminishing relationship with wealth, treated as absolute wealth. They also tend to confirm a positive relationship with relative wealth – that is, not simply how many dollars you have, but how many dollars you have relative to others in your society (either to the mean, median, or the wealthiest group). The differential importance of relative versus absolute wealth is not clear, though it is clear that relative wealth is important. Blanchflower and Oswald (2004) find that both absolute and relative income have positive diminishing effects on reported well-being; Layard *et al* (2010) find that the whole of the apparent effect of absolute income can be accounted for by the effect of relative income.

Empirical work on health also suggests a connection with relative wealth. There is a body of evidence that most aspects of health follow a ‘social gradient’: that is, they are worse for poorer people than for richer people (Marmot 2015). This may not be, at first glance, surprising – it follows the intuitive thought described above that wealth has a positive relationship with well-being. The interesting question is whether poorer people suffer from these problems due to their absolute or their relative poverty. There are good reasons to think it is the latter.

Firstly, lack of absolute wealth could explain why those with secure jobs and housing, who can afford medical care, education and three meals a day have better health and social outcomes than those without those things. But the social gradient for these problems runs all the way through society – it is not only that the poor suffer them more than the rich, but that the rich suffer more than the super-rich. Absolute wealth cannot plausibly account for differences between individuals who all have enough of it to meet these basic needs. As Michael Marmot puts it ‘Why should educated people with good stable jobs have a higher risk of dropping dead than people with a bit more education or slightly higher-status jobs? Is living in a five-bedroom house with three bathrooms better for your health than “crowding” the spouse and two children into a four-bedroom house with only two?’ (2015, 1–2)

Secondly, the hypothesis that poor health is caused by relative, rather than absolute, poverty explains Wilkinson and Pickett’s findings from cross-national comparisons (2009;

[88] This argument was crucial in motivating egalitarian policies in the mid-twentieth century, as social democratic parties drew on utilitarianism and the economics of the time (B. Jackson 2004).

see also Babones 2008). Amongst richer countries, inequality correlates with health problems better than national wealth: pre-austerity Greece and Spain fared better with respect to them than the UK and USA. Despite being significantly poorer, Greece and Spain had significantly more equal income distributions. So although poor Greeks and Spaniards were poorer, in absolute terms, than poor Britons and Americans, they were better off in relative terms. If health and social problems are caused by relative poverty, this would explain why Greece and Spain did better with respect to them (for a recent study showing a similar correlation in China, see Ting et al. 2022).

So, as with DMU, we have two reasonable if imperfect indicators of well-being (namely, self-reported well-being and health) suggesting that individual well-being is a function of relative wealth. The fact that the relatively poor fare worse than the relatively rich does not, by itself, give utilitarians a reason to favour flatter distributions of wealth. It could be that by closing the gap between the two, one makes the relatively poor better off but the rich worse off to a similar or greater extent, thereby doing nothing positive for the overall sum of well-being. What is needed to provide utilitarian reasons for wealth redistribution is that relative poverty negatively affects overall well-being more than the relative wealth that is its corollary promotes it.

Empirical studies show mixed results with respect to the relationship between income inequality and levels of subjective well-being across countries. Some suggest a positive correlation, some a negative one, and some none at all. One might expect that, if the negative effects of relative poverty on well-being outweighed the positives of relative wealth, this correlation would be negative. This empirical picture is not fatal to my argument, however. Firstly, I do not assume subjective well-being reports to perfectly correlate with actual well-being. I have assumed in some of my arguments that it is a reasonable indicator of it; for instance, the relationship between subjective well-being reports and absolute and relative income was taken as evidence for the relationship between these things and well-being. There is an important difference between this relationship and that in studies of subjective well-being and income inequality, however. The latter studies tend to compare countries; therefore, they are affected by differences in how people in different societies respond to subjective well-being surveys (Weimann, Knabe, and Schöb 2015, 66–67). Some cultures may simply encourage more positive responses to such questions than others – without this indicating that they have higher levels of actual well-being. (Wilkinson and Pickett's 'psychosocial effects' (see 3a) might suggest that more unequal societies might encourage this more, as admitting that your life is going badly indicates low status.) This suggests that in cross-national studies subjective well-being reports are a less good indicator of well-being than they are in within-country studies. Studies using time-series data from within one country (Oishi, Kesebir, and Diener 2011) or comparing cities (Hagerty 2000) or regions within one country (Bangham 2019) suggest the expected negative correlation (see Schröder 2018 for a comparison of between-country and within-country studies; but also Kang and Rhee 2021 for a recent between-country study showing that inequality is bad for subjective well-being). Secondly, as in an example I give below, in poorer countries inequality might indicate rapid economic development, which is likely to have positive well-being effects that outweigh the negatives implied by my arguments in this subsection and the last. This would explain the lack of robust cross-national negative correlations between inequality and subjective well-being.

Moreover, there is some evidence that relative poverty reduces the subjective well-being of the poor more than relative wealth increases the subjective well-being of the rich (Blanchflower and Oswald 2004; Clark, Frijters, and Shields 2008). This is consistent with the finding (Alesina, Di Tella, and MacCulloch 2004; Oishi, Kesebir, and Diener 2011) that the subjective well-being of poorer people is more adversely affected by income inequality (see also Buttrick, Heintzelman, and Oishi 2017). Looking at outcomes other than subjective well-being reports also gives us reasons to think that the adverse effects of relative poverty outweigh any benefits with respect to well-being. Richard Wilkinson and Kate Pickett find that 'almost all problems which are more common at the bottom of the social ladder are more common in more unequal societies' (2009, 18). In less equal societies, there is more relative poverty than in a more equal society: the poor are relatively poorer and the rich relatively richer. The social problems Wilkinson and Pickett discuss – violence, incarceration, low educational attainment, poor mental and physical health, early pregnancy and drug abuse – all plausibly adversely affect well-being, on any reasonable conception of well-being. The fact that they are more common in more unequal societies, that is, societies with higher rates of relative poverty and relative wealth, suggest that the harm done by the former outweighs the good of the latter.

Considerations of relative wealth and poverty open the utilitarian to more strongly egalitarian possibilities than DMU alone does. DMU suggests that redistributing a given amount of wealth from rich to poor will increase overall well-being. DMU also implies that in some cases egalitarian redistribution increases overall well-being even when it decreases the overall sum of wealth. But it could not justify redistribution that made everyone poorer.

However, if wealth has DMU *and* relative poverty matters to well-being in the way I have suggested, then egalitarian redistribution will increase overall well-being in some such cases. Consider the case of an unequal distribution. We make the rich a lot poorer, and the poor a little poorer. Making the rich a lot poorer may be bad for the rich, and making the poor a little poorer is bad for the poor. But making the rich a lot poorer is *good* for the poor, because it increases their relative wealth. By doing so it may also improve the quality of social relations and reduce the prevalence of social problems. These effects could make for an increase in well-being, even though the redistributive policy has made every person poorer in terms of wealth. As Goodin puts it (1995, 250): 'If your stock and use of resources actually impinges upon others' enjoyment of theirs, then it is easy enough to see how restrictions on your stock and use of resources might enhance overall social enjoyment - what you lose through such restrictions may be more than compensated by what others gain through them.'[89]

In this regard, utilitarianism may favour more equal distributions than other principles which are often thought to be more egalitarian. Firstly, take a maximin principle of wealth, which holds that distributions ought to maximise the lot of the worst off.[90] Secondly, take a principle that says wealth always matters, but matters more the poorer the person to whom it accrues, such that there is a positive but diminishing marginal moral value to goods.[91] These principles will always license inequalities that make everyone wealthier; utilitarianism, as we have just seen, will not.

Far from being indifferent to the distribution of wealth, then, DMU and relative poverty mean that utilitarians have reasons to prefer more equal distributions of wealth. Unlike some other theories, it can provide such reasons even where this reduces the total stock of wealth, and where it makes everyone poorer. Of course, utilitarians may sometimes recognise countervailing reasons, and will not always prefer more equal distributions of wealth to less equal ones – for instance, if everyone in the latter is much, much better off. But this is as the judgments of any plausible egalitarianism must be (see 4a).[92]

[89] Goodin's argument follows a similar pattern to mine, though he focuses on the distribution of many types of goods, not just wealth, and the importance of satisfying needs, rather than well-being. In Goodin's terms, this section has argued that wealth is an 'instrumental, competition-utilization resource' (1995, 251), such that overall well-being can be increased not only by giving it to those with less, but also by taking it away from those with more.

[90] This is similar to Rawls's Difference Principle, though that takes 'primary goods' as its currency, rather than wealth (1971). Since primary goods include 'the social bases of self-respect', Rawls's Difference Principle may issue in more equal distributions than the maximin principle mentioned here, for the same psychosocial reasons that utilitarianism does.

[91] This is akin to prioritarianism (Parfit 1991; Buchak 2017), but takes wealth rather than well-being as its currency.

[92] One further reason for utilitarians to favour more equal distributions of wealth has recently received treatment from Andreas Schmidt and Dann Juijn (2021). Recently some utilitarians have embraced 'longtermism': the thesis that the most important determinant of the moral status of our actions is their effects in the far future. There is a straightforward utilitarian case for this: there are lots more

3. Hierarchy

Here I present reasons for utilitarians to prefer social arrangements with less hierarchical social relations. I identify reasons both instrumental – hierarchy causes something that adversely affects well-being, and general – the very nature of hierarchy is such that it will tend to be suboptimal with respect to overall well-being.

a. Instrumental reasons

The classical utilitarians were often brave and leading opponents of traditional hierarchies. They justified this opposition on instrumental grounds. For instance, Bentham argued that the concentration of power in the hands of privileged classes, and the deference which they were shown, led to policy being made in accordance with their class interests, rather than to maximise overall well-being. This phenomenon, which he termed 'the sinister interest' and might today be called 'elite capture', converted Bentham to the cause of universal suffrage and opposition to monarchy, aristocracy, and the established church (Schofield 2006, chaps 5–7).[93] Mill advocated representative democracy on similar grounds, in addition emphasising the improving effect participating in politics had on individuals (2008a). Mill was also an outspoken opponent of hierarchical gender relations. Though he wrote, in a rhetorical flourish, that ''the legal subordination of one sex to the other – is wrong in itself' (2008b, 471), his feminism was largely justified on the grounds that such subordination had non-optimific results: women exposed to abuse in marriage; women's talents going to waste; men having a misplaced sense of superiority and an inability to enjoy the pleasures of interacting with women as equals.

people in the future than in the present, so there is a lot more potential to affect well-being by focusing there (Greaves and MacAskill 2021). There are many complications around this argument, in particular how to weigh future people's well-being and how to evaluate highly uncertain prospects. Schmidt and Juijn find that longtermist utilitarians have good reason to favour more equal distributions of wealth. This is largely because societies with high levels of wealth inequality tend to contribute more to climate change and conflict, and have worse governance: and these things increase the risk of human extinction, which would rob the long-term future of large amounts of well-being. I do not have much to say to elaborate on this argument here, and my arguments do not assume the controversial thesis of longtermism.

[93] Some recent work in comparative political science suggests that Bentham was right: countries with more 'inclusive' institutions do better than those with 'extractive' ones on a range of economic and social indicators (Acemoglu and Robinson 2012).

Recent work in social science gives further instrumental reasons for utilitarians to oppose hierarchy: it tends to reduce the well-being of those subject to it. One route towards this conclusion comes from asking: what could explain the adverse effects on well-being of relative poverty that we noted in the previous section?

Marmot (2015), and Wilkinson and Pickett (2009, chap. 3), appeal to 'psychosocial' mechanisms: relative wealth affects social relations, which affect us psychologically and thus affect our health and our assessments of our own well-being. It is easy to see why relative wealth would affect social relations. This is true from the most intimate to the more structural. It is hard to have close relationships with people who are much wealthier than you. They have different expectations about gifts, places to meet, things to do, and time to do it in. If you are absolutely poor, you may not be able to afford to take the bus into town to see your friends. If you are relatively poor, you may be able to afford the bus, but since your friends can all afford cars you feel ashamed to take it. The point is not that you cannot afford a car – that is a function of absolute wealth. It is that you cannot afford a car and they can – this has to do with relative wealth. Then, because they can afford it, they start taking expensive holidays together. You cannot join them, and miss out on gossip and memories. The friendship withers. The relatively poor are therefore limited in who they can be friends with, and friendships with relatively wealthier people will be more difficult to maintain, and likely of lower quality.

On a more structural scale, relative poverty makes it likely that one will be involved in certain bad socio-economic relations. If some have more wealth than others, then it is possible for relations of exploitation to emerge between them, with the wealthy able to buy up the means of production and the poor having to work for them. Large inequalities of wealth can also translate into domination. In the economic sphere, this comes hand-in-hand with exploitation: the poor worker has no means to resist or influence the commands of their employer. In the political sphere, this occurs when wealthier people can use their money to influence the state to a much greater degree than poor people, thus being able to – through the state – coerce poorer people in a way that they have no meaningful say over. In the symbolic sphere, since money is the primary way our kind of society recognises status, inequalities of wealth will lead to growing differences in, and importance of, status rankings.

So we have various ways in which unequal wealth affects social relations. Relative poverty makes one vulnerable to exclusion, low status, exploitation and domination. Now these are prominent amongst the relations to which 'relational egalitarians' object, the relations I have been calling 'hierarchical'. How could this help explain relative poverty's correlation with ill health and low life satisfaction?

Marmot proposes relevant psychosocial effects of two kinds: stemming from control over one's life, and stemming from participation in society. We can see how these align with the social relations discussed above: exploitation and domination reduce the control one has over one's life – they put control in the hands of bosses, dictators, abusive partners, or even inhuman structures such as markets; exclusion from intimate relationships or from positions that society seems to value reduce quantity and quality of social participation. And we can see how these things could affect health and self-reported happiness.

To take health first, people who lack control over their lives suffer from greater stress, which aside from being a mental health problem in its own right, is correlated with low HDL cortisol levels and thereby with greater risk of coronary heart disease. Stress can be reduced with strong social ties, so exclusion from social participation shares these effects. Studies have found that people with fewer social connections are more susceptible to diseases (S. Cohen et al. 1997) Whilst the causes of serious mental health conditions are less easy to pinpoint, empirical studies and common-sense suggest that the stressed, the pushed around and the lonely are more likely to suffer from them, too.

A further strand of research links stress to status. Dickerson and Kemeny (2004) found in laboratory experiments that large and reliable changes to cortisol level could be induced by exposing people to 'social evaluative threat': the prospect of being judged as failing by others – in particular, when failure and judgment were unavoidable. Status ranking in society is a way in which some people are judged as successful and some as failing, widely and generally, and often unavoidably, throughout their lives. Wilkinson and Pickett offer this as an explanation for the significant increase in anxiety across American society as wealth inequality has increased (Twenge 2000), as well as for their general finding that health and well-being are better in more equal societies. The bigger the differences in status, it is natural to think, the higher the stakes in being judged by others, and thus the greater the social evaluative threat. As they put it:

> 'If inequalities are bigger, so that some people seem to count for almost everything and others for practically nothing, where each one of us is placed becomes more important. Greater inequality is likely to be accompanied by increased status competition and increased status anxiety. It is not simply that where the stakes are higher each of us worries more about where he or she comes. It is also that we are likely to pay more attention to social status in how we assess each other.' (2009, 44; for a more recent restatement see their 2017)

This claim is supported by Paskov et al (2013), who find an association between income inequality and status anxiety, and by recent experimental data (Melita, Willis, and

Rodríguez-Bailón 2021). Stressed people are also, we can suppose, likely to report their well-being as less, because of both stress itself and its adverse impact on other health outcomes. Unemployment shows a strong negative correlation with self-reported happiness (which cannot be fully accounted for by its effect on income). Marriage correlates positively with self-reported happiness, and divorce negatively (Blanchflower and Oswald 2004). Work and marriages being sources of both friendship and status, these findings suggest that exclusion from social participation and respect has an adverse impact on self-reported happiness. This is also an attractive explanation for the effect of relative wealth: the relatively poor occupy a less respected position, and are excluded from certain kinds of social participation, and suffer great social evaluative threat, and therefore report their well-being less positively.[94]

Interestingly, an increase in the rate of unemployment also seems to have an adverse effect on self-reported happiness, even for those who do not lose their jobs – although one would think it puts more people below them in the hierarchy. This suggests firstly that people compare themselves upwards in the hierarchy[95]; also that control over one's life, just as for health, is important to self-reported happiness. When unemployment is higher, all workers are in a weaker bargaining position, so have less power vis-à-vis their bosses. Control would also explain why people are happier under more democratic governments. So this is also likely to be part of the explanation for the effect of relative wealth on self-reported happiness: the relatively wealthier have greater control over their lives.

In a review taking into account many of the studies cited above, and others, Buttrick and Oishi (2017, 1) find that:

> 'Living in highly unequal regimes is associated with both increased mistrust and increased anxiety about social status; these psychological mechanisms help explain some of the negative outcomes associated with income inequality, such as lower

[94] This could also explain the finding (Blanchflower and Oswald 2004) that African-Americans' self-reported happiness has increased over the second half of the twentieth century, as they have been permitted to participate in society to a greater degree, and disrespectful racist attitudes have become somewhat less prevalent, or at least less expressed.

[95] Blanchflower and Oswald (2004, 1378) find that self-reported well-being seems to correlate best with the ratio of one's income to the top quintile of the income distribution, 'consistent with the idea that people compare themselves more with well-off families, so that perhaps they get happier the closer their income comes to that of rich people around them… There is some sign… that individuals do not want to be far above the poorest people, that is, those in the bottom fifth of the income distribution. Although much remains to be understood, it may be that when people make relative-income comparisons they look primarily upward rather than downward.'

> happiness, lower social cohesion, weaker morality, higher mortality, worse health, and weaker governance.'

So it seems that there is social science suggesting that these hierarchical social relations are bad for well-being – thus, utilitarians have instrumental reasons to oppose them.

One might ask with respect to these relations, although they make life worse for those on one end of them, do they not make life similarly better for those on the other end? If so, utilitarians could not condemn them. There are several responses to this. Firstly, as noted above there is empirical research suggesting that relative poverty reduces the subjective well-being of the poor more than relative wealth increases the subjective well-being of the rich. Secondly, for some effects of relative poverty, there will always be more people on the wrong end than there could be on the 'right' one: there are more of us being exploited and dominated than there could be exploiters and dominators, by the nature of these relations, at least in any society at all like ours.

Thirdly, it could be that the dominators and exploiters, whilst they are happier than those they dominate and exploit, are not happier than they could be in a more egalitarian society. The rich are not immune to status anxiety and alienation from others, for instance, whilst in a more egalitarian society these things would be mitigated for everyone. How far this third point can be pushed will depend on one's account of well-being; in particular, how much one believes that alienation from others undermines well-being. Many socialists place great weight on this. As the Glaswegian trade unionist Jimmy Reid said, talking of capitalist elites:

> 'It is easy and tempting to hate such people. However, it is wrong. They are as much products of society and a consequence of that society, human alienation, as the poor drop-out. They are losers. They have lost essential elements of our common humanity. Man is a social being. Real fulfilment for any person lies in service to his fellow men and women.' (1972)

b. A general theory of hierarchy

One may worry that the instrumental reasons for utilitarians to oppose hierarchy are too contingent: is it mere accident that status anxiety reduces well-being? What if further empirical research showed that it did not, or that removing hierarchies brought about other pathologies? I think we can give a general account of why hierarchies – or at least those which egalitarians oppose – will be suboptimal according to utilitarianism, other things equal. There is something in the nature of hierarchies that makes them less than optimific. This general claim is less contingent than the arguments of the previous section.

So far we have mentioned relations such as status, exclusion, domination and exploitation, grouped them under the term 'hierarchy' and defined them extensionally as 'the relations with which relational egalitarians are concerned'. It would be useful to have a general account of hierarchy, one that explains what unifies these relations and allows us to judge whether other relations count as hierarchical. A recent proposal for a such an account runs as follows:

> 'a social position A is hierarchically ordered above social position B just when it is true for the participants in the relevant social context that if they display the socially expected complexes of attitude and behaviour, they thereby and to that extent value the occupants of position A more than the occupants of position B.' (van Wietmarschen 2021, 6)

There is much to recommend this account. It is true that hierarchies characteristically involve expectations – following Bicchieri (2017), 'expectations' in van Wietmarschen's account is meant in both empirical and normative senses – of behaviours and attitudes. For instance, in patriarchies women are expected to defer to men, and men to look down on women. The account also explains why hierarchies are important and persistent: people tend to care about and conform to social expectations. Thirdly, by specifying that hierarchies emerge within specific social contexts, van Wietmarschen's account avoids overproliferation of hierarchies. Compare with accounts on which hierarchies are constituted by mere inequalities of power or esteem (e.g. Kolodny, n.d.). Since, as van Wietmarschen points out, I am less powerful and less esteemed than the President of Chile but more than a Japanese factory worker, such accounts place me in between the two in a hierarchy. But this is an unnatural way of talking: if there is no social sphere of which we are all part, then we are not hierarchically ordered at all. Insofar as egalitarians are concerned about my relations with these two individuals, it will either be on the basis that we in fact do inhabit a shared social context (the global economy, for instance), or that we possess different shares in the distribution of wealth, opportunity, advantage and so on. Lastly, van Wietmarschen's account promises to explain why hierarchies, unlike other sets of social norms, have a valence: it is not just that some are treated differently to others, but that some are placed *above* others. This is accounted for by the fact that occupants of hierarchically ordered social positions are valued more and less than one another.

As well as explaining these general features of hierarchies, van Wietmarschen's account unifies the hierarchical relations we've mentioned so far: exclusion, status, domination and exploitation. People who suffer from exclusion are on the receiving end of social norms that see them disvalued as participants in co-operative activities. The achievements of those with higher status are trumpeted, showing their higher value as agents. The dominated are expected to show deference to their dominators, expressing their higher value as decision-makers. The exploited are expected to give up resources – time, goods, energy – to exploiters, who in turn are valued more highly in the collective enterprise (it is business owners who are said to 'create wealth', not workers; it is *Hadrian's* Wall, not the wall of those legionaries and slaves who spent their energies constructing it).

The unification of these relations is not total, though, and I worry that it is insufficient. I will now suggest an amendment to van Wietmarschen's account that better unifies different hierarchical relations, and in so doing make a case for my claim that utilitarianism will generally oppose hierarchy. 'Valuing', for van Wietmarschen, is valuing in a particular respect. As he puts it:

> 'Persons can be valuable in different ways, so the question naturally arises: which ways of being valuable are relevant for social hierarchy? My view is that social hierarchies can be built around *any* respect in which persons can be more valuable than others. In some social contexts, people are socially expected to treat different people as more or less honourable, pure, or courageous; in other contexts as better or worse tennis players, dog trainers, and so on.' (van Wietmarschen 2021, 7–8 his italics)

This is, I think, problematic. It places individuals in indeterminate positions with respect to the hierarchies they inhabit. Perhaps Constantine[96] is valued less than Warner[97] as a bearer of human rights because of the colour of his skin, but more than him as a cricketer because of his superior technique – both, we are supposing, within the single social context of interwar British society. On van Wietmarschen's account it appears that all we can say is that Constantine is placed above Warner in one hierarchy and below him in another. It nevertheless appears that there is a hierarchy in this society, and Warner is placed above Constantine in it. When a historian writes that in interwar British society aristocratic white men like Warner were at 'the top of the pile', we know what they mean and we know that it is true regardless of how many black men like Constantine were considered better cricketers.

[96] Learie Constantine, a black Trinidadian cricketer and grandchild of slaves who played in England in the interwar period and later became an antiracist campaigner and politician – and in fact a member of the House of Lords, though we are considering his pre-peerage days in this example.

[97] Sir Pelham 'Plum' Warner, also born in Trinidad and a great-grandchild of plantation owners, educated at Rugby School and the University of Oxford, captain of the England cricket team in the early twentieth-century and later President of the Marylebone Cricket Club, the most prestigious role in cricket administration.

Van Wietmarschen has at least three responses available. The first he discusses in his article:

> 'in such cases, the full set of operative social norms in the relevant context do not just require that individuals are valued as occupants of their distinct roles, but also that the occupants of some roles are valued more than others. The full set of social norms does not just require that peasants are appreciated as peasants and lords as lords; it also requires patterns of attitude and behaviour that amount to valuing lords more than peasants.' (2021, 8)

This may work for lords and peasants. But in the case of Constantine and Warner, it is not that social norms encourage valuing bearers of rights more than cricketers. Both Constantine and Warner are (and are acknowledged as) both bearers of rights and cricketers. It is rather that qua bearer of rights, Warner is on top, and qua cricketer, Constantine. The second response van Wietmarschen could make is that we are not talking about a single social context. On the cricket field, hierarchies are based around valuing as a cricketer, and Constantine stands above Warner. In a court of law, hierarchies are based around valuing as a bearer of rights and Warner stands above Constantine. This may well describe those contexts. But both are part of a wider social context (interwar British society) – neglecting this fact will impoverish our understanding of both cricket[98] and courts. In this wider social context, according to van Wietmarschen's account, it seems that Constantine lies both above and below Warner. But it is an uncontroversial fact and important explanandum for any account of hierarchy that in this wider social context Warner is placed above Constantine, and not the other way around.

In conversation van Wietmarschen has suggested a third response: that naming the context as 'interwar British society' makes certain hierarchies (such as those of bearers of rights) more salient than others (such as those of cricket). Perhaps this can be made to work, but we are owed an account of what makes one hierarchy more salient than another in a given social context. Maybe van Wietmarschen can save his account of hierarchy as far as description goes, avoiding unwanted indeterminacy. However, there is still an issue of normative force. Van Wietmarschen is explicit that his account of hierarchy is purely descriptive, and does not aim to explain the badness of hierarchies – indeed, he affirms that some hierarchies may be justified. However, *we* are interested in those social relations that relational egalitarians find problematic. These will probably be a subset of van Wietmarschen's hierarchies. The example at hand shows this. Relational egalitarians are

[98] 'What do they know of cricket, who only cricket know?', as CLR James asks (2005).

unconcerned with some people being considered better cricketers than others, but they are concerned with some people having more rights than others.

I propose an amendment to van Wietmarschen's account that will do what is needed in such cases. Descriptively, it will give us a determinate answer that Constantine stands below Warner in the interwar British social hierarchy. It will also, I think, pick out as hierarchies just those relations with which relational egalitarians are (or should be) concerned – in this case, it makes the race- and class-inflected hierarchy of rights a hierarchy, but not the talent-based ranking of cricketers. (It is therefore less capacious than van Wietmarschen's account; it does not aim to describe non-problematic hierarchies.) It will also deliver the verdict that utilitarianism is generally against hierarchies.

The amendment is that *one's valuing the occupants of some social position more than the occupants of some other causes one to give their well-being more weight in practical deliberation*. This understanding of valuing can then be plugged back into van Wietmarschen's definition of hierarchy. The resulting account states that the hierarchical relations with which egalitarians are concerned are such that if participants in the relevant social context display the socially expected complexes of attitudes and behaviour, they will weigh the well-being of those on one side of the relation more than that of those on the other side. Note that this is a claim about what hierarchical relations cause, not what constitutes them. Hierarchy exists in the social expectations, not in the well-being weights people use. The point is that the social expectations of concern are those that lead to the unequal weighting of well-being.

In the racist society of interwar Britain, social expectations determined that Warner's well-being was given more weight than Constantine's, regardless of how highly Constantine might be thought of as a cricketer. White aristocratic men were valued more than black men, however talented the latter. Now it might be that van Wietmarschen is right that all valuing is valuing in a particular way – Warner and Constantine were valued as bearers of rights, as gentlemen, as cricketers, and so on – but the aggregate effect of all these valuations was to give Warner's well-being more weight in most people's practical deliberation (as expected by social norms). This is why we can say he is valued *more*. This is why we can confidently rank Warner above Constantine in the social hierarchy, despite Constantine being his recognised superior in one particular field.[99]

[99] Would cricket fans not prefer Constantine to be fit and well, even at the expense of Warner? Perhaps. But this would not be as a result of giving weight to Constantine's well-being, but to that of the cricket-watching public. Moreover, the full set of social expectations in society extended far beyond those applicable to cricket fans qua cricket fans.

This account also unifies the relations of exclusion, status, domination and exploitation. When people are excluded from a culture their well-being will be underweighted: they have no voice to make their case; their needs are easily forgotten. Those of lower status have a voice but one that is discounted relative to those of higher status, and the latter occupy more prominent positions and thus their needs are more visible. Those who are dominated are limited in how they can press the case for their interests. The exploited must direct a portion of their energies towards the profit of another, a situation which would be hard to tolerate or justify if norms did not result in the exploiter's well-being appearing to matter more.[100]

One apparent counterexample to my amended version of van Wietmarschen's account is authority hierarchies. In say, a military chain of command who ranks above whom can be surmised from who has the authority to give orders to whom. There are social expectations that lower ranks obey the orders of higher ranks. But when they do so, does this cause them to weight the well-being of those in higher ranks above their own, or those below them? This is not obvious. Yet it is obvious that there is a hierarchy here.

My first response is that very often the norm that A should obey B's orders *does* cause B's well-being to be weighted more highly than A's, even if this is not essential to authority. People who give orders often do – even unconsciously – allow their own interests disproportionate weight in the orders they give. Subordinates, for their part, find it natural to associate the importance of following someone's orders with the importance of that someone's well-being. The effects can be seen in the more pleasant treatment that is typically enjoyed by the higher ranks of authority hierarchies. Now, in some authority hierarchies, considerations of anyone's well-being may be far from most practical deliberation – an army, for instance, might simply be trying to achieve a war aim, and neither higher nor lower ranks are thinking about the well-being of any soldier. Sometimes, even, a general might give orders that undermine his own well-being. Nevertheless, the social expectation that the lower ranks obey the higher ones will, in general, cause the well-being of the latter to receive more weight than that of the former.

Now, there may be 'pure authority hierarchies', in which A obeys B without any implications for the relative weightings of their well-being. These are most plausible where B's orders cover a fairly circumscribed area, in which their own interests are unaffected by the orders they give. An example would be a referee in a football match. Their decisions are

[100] A prominent relational egalitarian, Samuel Scheffler, espouses a similar view, writing that 'a society of equals is characterized by a reciprocal commitment on the part of each member to treat the equally important interests of every other member as exerting equal influence on social decisions.' (2015, 35–36).

almost always complied with, but nobody takes their well-being to be more important than that of the players. Partly this is because whether a referee makes one decision or another has no effect on their well-being (unlike an army officer who can order subordinates to clean his uniform, bring him food, protect him from the enemy and so on). Therefore players need make no association between the importance of the referee's orders and the importance of the referee's well-being (unlike the officer's subordinates who will naturally come to think that if they ought to look after the officer in these ways that nobody looks after them, this is because the officer's well-being more important than theirs). But the authority hierarchy between referee and players is not the kind of hierarchy to which egalitarians would tend to object. My amended version of van Wietmarschen's account does not cover this case, but that is as well, since my aim – unlike van Wietmarschen's – is not to pick out all hierarchies but all hierarchies that are problematic from an egalitarian standpoint. It is therefore a virtue rather than a failing of the account that it leaves out pure authority hierarchies, insofar as these are, like the referee case, unproblematic.

My amended account implies that utilitarianism will generally favour less hierarchical social arrangements. Utilitarianism recommends social arrangements that maximise overall well-being, where the latter is the sum of each individual's well-being, equally weighted. If there are hierarchies, people are expected to weigh the well-being of others unequally – giving more consideration to that of white people, rich people, men and so on. This will tend to lead them to decisions that do not maximise overall well-being, since they are deliberating with incorrect weights. For instance, take a decision between increasing Constantine's well-being by 5 utils or Warner's by 4. If the agent weighs Warner's well-being more heavily than Constantine's they will opt for the latter. But the former would have brought about more overall well-being. The most significant effects of hierarchies will be at the level of public policy, where policies that benefit groups higher in the hierarchy will be preferred to those that benefit groups lower in the hierarchy by a greater amount. As far as the utilitarian is concerned, this would be inefficient – and thus morally impermissible.

This is not to say that utilitarianism requires individuals to always weight each person's well-being equally. It may maximise overall well-being if we each (sometimes, at least) gave more weight to the well-being of ourselves or those close to us. This is because we are better placed to help them, and will do so more effectively without taking effects on more distant others into account. More closely related to hierarchy, it may be that some individuals habitually overweighting the well-being of some oppressed groups would lead to more overall well-being, given that others will underweight their well-being. But we are concerned in this chapter primarily with social arrangements, not individual action. A social arrangement in which individuals are expected to overweight the well-being of some and underweight that of others will be suboptimal, even if the best way for individuals to correct it involves reversing those valences.

So utilitarianism has a very general reason to oppose hierarchies, defined according to the amended van Wietmarschen account. Hierarchies (or at least those that relational egalitarians find problematic), by definition, cause the over- and underweighting of well-being and thus they will tend to miscalculation and suboptimal decision-making. Sometimes hierarchies might be justified, on utilitarian grounds, as the necessary means to a greater good. But always, by their very nature, they will be regrettable.[101, 102]

4. Is utilitarianism a form of egalitarianism?

I have shown that utilitarianism will typically find reasons to oppose inequalities of wealth and hierarchy. This shows it can accommodate the most important egalitarian concerns. Ought it then – contrary to common taxonomies – be considered a form of egalitarianism? I want to argue that it ought. I will do this by reviewing possible criteria for a

[101] This account also explains why the causal connections between hierarchy and unequal wealth distributions can be expected to run in both directions. As we have seen, inequalities in wealth lead to hierarchy. However, it is also the case that hierarchy, insofar as it causes an overvaluation of the rich and undervaluation of the poor, legitimises wealth inequality. As Piketty puts it: 'Every human society must justify its inequalities: unless reasons for them are found, the whole political and social edifice stands in danger of collapse. Every epoch therefore develops a range of contradictory discourses and ideologies for the purpose of legitimising the inequality that already exists or that people believe should exist.' (2020, 1) An unequal distribution would appear optimal, to a mind biased by hierarchy to overvalue the rich, and thus serve this ideological function.

[102] So far I have not tended to make clear whether the inequalities under discussion are within a single state or international. It is one of the most significant facts about the modern world, however, that inequalities between countries are very large, dwarfing – at least in terms of wealth and income – within-country inequality (Milanovic 2015). However, most of the empirical studies cited above take within-country inequality as the independent variable, and suggest that lower within-country inequality is better for well-being. It is a virtue of utilitarianism, I think, that it can vindicate and explain the egalitarian thought that both global and within-country inequality matter. Global inequality is suboptimal, as far as well-being goes, because of DMU and the undervaluing of people in poorer countries leading to inefficient distributions of resources (consider, for example, Covid-19 vaccines). But as long as we primarily interact with and compare ourselves to those within our own countries, domestic inequalities of both wealth and relations will also have a large adverse effect on overall well-being through psychosocial effects.

theory to be considered a form of egalitarianism, and either arguing that they are flawed as criteria or that utilitarianism meets them.

a. Distribution and pluralism

One such criterion has already been mentioned:

(i) Egalitarian theories favour more equal distributions of well-being.

This criterion is flawed: it does not capture our intuitive commitment to equality, or the aims of real-world egalitarian movements, as is shown by WEALTHY MAN. Those aims and commitments, as well as the range of philosophical egalitarian theories, agree that the distribution of wealth is more important than that of well-being. This suggests a second criterion:

(ii) Egalitarian theories favour more equal distributions of wealth.

This criterion may be read in stronger or weaker forms. The stronger form is:

(ii) Egalitarian theories always favour more equal distributions of wealth.*

Utilitarianism falls foul of (ii*), as some more equal distributions of wealth will instantiate less well-being than some less equal ones, for all that has been said above. It could be that permitting a degree of wealth inequality increases the total wealth of a society, for example, by providing incentives for individuals to work hard or take risks. In this case, given that overall well-being tends to increase with total wealth, that increase could in theory be enough to outweigh the negative effects of inequality on well-being mentioned in the previous section, and utilitarians may prefer to permit such inequality. It might also be that some hierarchies are overall beneficial, despite their general tendency to lead to the miscalculation of value.

However, (ii*) is implausible. It is embarrassed by 'levelling down' cases: comparisons of more equal societies in which every individual is poor and miserable with less equal ones in which everyone is much better off. In such cases any plausible theory will favour the latter (at least if we ramp up the misery to sufficient levels). And theories that do so are commonly called 'egalitarian'. Such cases are part of the motivation for Rawls's Difference Principle, often referred to as part of the 'liberal egalitarian' view, endorses inequalities that improve the conditions of the worst-off. Other egalitarian theories are *pluralistic*: they place some moral weight on distributive equality, but endorse other values as well.[103] We saw a slightly different motivation for this in WEALTHY MAN. Relieving the boss's headache does not promote any egalitarian value. But such relief of suffering seems justified. So egalitarians must endorse some non-egalitarian values, on pain of implausibility. Thus, where a more equal society instantiates much less of those other values, such egalitarians may prefer the more unequal option.

A weaker and more plausible reading of (ii) is:

> *(ii**) Egalitarian theories consistently provide reasons to favour more equal distributions of wealth.*

Utilitarianism meets this criterion, as was shown in Section 2. DMU and the effects of relative wealth and poverty mean that egalitarian distribution will tend to increase well-being, other things equal. Thus there are always utilitarian reasons to do so, even if these are sometimes outweighed. And as we just saw, such outweighing is inevitable for plausible egalitarian theories.

It might be thought that the reasons utilitarianism gives in favour of more equal distributions of wealth are the wrong kind of reasons to count as egalitarian. For instance, distributive egalitarians who believe it would be better if every person had the same amount of primary goods, access to advantage, capabilities, and so on will also consistently give reasons for more equal distributions of wealth. But they do this because they place intrinsic value on the distribution of some other good. Utilitarianism places no intrinsic value on the distribution of anything – simply on the sum of well-being. This suggests a criterion that utilitarianism would fail to meet:

> *(iii) Egalitarian theories favour more equal distributions of something intrinsically.*

Temkin seems to suggest something like this criterion (2001). However, as I mentioned above, there are respectable egalitarian theories that do not meet (iii). These are the 'relational egalitarian' theories, which focus on the elimination of hierarchical social relations and endorse equal distribution merely instrumentally, as a means to this elimination. The clearest statement of the relationship between hierarchy and wealth distribution, from a relational egalitarian perspective, is Martin O'Neill's theory, which he calls 'non-intrinsic egalitarianism' (NIE). As the name suggests, it falls foul of (iii). In the

[103] Sometimes, 'levelling down' is held to embarrass even pluralistic egalitarian theories, since they hold that there is *something* good about levelling down, even if it is bad overall. I do not find this objection compelling: it seems correct that there is something good but outweighed in such cases.

next section I will introduce NIE, and subsequently I will show that utilitarianism meets further criteria for a theory's being a form of egalitarianism at least as well as NIE.

b. Non-Intrinsic Egalitarianism

O'Neill's theory is of special interest for two reasons. One is that it acknowledges egalitarianism's connection with the two aspects of wealth distribution and non-hierarchical relations. The second, which may be part of the reason for the first, is that O'Neill is not simply a philosopher, but also someone in close connection with egalitarian (in particular, democratic socialist) political movements. Thus, in keeping with the ideas set out in Section 1 of this chapter, his theory is a good place to look for criteria of a theory's being egalitarian.

For O'Neill, unequal distributions of wealth make for worse states of affairs. This is explained by the consequences that usually arise from inequality. (O'Neill uses 'equality' and its cognates to refer to the distribution of wealth, as I will from here, unless otherwise stated.) These consequences are (2008, 121–23):

(a) the avoidable frustration of basic needs,

(b) status hierarchies,

(c) the domination of one group by another,

(d) weakened self-respect of the poor,

(e) servility and deferential behaviour,

(f) the undermining of healthy fraternal social relations and attitudes.

NIE holds that inequality is bad because it leads to (some subset of) consequences (b)-(f) (2008, 123) – which are examples of the hierarchical social relations to which relational egalitarians are averse. (O'Neill also believes that inequality is bad because it leads to (a), but he does not believe this qua non-intrinsic egalitarian, but rather as an independently plausible and not distinctively egalitarian claim.) If inequality occurred without leading to such consequences, NIE would find no fault with it. NIE, then, endorses equality conditional upon the empirical claim that inequality has these consequences.

Utilitarianism also holds that inequality is bad insofar as it has certain consequences, namely:

(g) less overall well-being than there could otherwise be.

If inequality occurred without leading to (g), utilitarianism would find no fault with it.[104] Utilitarianism's commitment to equal distributions, then, depends on an empirical claim that distributive inequality has this consequence. In Section 2 I showed that this empirical claim is well-founded: inequality will lead to (g), at least other things equal.

So, utilitarianism seems to have the same structure as NIE with regard to unequal distributions of wealth: such inequality is bad when and because it leads to a certain consequence. Furthermore, the consequences are not entirely independent. O'Neill's (b)-(f) are effectively the hierarchical relations with which relational egalitarians are concerned – which are also a large part of the explanation for utilitarianism's condemnation of distributive inequality, since they lead to suboptimal outcomes as shown in Section 3. Therefore, both NIE and utilitarianism explain the badness of distributive inequality at least in part through relational egalitarian concerns.

Despite these similarities, O'Neill argues that utilitarianism is only 'weakly' egalitarian, in contrast to NIE, which is 'strongly' egalitarian (2008, 125). In the next two sections, I will show that his argument for this conclusion fails: utilitarianism is as egalitarian as NIE.

c. Contingency and reliability

Like utilitarianism, NIE endorses distributive equality for the sake of other goals, and conditional upon contingent empirical claims. This does not, according to O'Neill, make it any less strongly egalitarian. He writes:

> 'the connection between distributive egalitarianism and these broader egalitarian goals and values [i.e. (b)-(f)], although in some sense contingent, is not a weak one. If this connection were weak, then we might... question whether Non-Intrinsic egalitarian views are strongly egalitarian at all. Yet it seems plausible to think that it is a deep social fact that we can realize the values embedded in the egalitarian considerations (b)-(f) only where substantial inequalities of condition have been eliminated... If this "deep social fact" really does obtain, then Non-Intrinsic egalitarianism will reliably mandate the elimination of inequalities of condition...' (2008, 131)

[104] On some approaches to measuring inequality (e.g. Atkinson 1970), it is analytically true that more unequal distributions lead to less 'social welfare'. But the concept of social welfare in such approaches is not equivalent to the utilitarian maximand of 'overall well-being' (though there is a social welfare function that is equivalent to utilitarianism) – social welfare is not simply an aggregate of each individual's well-being, but encodes social values (such as an aversion to inequality) in addition.

Utilitarianism, according to O'Neill, is only weakly egalitarian. He seems to be appealing to the following criterion for a theory's being (strongly) egalitarian:

> *(iv) Where egalitarian theories favour more equal distributions contingently, they favour them reliably, in virtue of deep social facts.*

If NIE meets (iv) and utilitarianism does not, then the connection between distributive inequality and (g) must be less reliable than that between distributive inequality and (b)-(f), and the social fact that equality correlates with overall well-being must be less 'deep' than the fact that it correlates with improvements with respect to (b)-(f). Is this true? O'Neill explains the depth of his deep social fact thus:

> 'firstly, reductions in inequality almost always bring about improvements in states of affairs of the sort favoured under considerations (b)-(f); secondly, such improvements are generally possible only where inequalities are reduced, and greater distributive equality is achieved.' (2008, 131)

Now, if O'Neill's claims here are true, there is nothing to stop a utilitarian agreeing with them. Furthermore, since (b)-(f) overlap to a large degree with those hierarchical social relations whose psychosocial effects are inimical to well-being (see Section 3), improving states of affairs with respect to (b)-(f) is one way to improve them with respect to (g). The utilitarian can then draw this conclusion from O'Neill's claims: there is a way in which things can be improved with respect to (g) (the increase of well-being via improvements with respect to (b)-(f)) which: (i) almost always follows from the reduction of inequality; (ii) is generally possible only where inequalities are reduced.

It might seem that NIE is nevertheless more egalitarian than utilitarianism. For NIE's concerns seem to track distributive inequalities more closely: NIE does not merely say that there is *a* way to improve things that necessitates reducing inequalities, but that this is *the* way to improve things, since well-being could be increased in many other ways but the values encoded in (b)-(f) could not. But this appearance neglects the fact that NIE is a pluralistic theory. As O'Neill recognises (2008, 143–44), it is not plausible to hold that only (b)-(f) matter. NIE, fully spelled out, will have to admit that it is good to (for instance) cure a rich man's headache, although this has no effect on (b)-(f). Thus NIE's adherents also think that there is a way in which things can be improved (improvements with respect to (b)-(f)) which usually follows from and is generally only possible where inequalities are reduced. For both them and utilitarians this is just one way amongst others in which things can be improved.

In fact, I do not think O'Neill's claim that only the reduction of distributive inequality can improve things with respect to (b)-(f) is true. There is very likely a strong correlation between distributive inequality and (b)-(f). But we can conceive of ways in which improvements could be made with respect to (b)-(f) without the distribution of wealth being touched. Servility and deference could be reduced by changes to norms and structures which could occur whilst wealth distribution is unchanged. O'Neill himself writes, in a review of their 2009 book *The Spirit Level*, that Wilkinson and Pickett's

> 'emphasis… is perhaps excessively on the way in which these policies could lead to greater income equality through compressing wage differentials. An alternative approach that de-emphasized *income* inequalities to some degree could give more emphasis to the ways in which various forms of economic democracy can act in more direct ways to bolster individual status and self-respect, and to transform the character of social relations at work, through granting more power to individual workers. Thus… their approach prevents them from seeing the full range of egalitarian arguments that might be deployed in favour of such policies.' (M. O'Neill 2010, 405– his italics)

The implication is that measures such as workplace democracy could improve things with respect to egalitarian criteria, though they may not involve any redistribution of wealth or income. But then, he should not worry that the possibility of such cases would undermine the claim of NIE to be a kind of egalitarianism. It seems that his view is that egalitarians should want to eliminate or reduce (b)-(f), and this gives them reason to eliminate or reduce distributive inequalities. The fact that it could give them reason to promote non-redistributive policies, such as workplace democracy, does not threaten the egalitarianism of the view. Utilitarianism has the same attitude to such cases. It can endorse improvements with respect to (b)-(f), which will almost always lead to increases in overall well-being, whether they are associated with changes in the distribution of wealth or not. If this doesn't threaten the egalitarian credentials of O'Neill's view, then it doesn't threaten the egalitarian credentials of utilitarianism either.

These cases aside, it seems plausible that one good way of improving things with respect to (b)-(f) will be closing inequalities of wealth. This is the grounds of O'Neill's claim that NIE will 'reliably mandate' such policies. Will utilitarianism be less reliable on this score?

There will be circumstances in which utilitarianism does not endorse closing inequalities even when doing so would lead to improvements with respect to (b)-(f). Utilitarianism would fail to endorse such policies in cases where overall well-being decreases as equality increases (that is, 'levelling down' cases). We might think that cases of improvements with respect to (b)-(f) without improvements with respect to overall well-being are rare. Anyway, in such cases, inequalities typically lead to other values which pluralist non-intrinsic egalitarians ought to be concerned with.[105] Take, for example, a country that is rapidly industrialising its economy. It may be necessary for this transformation that wealth is unequally distributed. If this transformation satisfies the utilitarian, it does not lead to (g) – that is, it produces the greatest possible sum of well-being. If the pluralist non-intrinsic egalitarian values the sum of well-being as well as (b)-(f), they may ultimately endorse this inequality. And even if they don't value the sum of well-being, they could still endorse it. For if the greatest possible sum of well-being is produced, this is because other values are met, such as: suffering is reduced; and/or people enjoy greater freedom; and/or new technologies are invented; and/or needs are better satisfied, and so on. A plausible pluralistic NIE will value at least some of these things alongside concern for (b)-(f). In that case, if the value of these things is great enough, non-intrinsic egalitarians could endorse inequality in this case – even whilst continuing to believe that (b)-(f) are bad. To know how reliably a pluralistic non-intrinsic egalitarian will mandate reducing inequalities, one must know their other values, and their method of resolving conflicts between them. Without knowing this, we can just say that both utilitarians and non-intrinsic egalitarians will agree that reducing inequality is often (but not always) all-things-considered good, and that this is largely because of the considerations about (b)-(f).

There are also circumstances in which utilitarianism will mandate the closing of inequalities and NIE will not. This is obvious for some versions of pluralistic NIE: consider a view that favoured distributive equality on the grounds of (b)-(f), but also respected Nozickean entitlements to private property (2013), making redistributive taxation impermissible. In a society with large inequalities and in which most resources were held privately, this form of pluralist NIE would not mandate very much reduction of distributive inequalities; utilitarianism is likely to mandate more. The general point, again, is that there is no reason to think that utilitarianism will be less reliable in mandating equality than a pluralistic egalitarian theory, and all egalitarian theories must be pluralistic to be plausible.

Even if NIE were not pluralistic, there are cases in which utilitarianism would mandate reductions in inequality that such 'pure' NIE would not. Consider a version of the Divided World case introduced by Parfit (1991, 6). In this version of the case, there are two groups of people who have no interactions and are unaware of each other's existence, and one group is

[105] Avoiding suggesting counterintuitively egalitarian policies in levelling down cases is a major motivation for O'Neill's claim that egalitarians should be pluralists (2008, 143–44).

much wealthier than the other.[106] As O'Neill says, 'none of the egalitarian reasons covered by (b)–(f) can obtain' in such cases, so pure NIE does not mandate closing the inequality (2008, 136). But utilitarianism could. Given diminishing marginal utility, (g) would probably obtain: if the better off group had less and the worse off group more, overall well-being could be greater.

It seems, then, that the social fact linking distributive inequality to overall well-being is no less deep than that linking it to O'Neill's concerns for (b)-(f). Furthermore, whilst there could be cases in which NIE mandates more equal distributions than utilitarianism, they may be rare, particularly if NIE takes a plausible pluralistic form. There are also cases in which utilitarianism advocates more equal distributions than pluralist and pure NIE.

d. Distinctiveness and intrinsicness

O'Neill claims that concern for consequences (b)-(f) is 'distinctively egalitarian' (2008, 124–26). What makes these concerns distinctively egalitarian, for O'Neill, is their connection with life in a 'society of equals'. If (b)-(f) arose in such a society, it would cease to be a society of equals. In contrast, if every individual there became afflicted by chronic pain, then it would be worse with respect to (g), but no less a society of equals. So, (g) is not a distinctively egalitarian concern. O'Neill may be interpreted as suggesting the following criterion:

> *(v) Where egalitarian theories favour more equal distributions contingently, they favour non-hierarchical social relations.*

Does this provide a reason to judge that utilitarianism is not a form of egalitarianism? No, because utilitarians also have concerns for (b)-(f), and so do favour non-hierarchical social relations. As argued above, improvements with respect to (b)-(f) will tend to lead to increases in overall well-being. So, utilitarianism *is* concerned with (b)-(f), and if these concerns are distinctively egalitarian, then utilitarianism shares distinctively egalitarian concerns. For sure, utilitarians care about things other than (b)-(f) – for example chronic pain, existential risk, and overall well-being itself (i.e. (g)) – that are not distinctively egalitarian. But as O'Neill believes that egalitarians should be pluralists, he does not think that having non-egalitarian concerns compromises one's egalitarianism.

[106] In Parfit's original case, the inequality is of well-being rather than wealth. Utilitarianism will not mandate closing such an inequality, since doing so would alter only the distribution and not the sum of well-being. But we have already seen that the distribution of well-being is not a necessary concern for egalitarians. Moreover, NIE will not mandate closing this inequality either: since there is no interaction between the groups, none of (b)-(f) can follow from it.

Perhaps, however, O'Neill would argue not that utilitarianism does not share these distinctively egalitarian concerns, but rather that it does not have these concerns in a distinctively egalitarian way. O'Neill – despite the name he gives his theory – believes in the *intrinsic* badness of consequences (b)-(f) (2008, 130). Utilitarians do not: they believe that, like all consequences, (b)-(f) are bad only insofar as they lead to less overall well-being than there could otherwise be. So O'Neill may have in mind something more like this criterion:

> *(vi) Where egalitarian theories favour more equal distributions contingently, they regard hierarchical social relations as intrinsically bad.*

Utilitarianism fails to meet this criterion, since it opposes hierarchy only as a means to maximising well-being. (As we saw in Section 3, it may be very generally true that it is such a means.) I want to claim that if utilitarianism fails to meet (vi), this is unimportant. The main significance of considering some consequences to be intrinsically bad, rather than bad because they fail to maximise well-being, is that one can condemn states of affairs that instantiate those consequences even when they maximise well-being. For the difference between NIE and utilitarianism to be important, then, there must be cases in which inequalities lead to (b)-(f) but not to (g). These are the cases such as the industrialising country mentioned above. But, as we saw there, given their pluralism, NIE is not necessarily committed to more egalitarian policies here than utilitarianism – and there are also cases in which utilitarians will condemn inequalities that non-intrinsic egalitarians, pluralistic or pure, will not.

One thing that might be said for the stronger egalitarian credentials of NIE is that in cases of levelling down, even where both utilitarianism and (pluralistic) NIE give the same answer, the latter always places some disvalue on (b)-(f).[107] Thus, O'Neill can hold that it is always regrettable that (b)-(f) occur, even in levelling down cases where they ultimately, on the grounds of other values, endorse states of affairs where (b)-(f) occur. It might be claimed that utilitarians, in contrast, could not regret the occurrence of (b)-(f) when they form part of the state of affairs that maximises well-being. Surely utilitarianism cares only about maximising well-being – how could utilitarians see any part of an optimific state of affairs as

[107] The claim that pluralistic NIE always places disvalue on (b)-(f) depends on how the pluralistic structure of the view works. Pluralism need not involve a set of commensurable pro tanto reasons that are weighed against one another. Perhaps some values undercut others, so that, for example, a pluralistic view that is concerned with (b)-(f) and with property rights places no disvalue on (b)-(f) insofar as they are unavoidable consequences of such rights. But let's grant that concern for (b)-(f) is not undercut like this in pluralistic NIE.

regrettable? To put it in other words, it appears that NIE meets the following criterion, and utilitarianism does not:

> *(vii) Where egalitarian theories favour more equal distributions contingently, they always find the presence of hierarchical relations regrettable.*

But utilitarianism can, I think, meet (vii). The claim that utilitarians cannot regret any aspect of an optimific state of affairs would, if true, cause a problem for utilitarianism across a wide range of cases. Consider the trolley problem (Foot 1967). Utilitarianism (with common sense) suggests switching the trolley to save five lives at the cost of one. The death of the one is surely seen as something to regret – as is typically expressed by utilitarians talking of it as 'a cost', or something to be 'traded off'. But the death of the one is part of the outcome that maximises well-being. So, if utilitarians cannot regret any aspect of an optimific state of affairs, they cannot regret the death of the one.

I propose that the utilitarian can respond to this worry as follows. One ought to bring about the inequality and the death of the one, in the circumstances described, since doing so would maximise well-being. However, it would be better if the circumstances were not as described. It would be better, in particular, if one could reap the rewards of industrialisation (or save the five) without increasing inequalities (or killing the one). Utilitarianism implies that these circumstances would be better than those in the original cases because more well-being could be realised in them: other things equal, saving five people without killing the one or industrialising an economy without increasing inequalities would increase well-being more than saving the five and killing the one or industrialising and increasing inequalities. This is what makes killing the one or increasing inequalities 'costs': they are costs relative to what we could realise in better circumstances.

This account also has the virtue of encapsulating a common aspect of regret in cases where one does what one ought: the thought that 'it would have been better if I hadn't had to do that'. In fact, non-utilitarians are likely to share exactly this thought when switching the trolley or (as a follower of NIE would) pursuing inegalitarian policies in the levelling down cases described above. It is hardly controversial that it would be better if nobody had to be hit by a runaway trolley, and if economic development and egalitarian social relations were always mutually compatible. This shared intuition, and thus this kind of regret can be explained, I have suggested, with the resources of utilitarianism: the reason that it would be better if I hadn't had to do that is that in such circumstances, overall well-being would be greater. So utilitarians can regret aspects of what, in their circumstances, is the optimific state of affairs.

The considerations of the past two subsections show that there is no important difference between NIE and utilitarianism with respect to their egalitarian credentials. Both can have distinctively egalitarian concerns (b)-(f). Both rely on contingent empirical claims to mandate the reduction of inequalities of wealth, and where those claims go through for O'Neill they are highly likely to go through for the utilitarian. Both will, sometimes, endorse inequality that produces (b)-(f), and both will see this consequence as a cost. The times at which they do so might differ, but there is no guarantee that NIE, given its pluralism, will endorse inequality less often than utilitarianism does. The fact that one and not the other theory values (b)-(f) intrinsically, then, does not seem to issue in important differences. Both are equally forms of egalitarianism.

Furthermore, there may be advantages, from the egalitarian point of view, to not valuing egalitarian concerns intrinsically. In particular, if one values them because of their effects on well-being, one has a principled way of establishing the degree to which they are valuable: they are valuable to the degree to which they affect overall well-being. O'Neill writes that 'Any plausible view of the value of equality, which acknowledges the force of considerations (b)–(f), will also acknowledge the great significance of those considerations' (2008, 144). But we might want an argument as to why these considerations have such force. This is especially pressing for a pluralistic NIE, since it acknowledges other concerns which might conflict with egalitarian ones. If he cannot show that (b)-(f) have significant force relative to those other considerations, then we might doubt that O'Neill has given an argument for egalitarianism at all (Brown 2014). Utilitarians can provide a rationale for the relative significance of egalitarian concerns: they can appeal to the arguments of sections 2 and 3 which showed that equality has large effects on well-being. Therefore, the utilitarian can say, since overall well-being is the measure of moral significance, equality is morally significant. The utilitarian can also work out how much weight these concerns have relative to others in cases of conflict – by assessing the impact on well-being of those other concerns and comparing it with the impact on well-being of egalitarian concerns.

e. Marx's view

I have argued that, according to utilitarianism there are *pro tanto*, overridable reasons, that is, to favour more egalitarian social arrangements. This implies that (a) inequalities may sometimes be justified to produce better outcomes; (b) when this is so, it is a cause for regret, in the sense that it would be better if those outcomes could be produced without inequality. Karl Marx agreed. In his 1875 *Critique of the Gotha Programme* (2000a, 614–15), perhaps the clearest positive proposal for future society he produced, he contrasted 'a communist society… just as it emerges from capitalist society' (sometimes called 'low communism') with 'a higher phase of communist society'. In low communism, there would be inequality – seemingly in order to incentivise work and to enhance the legitimacy of the system. As he puts it:

> 'the individual producer receives back from society – after the deductions [for public goods, investment, insurance, pensions and so on] have been made – exactly what he gives to it… He receives a certificate from society that he has furnished such and such an amount of labour (after deducting his labour for the common funds), and with this certificate he draws from the social stock of means of consumption the same amount of labour[108].' (2000a, 614)

By connecting each individual's possibilities for consumption to the labour they contribute, such a system would incentivise work. It would also accord with the moral views of a society that has recently been a capitalist one:

> 'the same principle prevails as in the exchange of commodity-equivalents: a given amount of labour in one form is exchanged for an equal amount of labour in another form. Hence, equal right here is still in principle – bourgeois right…' (2000a, 614)

Thus, the low communist system would not offend capitalistic sensibilities too much.

However, Marx believed low communism would be imperfect, because of the inequality it would involve:

> 'But one man is superior to another physically or mentally and so supplies more labour in the same time, or can labour for a longer time… Further, one worker is married,[109] another not; one has more children than another, and so on and so forth. Thus, with an equal performance of labour, and hence and equal share in the social consumption fund, one will be richer than another…' (2000a, 615)

After low communism has expanded production and done away with class division, Marx believed, a better possibility would become available.

[108] Marx means by the latter instance of 'labour' here consumer products produced by a certain amount of labour (which his economic theory holds to be equivalent). The controversial labour theory of value is of no importance to the argument as a whole.

[109] This line gives away Marx's uncharacteristically conservative views about gender (MacKinnon 1989, chap. 1): if both men and women work equally, why should it make a difference to consumption needs that some workers are married? However, the general point, that due to caring responsibilities some need more resources than others to achieve the same living standards is surely correct.

> ‘In a higher phase of communist society, after the enslaving subordination of the individual to the division of labour, and therewith also the antithesis between mental and physical labour, has vanished; after labour has become not only a means of life but life’s prime want; after the productive forces have also increased with the all-round development of the individual, and all the springs of co-operative wealth flow more abundantly – only then can the narrow horizon of bourgeois right be crossed in its entirety and society inscribe on its banners: from each according to his ability, to each according to his need!’ (2000a, 615)

In high communism, then, people’s attitudes towards work and their fellow citizens will have altered such that there is no need to incentivise work, or to legitimise the disbursement of funds to people by reference to their contribution. We can then distribute according to need – and a more equal distribution of wealth will occur. Yes, unequal needs mean that some will get more of the social pot than others, but since they will consume this in meeting their needs none would end up richer than another, as they would in low communism.

In low communism we couldn’t have had the desirable outcomes of economic expansion and political legitimacy without licensing some inequality of wealth. In high communism, we get at least as good economic and political results without wealth inequality. Since Marx believed that low communism was a better social form in some circumstances, but that high communism, if it were possible, would be better still, it seems that he agrees with the utilitarian on points (a) and (b). Inequalities are sometimes justified as necessary means to other values, but it would be better if they weren’t necessary.

Moreover, the abilities/needs principle that Marx endorses for high communism is one that utilitarians should also inscribe on their banners in similar circumstances. Utilitarians want to maximise well-being. An important part of an individual’s well-being, on any plausible account, is their material consumption, and the most important part of that (as the phenomenon of DMU suggests) is that their basic material needs are met. So, to a decent approximation, maximising well-being will involve maximising the meeting of needs. For a given social product, distributing it according to needs looks like a good utilitarian strategy.

Now, as is often pointed out – by Rawls, by Marx, by Hume (1965, 44–45) and by right-wing economists (Wanniski 1978) – we don’t have a given social product: distributive policies can affect the ‘supply side’ of the economy, and thus the size of the product we have to distribute. The most significant mechanism for this is incentives: if the proceeds of your labour will be distributed away from you, and your needs met unconditionally, why work? This being the case, distribution according to needs may be suboptimal for meeting needs.

Marx's low communism, like Rawls's Difference Principle, permits less equal forms of distribution in order to incentivise the expansion of production.

However, as G. A. Cohen points out (2008, pt. I), these incentives are necessary only because people have a certain pattern of motivation, roughly: self-interested preferences for as much consumption and as little work as possible. The important feature of Marx's high communism is that people no longer have such preferences. They enjoy working (because the alienation present in work under capitalism has been diminished) and they are not self-interested in respect of consumption (because wealth is abundant, and so non-rivalrous, and because class divisions have been abolished allowing them to respect others as equally deserving). Thus in high communism inequality is not necessary to incentivise work. People will contribute what their abilities allow them to, that is, as much as they can.

Why should a utilitarian prefer high communism, with its abilities/needs principle? Because needs would be best met in such a society. The social product would be optimised, because everyone would contribute to it as far as their abilities could help them. Perhaps such optimal contributions could be encouraged through incentives. But if it could be encouraged without incentives, the social product could then be distributed in the way (according to needs) that, as we saw, best meets needs and thus probably well-being for a given social product. Thus this is the distributive policy that maximises utility. It is no accident then, that the arch-utilitarian Mill endorses a similar principle: 'that all should work according to their capacity, and receive according to their wants' (1965, 203).[110] Socialists have typically defended the Abilities/Needs Principle on the grounds that it embodies the values of solidarity, fairness, sensitivity to difference and meaningful work (Gilabert 2015). I suggest that it may also be defended on utilitarian grounds.

Now, I am not saying that utilitarians ought to join the communist movement. It may be that high communism is impossible – humans will always be too self-interested and work-shy. It may be that it is possible but the costs of bringing it about are too high. All I am claiming is that a society abiding by the abilities/needs principle would be an optimal social

[110] Mill also emphasises that for such a distribution to be optimal, the social ethos would have to change in a more solidaristic direction. He writes: 'History bear witness to the success with which large bodies of human beings may be trained to feel the public interest their own. And no soil could be more favourable to the growth of such a feeling, than a Communist association, since all the ambition, and the bodily and mental activity, which are now exerted in the pursuit of separate and self-regarding interests, would require another sphere of employment, and would naturally find it in the pursuit of the general benefit of the community... A contest, who can do the most for the common good, is not the kind of competition which Socialists repudiate.' (Mill 1965, 205; for a recent assessment of Mill's socialistic inclinations, see McCabe 2021).

arrangement, with respect to distribution at least. Thus, utilitarians should join Marx in regretting that it is not possible in the near term.

5. Conclusion

Utilitarianism has often been criticised for its neglect of egalitarian concerns. I have shown that utilitarians will consistently find reasons to favour what egalitarians (ought to) care about: more equal distributions of wealth (Section 2), and less hierarchical social relations (Section 3). It is true that these reasons are somewhat contingent, instrumental, and sometimes outweighed. But as I have argued in Section 4, this does not make utilitarianism any less an egalitarian theory than some leading contemporary philosophical theories typically thought of as egalitarian. It is also consonant with Marx's views on distribution and equality.

In making this case I have emphasised, once again, the importance of social relations to individual well-being. How we interact with others affects our well-being, our health, and our capacity to affect the very long-term future. Thus utilitarians have good reasons to give concerns that have traditionally been the preserve of socialists and egalitarians further consideration.

Chapter Eight: Conclusions

0. Abstract

In this concluding chapter I summarise the thesis and suggest directions for future work.

1. Summary of the thesis

Utilitarianism currently has a low stock amongst philosophers, despite its prominence. That prominence is, I believe, warranted, for utilitarianism is an attractively well-motivated, simple and general theory, which offers a standard of moral assessment significantly distinct from common sense to issue in radical and interesting conclusions. Its stock amongst socialists is if anything lower – and it is prominent largely as a bourgeois ideology attacked, in passing, by Marx. This thesis provides considerations that should raise utilitarianism's stock amongst both groups. A thread running through it has been that utilitarianism is more defensible when we acknowledge insights emphasised by the socialist tradition; in particular, the socially connected nature of persons.

The primary aim of the thesis was to respond to Bernard Williams's objections to utilitarianism. His Integrity Objection has been very influential and very variously interpreted. In Chapter Two I presented a new exegesis of it and compared this with others in the literature. In Chapter Three I examined one common response to it, and found that it could give at most a partial response. In Chapter Four and Chapter Five I gave my full response to the Integrity Objection; a response that emphasised that individuals are mutually interdependent, that they identify with communities and shared properties, and that alienation from others is a serious problem. In Chapter Six I investigated another problem for utilitarianism drawn from Williams's cases, and showed how a revised utilitarianism, which applies utilitarian obligations to groups, could provide a solution. Williams also alleges that utilitarianism is at odds with egalitarianism, a criticism that has become common sense in philosophy and accounts in part for socialist suspicions of the theory. I argued in Chapter Seven that this allegation is false.

To summarise my key conclusions:

1. The core of Williams's Integrity Objection is that agents who accept utilitarianism cannot be properly committed to their projects, because they must regard them impartially and as dispensable.
2. Utilitarianism is probably not self-effacing for individuals, but may be for groups.
3. A moral theory ought not be dismissed on the grounds that it requires agents to regard their projects as dispensable; in fact, they probably should.

4. The impartiality that utilitarianism requires of agents need not alienate them from themselves.
5. Utilitarianism must be revised to account for pre-emption cases; one attractive revision is what I have called ‘collectivist utilitarianism’, in which utilitarianism is applied to groups as well as to individuals.
6. Utilitarianism will tend to favour more egalitarian distributions of wealth and less hierarchical social relations.
7. Utilitarianism ought to be considered a form of egalitarianism.

As a by-product of arguing for these conclusions, the thesis has expounded a possible kind of utilitarianism we might call ‘left-utilitarianism’. Building on a classical act-utilitarian picture, it is concerned with avoiding alienation from others; it emphasises our social connections with others rather than the point of view of the universe as the root of morality; it is applied to group agents and, drawing on social science consistently finds reasons to oppose unequal distributions of wealth and hierarchical social relations.

2. Directions for future work

There are various questions raised by this thesis that deserve further treatment. To close, let me mention some of them.

a. Effective altruism

Many contemporary utilitarians are associated with the effective altruist movement. Effective altruists are not all utilitarians (though many are), but they typically endorse principles like Rescue, and possibly Responsibility, as well as trying to take an impartial approach to their work and charitable giving. Therefore, effective altruism probably does preclude commitment, in Williams’s sense. My claim that this is not so problematic as Williams thought offers support to effective altruism.

However, it might also encourage effective altruists to reassess some of their rhetoric. They often make their arguments in individualistic terms, and have been accused, not unreasonably, of being ‘comfortable with ways of talking that are familiar from the exponents of global capitalism’ and thus ‘speak[ing] in the proprietary language of the illness – global inequality – whose symptoms [they] propose to mop up’ (Srinivasan 2015). If my arguments in chapters Four and Five are convincing, effective altruists might better defend themselves by emphasising our mutual interdependence, and being more comfortable with ways of talking that are familiar from the socialist tradition. To what extent would this be effective for the movement, in comparison to other ways of dealing with concerns about integrity? Can other effective altruist positions be defended in this way?

The discussion in Chapter Six might have more substantive implications for effective altruism. The movement tends to encourage people to think 'on the margin', sticking to the standard act-utilitarian line that only the difference individuals make matters. They are alive to the phenomenon of pre-emption, which effective altruist career consultants often call 'replaceability', and it has been discussed in the case of career choice (if I go to medical school, I'm just taking the place of the best applicant who'll miss out, not increasing the supply of doctors (Macaskill 2015; Todd 2012)) and charitable giving (some charities have billionaires who will fund any shortfall they have, such that individual donations to them are merely pre-emptive (Budolfson and Spears 2019)). However, they tend to recommend the standard act-utilitarian behaviour akin to George taking the CBW job, which, as we saw, is inadequate on its own at capturing the concern utilitarians should have for well-being.

If effective altruists were to adopt collectivist utilitarianism, how would their recommendations change?[111] Their recommendations to individual agents, in terms of ethical oughts, should not. But perhaps they should shift their focus towards recommendations for groups – in the knowledge that these may not always cohere with those for individuals. And if effective altruists began to take groups more seriously as moral agents, they may be more ambitious, since groups can do things that individuals cannot. Effective altruism has been criticised from the left for placing a low priority on structural change (Herzog 2016; Syme 2019); this is an understandable implication of focusing on individual actions, given that these (for reasons of pre-emption or otherwise) are likely to be ineffective at bringing such change about. Groups such as classes and states, on the other hand, *can* bring about structural change – which is why socialists are interested in them. An effective altruism aimed at groups may look much more like a political movement – and, as I argued in Chapter Seven, insofar as it is a utilitarian political movement it will be one with egalitarian concerns. What other concerns might it have, and how would it negotiate the contrasting demands of giving individual and group-level recommendations without appearing inconsistent?

[111] For others asking similar questions, see Collins (2019a) and Dietz (2019).

b. Longtermism

Recently many effective altruists have embraced ‘longtermism’: the thesis that the most important determinant of the moral status of our actions is their effects in the far future. There is a straightforward utilitarian case for this: there are lots more people in the future than in the present, so there is a lot more potential to affect well-being by focusing there (Greaves and MacAskill 2021). One of the most pressing challenges to longtermism is that the effects of individual actions on the far future are likely to ‘wash out’ – what I do today will have various effects on the world from now until the end of time, and some will be good and some bad, and we have little way of assessing these. Another kind of washing out stems from pre-emption: even if I do something that has huge effects on the far future, like invent a cure for cancer such that nobody dies from it ever again, it may well be the case that if I hadn’t done it, someone else in a similar lab would have done a few years later, so that my acting only made a difference in the short term. If I’m right about the prevalence of pre-emption, should this reduce our credence in longtermism?

If collectivist utilitarianism is correct, then longtermism might be plausible for groups even if it is not for individuals. As mentioned in the previous subsection, groups can bring about structural changes that individuals cannot. And structural changes much more plausibly have an impact on the far future. Consider: our present society is surely marked by the fact that feudalism was the dominant economic system in Europe in 1000 AD, but it is less obvious that the actions of any particular European in 1000 AD had a significant, predictable effect upon us. So if utilitarianism’s moral obligations apply to groups, the case for longtermism might be strengthened – at least, for groups. Is this correct, and if so, which groups should longtermists be interested in, and which actions should they recommend they take?

A note of caution about longtermism was raised, nonetheless, in Chapter Five. Insofar as worries about utilitarianism’s alienation are avoided by appeal to our interdependence, communal relations and similarity with others, it is not obvious how far this could extend to future people. As I said there, I find a waning sympathy for those sentient beings living millions of years, or millions of light-years away, and those of little or no social relation to me, especially if they are supposed to be different species, or artificial, or have a well-being that consists in something quite different to mine. I find ‘longtermist’ arguments more plausible when I consider that future generations will either have lives, hopes and dreams, pains and projects quite like mine, and that they will be part of a collective human story to which I also belong (see also Scheffler 2018). Is this a mere prejudice of mine or the start of a cogent argument against longtermism?

c. Socialism and the good life

Perhaps surprisingly for a thesis about utilitarianism, I have said very little about what well-being is. This is deliberate. Many of my results are such that they hold given any plausible theory of well-being, and I would like my arguments to be as ecumenical as possible in this regard. However, a further direction for research in the 'left-utilitarian' vein would be developing a socialist-infused account of the good life.

Since Parfit (1984, 493–502) it has been traditional to divide theories of well-being into hedonistic, desire-satisfaction and objective list theories. Socialists are typically sceptical of theories of the first two kinds, I think, given the influence of ideology on what we find pleasing and desirable. So which good would be on the socialist 'objective list', as constitutive elements of well-being? Socialists tend to emphasise (often implicitly) as elements of well-being: standing in fraternal social relations, the meeting of basic needs, the capacity to serve others, being part of a larger societal project, autonomy over productive activities, and freedom from domination. They tend to place less weight than others on individual advancement and success, on material prosperity beyond the meeting of needs, on meeting traditional social expectations, and on supernatural and spiritual connections.

Working out and defending a socialist account of well-being would be a crucial part of the left-utilitarian project. In order to apply utilitarianism, in policy or everyday life, we need to have some idea of what well-being is. Different theories of well-being render different interventions 'effective': for instance, Tyler Cowen (2007) argues that utilitarians who care about future generations should maximise the rate of economic growth at the expense of other social justice considerations. Would this claim be true given a socialist account of well-being? Which other social arrangements would be utility-maximising given such an account? And if individual well-being involves serving others, how far could this soften the blow of utilitarianism's demands to sacrifice one's own interests for those of others?

d. What do we want from a moral theory?

I granted to Williams that a moral theory should not be unreasonably alienating, and argued that utilitarianism's incompatibility with commitment does not make it so. But is this a good desideratum for a moral theory? Partly, the answer will depend on one's metaethical view. It seems likely that moral realists who believe in mind-independent moral truths are unlikely to dismiss moral theories on the grounds that accepting them might alienate us – the truths are mind-independent, after all – whereas those who think that moral theory is instead a way of understanding our own minds and philosophical commitments with respect to how we treat one another will be more moved by alienation worries. Are these suspicions correct?

In Chapter Six, I proposed a position that fragmented some of our key moral concepts: in particular, what one ought not to do was distinguished from what one ought to be blamed or punished for, or feel guilt about. Such a division is controversial amongst philosophers, with some holding that rightness and blameworthiness are closely conceptually connected (Wallace 2019; Darwall 2013). But how far is this distinction sustainable, and how far is it at odds with folk views of morality? What might be the further implications of such a fragmentation? Does utilitarianism's recourse to such revisions of concepts reveal its inadequacy, or could it be the beginning of a utilitarian response to a range of other objections (Railton 1988)? Again, the answers here will depend on metaethical views: do we think of morality as primarily a set of norms for a community, or guidance for individuals in specific circumstances, and is there a useful sense of the latter distinct from the former?

These questions will continue to occupy my mind, though it will be on another occasion that my answers occupy space on the page.

Bibliography

Acemoglu, Daron, and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity and Poverty*. London: Profile.

Adler, Matthew D. 2019. *Measuring Social Welfare: An Introduction*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780190643027.001.0001.

Alesina, Alberto, Rafael Di Tella, and Robert MacCulloch. 2004. 'Inequality and Happiness: Are Europeans and Americans Different?' *Journal of Public Economics* 88 (9): 2009–42. https://doi.org/10.1016/j.jpubeco.2003.07.006.

Anderson, Elizabeth S. 1999. 'What Is the Point of Equality?' *Ethics* 109 (2): 287–337.

Arneson, Richard J. 1989. 'Equality and Equal Opportunity for Welfare'. *Philosophical Studies* 56 (1): 77–93. https://doi.org/10.1007/BF00646210.

Ashford, Elizabeth. 2000. 'Utilitarianism, Integrity, and Partiality'. *Journal of Philosophy* 97 (8): 421–39. https://doi.org/10.5840/jphil200097834.

———. 2005. 'Utilitarianism with a Humean Face'. *Hume Studies* 31 (1): 63–92. https://doi.org/10.1353/hms.2011.0177.

———. 2007. 'The Duties Imposed by the Human Right to Basic Necessities'. In *Freedom From Poverty as a Human Right: Who Owes What to the Very Poor? Co-Published with Unesco*, edited by Thomas Pogge. Oxford University Press.

———. 2021. 'Individualist Utilitarianism and Converging Theories of Rights'. In *Principles and Persons*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780192893994.003.0008.

Atkinson, Anthony B. 1970. 'On the Measurement of Inequality'. *Journal of Economic Theory* 2 (3): 244–63. https://doi.org/10.1016/0022-0531(70)90039-6.

Babones, Salvatore J. 2008. 'Income Inequality and Population Health: Correlation and Causality'. *Social Science & Medicine (1982)* 66 (7): 1614–26. https://doi.org/10.1016/j.socscimed.2007.12.012.

Bacharach, Michael. 2018. *Beyond Individual Choice*. Edited by Robert Sugden and Natalie Gold. Princeton University Press. https://www.degruyter.com/document/doi/10.1515/9780691186313/html.

Bakunin, Mikhail. 1973. 'God and the State'. In *Bakunin on Anarchy: Selected Works by the Activist-Founder of World Anarchism*, edited by Sam Dolgoff, 225–42. London: George Allen & Unwin Ltd.

Bangham, George. 2019. 'Happy Now? Lessons for Economic Policy Makers from a Focus on Subjective Well-Being'. *Resolution Foundation* (blog). 2019. https://www.resolutionfoundation.org/publications/happy-now-lessons-for-economic-policy-makers-from-a-focus-on-subjective-well-being/.

Benn, Tony. 1998. '"Don't Arab and Iraqi Women Weep When Their Children Die?"' Parliamentary Speech, House of Commons, February 17. https://api.parliament.uk/historic-hansard/commons/1998/feb/17/iraq#S6CV0306P0_19980217_HOC_220.

Bicchieri, Cristina. 2017. *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190622046.001.0001.

Blair, Tony. 1995. 'Leader's Speech'. Brighton, October 3. http://www.britishpoliticalspeech.org/speech-archive.htm?speech=201.

Blanchflower, David G., and Andrew J. Oswald. 2004. 'Well-Being over Time in Britain and the USA'. *Journal of Public Economics* 88 (7–8): 1359–86. https://doi.org/10.1016/S0047-2727(02)00168-8.

Bostrom, Nick. 2003. 'Astronomical Waste: The Opportunity Cost of Delayed Technological Development'. *Utilitas* 15 (3): 308–14. https://doi.org/10.1017/S0953820800004076.

Bourget, D., and D. J. Chalmers. ms. 'Philosophers on Philosophy: The PhilPapers 2020 Survey'. https://survey2020.philpeople.org/.

Broome, John. 2004. *Weighing Lives*. Oxford University Press.

———. 2019. 'Against Denialism'. *The Monist* 102 (1): 110–29. https://doi.org/10.1093/monist/ony024.

Brown, Alexander. 2014. 'What Should Egalitarians Believe If They Really Are Egalitarian? A Reply to Martin O'Neill'. *European Journal of Political Theory* 13 (4): 453–69. https://doi.org/10.1177/1474885113506710.

Brudney, Daniel. n.d. 'On Productivity Holism'. *European Journal of Philosophy* n/a (n/a). Accessed 1 April 2022. https://doi.org/10.1111/ejop.12776.

Buchak, Lara. 2017. 'Taking Risks behind the Veil of Ignorance'. *Ethics* 127 (3): 610–44. https://doi.org/10.1086/690070.

Budolfson, Mark. 2019. 'The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response'. *Philosophical Studies* 176 (7): 1711–24. https://doi.org/10.1007/s11098-018-1087-6.

Budolfson, Mark, and Dean Spears. 2019. 'The Hidden Zero Problem: Effective Altruism and Barriers to Marginal Impact'. In *Effective Altruism: Philosophical Issues*, edited by Hilary Greaves and Theron Pummer, 184–281. https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198841364.001.0001/oso-9780198841364-chapter-12.

Bueno, David. 2011. 'The Relationship between Income, Health Status, and Health Expenditures in the United States'. Thesis, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/65780.

Buttrick, Nicholas R, Samantha J Heintzelman, and Shigehiro Oishi. 2017. 'Inequality and Well-Being'. *Current Opinion in Psychology*, Inequality and social class, 18 (December): 15–20. https://doi.org/10.1016/j.copsyc.2017.07.016.

Buttrick, Nicholas R., and Shigehiro Oishi. 2017. 'The Psychological Consequences of Income Inequality'. *Social and Personality Psychology Compass* 11 (3): e12304. https://doi.org/10.1111/spc3.12304.

Calhoun, Cheshire. 2009. 'What Good Is Commitment?' *Ethics* 119 (4): 613–41. https://doi.org/10.1086/605564.

Cassidy, John. 2010. *How Markets Fail: The Logic of Economic Calamities*. London: Penguin Books.

Chappell, Richard Yetter. 2015. 'Value Receptacles'. *Noûs* 49 (2): 322–32. https://doi.org/10.1111/nous.12023.

———. 2022. 'Level-Up Impartiality'. Substack newsletter. *Good Thoughts* (blog). 19 May 2022. https://rychappell.substack.com/p/level-up-impartiality.

Clark, Andrew E., Sarah Flèche, Richard Layard, Nattavudh Powdthavee, and George Ward. 2018. 'Income'. In *The Origins of Happiness*, NED-New edition, 33–50. The Science of Well-Being over the Life Course. Princeton University Press. https://doi.org/10.2307/j.ctvd58t1t.6.

Clark, Andrew E., Paul Frijters, and Michael A. Shields. 2008. 'Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles'. *Journal of Economic Literature* 46 (1): 95–144. https://doi.org/10.1257/jel.46.1.95.

Cohen, G. A. 1989. 'On the Currency of Egalitarian Justice'. *Ethics* 99 (4): 906–44. https://doi.org/10.1086/293126.

———. 2008. *Rescuing Justice and Equality*. Harvard University Press.

Cohen, S., W. J. Doyle, D. P. Skoner, B. S. Rabin, and J. M. Gwaltney. 1997. 'Social Ties and Susceptibility to the Common Cold'. *JAMA* 277 (24): 1940–44.

Collins, Stephanie. 2019a. 'Beyond Individualism'. In *Effective Altruism: Philosophical Issues*, by Hillary Greaves and Theron Pummer, 202–16. Oxford: Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/oso/9780198841364.001.0001/oso-9780198841364-chapter-13.

———. 2019b. *Group Duties: Their Existence and Their Implications for Individuals*. Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/oso/9780198840275.001.0001/oso-9780198840275.

Costa, P. T., and R. R. McCrae. 1988. 'Personality in Adulthood: A Six-Year Longitudinal Study of Self-Reports and Spouse Ratings on the NEO Personality Inventory'. *Journal of*

Personality and Social Psychology 54 (5): 853–63. https://doi.org/10.1037//0022-3514.54.5.853.

Cowen, Tyler. 2007. 'Caring about the Distant Future: Why It Matters and What It Means'. *University of Chicago Law Review* 74 (1). https://chicagounbound.uchicago.edu/uclrev/vol74/iss1/2.

Cowherds, The. 2015. *Moonpaths: Ethics and Emptiness*. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780190260507.001.0001/acprof-9780190260507.

Cripps, Elizabeth. 2013. *Climate Change and the Moral Agent: Individual Duties in an Interdependent World. Climate Change and the Moral Agent*. Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199665655.001.0001/acprof-9780199665655.

Crisp, Roger. 1992. 'Utilitarianism and the Life of Virtue'. *The Philosophical Quarterly (1950-)* 42 (167): 139–60. https://doi.org/10.2307/2220212.

———. 2003. 'Equality, Priority, and Compassion'. *Ethics* 113 (4): 745–63. https://doi.org/10.1086/373954.

———. 2018. 'Against Partiality'. The University of Kansas, April 19. http://hdl.handle.net/1808/26747.

Darwall, Stephen. 2013. *Morality, Authority, and Law: Essays in Second-Personal Ethics I.* Oxford University Press.

Dickerson, S. S., and M. E. Kemeny. 2004. 'Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research 130 (3): 3 5 5 -9 1.' *Psychological Bulletin* 130 (3): 355–91.

Dietz, Alexander. 2016. 'What We Together Ought to Do'. *Ethics* 126 (4): 955–82. https://doi.org/10.1086/686002.

———. 2019. 'Effective Altruism and Collective Obligations'. *Utilitas* 31 (1): 106–15. https://doi.org/10.1017/s0953820818000158.

Driver, Julia. 2005. 'Consequentialism and Feminist Ethics'. *Hypatia* 20 (4): 183–99. https://doi.org/10.1111/j.1527-2001.2005.tb00543.x.

Dworkin, Ronald. 1981. 'What Is Equality? Part 2: Equality of Resources'. *Philosophy and Public Affairs* 10 (4): 283–345.

———. 2000. *Sovereign Virtue*. Cambridge, MA and London, England: Harvard University Press.

Easwaran, Kenny. 2021. 'A New Method for Value Aggregation'. *Proceedings of the Aristotelian Society* 121 (3): 299–326. https://doi.org/10.1093/arisoc/aoab008.

Ecob, Russell, and George Davey Smith. 1999. 'Income and Health: What Is the Nature of the Relationship?' *Social Science & Medicine* 48 (5): 693–705. https://doi.org/10.1016/S0277-9536(98)00385-2.

Eggleston, Ben. 2013. 'Rejecting The Publicity Condition: The Inevitability of Esoteric Morality'. *Philosophical Quarterly* 63 (250): 29–57. https://doi.org/10.1111/j.1467-9213.2012.00106.x.

Elster, Jon. 1983. *Sour Grapes: Studies in the Subversion of Rationality*. Editions De La Maison des Sciences De L'Homme.

Estlund, David. 2020. *Utopophobia: On the Limits (If Any) of Political Philosophy*. Princeton, New Jersey: Princeton University Press. https://www.jstor.org/stable/10.2307/j.ctvhrd1gx.

Feldman, Fred. 1980. 'The Principle of Moral Harmony'. *Journal of Philosophy* 77 (3): 166–79. https://doi.org/10.2307/2025668.

Financial Times. 2007. 'Citigroup Chief Stays Bullish on Buy-Outs', 9 July 2007.

Foot, Philippa. 1967. 'The Problem of Abortion and the Doctrine of Double Effect'. *Oxford Review* 5: 5–15.

———. 1983. 'Utilitarianism and the Virtues'. *Proceedings and Addresses of the American Philosophical Association* 57 (2): 273–83. https://doi.org/10.2307/3131701.

Frankena, William K. 1973. *Ethics*. 2d ed. Prentice-Hall Foundations of Philosophy Series. Englewood Cliffs, N.J.: Prentice-Hall.

Frankfurt, Harry. 1987. 'Equality as a Moral Ideal'. *Ethics* 98 (1): 21–43. https://doi.org/10.1086/292913.

Frankfurt, Harry G. 1988. 'Rationality and the Unthinkable'. In *The Importance of What We Care About: Philosophical Essays*, 177–90. Cambridge University Press. https://doi.org/10.1017/CBO9780511818172.014.

Garfield, Jay L., Stephen Jenkins, and Graham Priest. 2015. 'The Śāntideva Passage: Bodhicaryāvatāra VIII.90–103'. In *Moonpaths: Ethics and Emptiness*, by The Cowherds. Oxford: Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780190260507.001.0001/acprof-9780190260507-chapter-5.

Gauthier, David P. 1963. *Practical Reasoning; the Structure and Foundations of Prudential and Moral Arguments and Their Exemplification in Discourse*. Oxford, Clarendon Press. http://archive.org/details/practicalreasoni0000gaut.

Gilabert, Pablo. 2015. 'The Socialist Principle "From Each According To Their Abilities, To Each According To Their Needs"'. *Journal of Social Philosophy* 46 (2): 197–225. https://doi.org/10.1111/josp.12096.

Gilbert, Margaret. 1989. *On Social Facts*. London: Routledge.

———. 1994. ‘Remarks on Collective Belief’. In *Socializing Epistemology: The Social Dimensions of Knowledge*, edited by Frederick F. Schmitt, 235–56. Rowman & Littlefield.

———. 2006. *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford: Oxford University Press. https://doi.org/10.1093/0199274959.001.0001.

———. 2013. *Joint Commitment: How We Make the Social World*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199970148.001.0001.

Goldings, Herbert Jeremy. 1954. ‘On the Avowal and Projection of Happiness’. *Journal of Personality* 23 (1): 30–47. https://doi.org/10.1111/j.1467-6494.1954.tb02336.x.

Goodin, Robert E. 1988. ‘What Is So Special about Our Fellow Countrymen?’ *Ethics* 98 (4): 663–86.

———. 1995. *Utilitarianism as a Public Philosophy*. Cambridge University Press.

Goodman, Charles. 2017. ‘Śāntideva’s Impartialist Ethics’. The Oxford Handbook of Indian Philosophy. 2017. https://doi.org/10.1093/oxfordhb/9780199314621.013.23.

Greaves, Hilary. 2020. ‘Global Consequentialism’. In *The Oxford Handbook of Consequentialism*, edited by Douglas W. Portmore, 422–40. New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190905323.013.11.

Greaves, Hilary, and William MacAskill. 2021. ‘The Case for Strong Longtermism’. *Global Priorities Institute Working Papers* (blog). 14 June 2021. https://globalprioritiesinstitute.org/hilary-greaves-william-macaskill-the-case-for-strong-longtermism-2/.

Greaves, Hilary, William MacAskill, Rossa O’Keeffe-O’Donovan, Philip Trammell, Benjamin Tereick, Andreas Mogensen, Christian Tarsney, Gustav Alexandrie, and Maxime Cugnon de Sévricourt. 2020. ‘A Research Agenda for the Global Priorities Institute’. Global Priorirites Institute. https://globalprioritiesinstitute.org/research-agenda/.

Gross, Daniel A. 2021. ‘Peter Singer Is Committed to Controversial Ideas’. The New Yorker. 25 April 2021. https://www.newyorker.com/culture/the-new-yorker-interview/peter-singer-is-committed-to-controversial-ideas.

Gustafsson, Johan E. 2021. ‘Utilitarianism Without Moral Aggregation’. *Canadian Journal of Philosophy* 51 (4): 256–69.

Gyekye, Kwame. 2010. ‘Person and Community in Akan Thought’. In *Person and Community: Ghanaian Philosophical Studies, I*, edited by Kwasi Wiredu and Kwame Gyekye, 101–22. Cultural Heritage and Contemporary Change. Series II: Africa. 1. Washington, D. C.: The Council for Research in Values and Philosophy.

Hagerty, Michael R. 2000. ‘Social Comparisons of Income in One’s Community: Evidence from National Surveys of Income and Happiness’. *Journal of Personality and Social Psychology* 78 (4): 764–71. https://doi.org/10.1037/0022-3514.78.4.764.

Harcourt, Edward. 1998. ‘Integrity, Practical Deliberation and Utilitarianism’. *The Philosophical Quarterly* 48 (191): 189–98. https://doi.org/10.1111/1467-9213.00091.

Hare, R. M. 1981. *Moral Thinking: Its Levels, Method, and Point*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/0198246609.001.0001/acprof-9780198246602.

Harman, Gilbert. 1986. *Change in View*. Cambridge, Mass.: MIT Press.

Harsanyi, John C. 1977. ‘Morality and the Theory of Rational Behavior’. *Social Research* 44 (4): 623–56.

Helliwell, John F., Richard Layard, Jeffrey Sachs, and Jan-Emmanuel De Neve, eds. 2020. ‘World Happiness Report 2020’. New York: Sustainable Development Solutions. https://worldhappiness.report/ed/2020/#read.

Herzog, Liza. 2016. ‘Can “Effective Altruism” Really Change the World?’ OpenDemocracy. 2016. https://www.opendemocracy.net/en/transformation/can-effective-altruism-really-change-world/.

Hobbes, Thomas. 1997. *On the Citizen (1642).* Translated by Richard Tuck and Michael Silverthorne. Cambridge Texts in the History of Political Thought. Cambridge: University Press. https://doi.org/10.1017/CBO9780511808173.

Hooker, Brad. 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Clarendon Press.

Humberstone, I. L. 1971. ‘Two Sorts of ’Ought’s’. *Analysis* 32 (1): 8–11. https://doi.org/10.1093/analys/32.1.8.

Hume, David. 1965. ‘An Enquiry Concerning the Principles of Morals (1751)’. In *Hume’s Ethical Thought*, edited by Alasdair MacIntyre, 23–156. Classics in the History of Thought. London: Macmillan.

Hyman, John. 2015. *Action, Knowledge, and Will. Action, Knowledge, and Will.* Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780198735779.001.0001/acprof-9780198735779.

Jackson, Ben. 2004. ‘The Uses of Utilitarianism: Social Justice, Welfare Economics and British Socialism, 1931–48’. *History of Political Thought* 25 (3): 508–35.

Jackson, Frank. 1987. ‘Group Morality’. In *Metaphysics and Morality: Essays in Honour of J.J.C. Smart*, edited by J. J. C. Smart, Philip Pettit, Richard Sylvan, and Jean Norman. Blackwell.

———. 1991. ‘Decision-Theoretic Consequentialism and the Nearest and Dearest Objection’. *Ethics* 101 (3): 461–82.

James, C. L. R. 2005. *Beyond a Boundary (1963)*. Yellow Jersey Press.

Kandiyali, Jan. 2020. ‘The Importance of Others: Marx on Unalienated Production’. *Ethics* 130 (4): 555–87. https://doi.org/10.1086/708536.

Kang, Haejo, and Dong-Eun Rhee. 2021. ‘Does Income (Re)Distribution Matter for Subjective Well-Being? Evidence from Cross-Country Panel Data’. *Social Science Quarterly* 102 (2): 706–21. https://doi.org/10.1111/ssqu.12943.

Karpus, Jurgis, and Natalie Gold. 2017. ‘Team Reasoning: Theory and Evidence’. In *The Routledge Handbook of Philosophy of the Social Mind*, edited by Julian Kiverstein, 400–417. New York, USA: Routledge.

Kingston, Ewan, and Walter Sinnott-Armstrong. 2018. ‘What’s Wrong with Joyguzzling?’ *Ethical Theory and Moral Practice* 21 (1): 169–86. https://doi.org/10.1007/s10677-017-9859-1.

Kolodny, Niko. n.d. ‘The Pecking Order: Social Hierarchy as a Philosophical Problem’. https://www.ocf.berkeley.edu/~ngkolodny/The_Pecking_Order.docx.

Lackey, Jennifer. 2016. ‘What Is Justified Group Belief?’ *The Philosophical Review* 125 (3): 341–96. https://doi.org/10.1215/00318108-3516946.

———. 2020. ‘Group Belief: Lessons from Lies and Bullshit’. *Aristotelian Society Supplementary Volume* 94 (1): 185–208. https://doi.org/10.1093/arisup/akaa007.

Layard, R., G. Mayraz, and S. Nickell. 2008. ‘The Marginal Utility of Income’. *Journal of Public Economics* 92 (8–9): 1846–57. https://doi.org/10.1016/j.jpubeco.2008.01.007.

———. 2010. ‘Does Relative Income Matter? Are the Critics Right?’ In *International Differences in Well-Being*, by Ed Diener, Daniel Kahneman, and John F. Helliwell. Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199732739.001.0001/acprof-9780199732739-chapter-6.

Lazari-Radek, Katarzyna de, and Peter Singer. 2010. ‘Secrecy in Consequentialism: A Defence of Esoteric Morality’. *Ratio* 23 (1): 34–58. https://doi.org/10.1111/j.1467-9329.2009.00449.x.

Leopold, David. 2018. ‘Alienation’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2018. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/alienation/.

List, Christian, and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press.

Little, Margaret Olivia. 2007. ‘Abortion and the Margins of Personhood’. *Rutgers Law Journal* 39: 331.

Macaskill, William. 2015. *Doing Good Better: How Effective Altruism Can Help You Make a Difference*. New York, N.Y: Avery Publishing Group.

Mackenbach, Johan P., Pekka Martikainen, Caspar WN Looman, Jetty AA Dalstra, Anton E. Kunst, and Eero Lahelma. 2005. 'The Shape of the Relationship between Income and Self-Assessed Health: An International Study'. *International Journal of Epidemiology* 34 (2): 286–93. https://doi.org/10.1093/ije/dyh338.

MacKinnon, Catharine A. 1989. *Toward a Feminist Theory of the State*. Cambridge, Mass.: Harvard University Press.

Maltais, Aaron. 2013. 'Radically Non-Ideal Climate Politics and the Obligation to at Least Vote Green'. *Environmental Values* 22 (5): 589–608. https://doi.org/10.3197/096327113X13745164553798.

Marmot, Michael. 2015. *Status Syndrome: How Your Place on the Social Gradient Directly Affects Your Health*. 2nd ed. London: Bloomsbury.

Marx, Karl. 1988. *The Economic and Philosophic Manuscripts of 1844*. Translated by Martin Milligan. Great Books in Philosophy. Amherst, New York: Prometheus Press.

———. 1990. *Capital, Volume I (1867)*. Translated by Ben Fowkes. London: Penguin Classics.

———. 2000a. 'Critique of the Gotha Programme (1875)'. In *Karl Marx: Selected Writings*, edited by David McLellan, 2nd ed., 610–16. Oxford: Oxford University Press.

———. 2000b. 'On James Mill (1844)'. In *Karl Marx: Selected Writings*, edited by David McLellan, 2nd ed., 124–33. Oxford, UK: Oxford University Press.

———. 2000c. 'The Eighteenth Brumaire of Louis Bonaparte (1852)'. In *Karl Marx: Selected Writings*, edited by David McLellan, Second Edition, 329–55. Oxford: Oxford University Press.

Marx, Karl, and Friedrich Engels. 2000a. 'The Communist Manifesto (1848)'. In *Karl Marx: Selected Writings*, edited by David McLellan, 2nd ed., 245–72. Oxford, UK: Oxford University Press.

———. 2000b. 'The German Ideology (1846)'. In *Karl Marx: Selected Writings*, edited by David McLellan, 2nd ed., 175–209. Oxford, UK: Oxford University Press.

McCabe, Helen. 2021. *John Stuart Mill, Socialist*. Montreal ; Kingston ; London ; Chicago: McGill-Queen's University Press.

Melita, D., G. B. Willis, and R. Rodríguez-Bailón. 2021. 'Economic Inequality Increases Status Anxiety Through Perceived Contextual Competitiveness'. *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.637365.

Mellor, Rowan. n.d. 'Joint Ought'. *Unpublished*.

Milanovic, Branko. 2015. 'Global Inequality of Opportunity: How Much of Our Income Is Determined by Where We Live?' *The Review of Economics and Statistics* 97 (2): 452–60. https://doi.org/10.1162/REST_a_00432.

Mill, John Stuart. 1965. *Principles of Political Economy with Some of Their Applications to Social Philosophy (1848)*. Edited by John M. Robson. Toronto; London: University of Toronto Press; Routledge and Kegan Paul.

———. 2008a. 'Considerations on Representative Government (1861)'. In *On Liberty and Other Essays*, 205–470. Oxford, UK: Oxford University Press.

———. 2008b. 'The Subjection of Women (1869)'. In *On Liberty and Other Essays*, 471–582. Oxford, UK: Oxford University Press.

———. 2008c. 'Utilitarianism (1863)'. In *On Liberty and Other Essays*. Oxford: Oxford University Press.

Mitchell, Polly. 2018. 'Adaptive Preferences, Adapted Preferences'. *Mind* 127 (508): 1003–25. https://doi.org/10.1093/mind/fzy020.

Mortensen, Laust H, Johan Rehnberg, Espen Dahl, Finn Diderichsen, Jon Ivar Elstad, Pekka Martikainen, David Rehkopf, Lasse Tarkiainen, and Johan Fritzell. 2016. 'Shape of the Association between Income and Mortality: A Cohort Study of Denmark, Finland, Norway and Sweden in 1995 and 2003'. *BMJ Open* 6 (12). https://doi.org/10.1136/bmjopen-2015-010974.

Mulgan, Tim. 2001. *The Demands of Consequentialism*. Oxford: Clarendon.

Nagel, Thomas. 1970. *The Possibility of Altruism*. Oxford Clarendon Press.

———. 1971. 'The Absurd'. *The Journal of Philosophy* 68 (20): 716–27. https://doi.org/10.2307/2024942.

Nefsky, Julia. 2015. 'Fairness, Participation, and the Real Problem of Collective Harm'. *Oxford Studies in Normative Ethics* 5: 245–71.

———. 2017. 'How You Can Help, Without Making a Difference'. *Philosophical Studies* 174 (11): 2743–67. https://doi.org/10.1007/s11098-016-0808-y.

———. 2019. 'Collective Harm and the Inefficacy Problem'. *Philosophy Compass* 14 (4): e12587. https://doi.org/10.1111/phc3.12587.

Ng, Yew-Kwang. 2000. 'From Separability to Unweighted Sum: A Case for Utilitarianism'. *Theory and Decision* 49 (4): 299–312. https://doi.org/10.1023/A:1026432128221.

Nozick, Robert. 2013. *Anarchy, State, and Utopia*. Reprint edition. New York: Basic Books.

Nussbaum, Martha C. 2001. 'Symposium on Amartya Sen's Philosophy: 5 Adaptive Preferences and Women's Options'. *Economics & Philosophy* 17 (1): 67–88. https://doi.org/10.1017/S0266267101000153.

Oishi, Shigehiro, Selin Kesebir, and Ed Diener. 2011. 'Income Inequality and Happiness': *Psychological Science*, August. https://doi.org/10.1177/0956797611417262.

Oliver, Alex, and Timothy Smiley. 2013. *Plural Logic*. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199570423.001.0001/acprof-9780199570423.

O’Neill, Martin. 2008. ‘What Should Egalitarians Believe?’ *Philosophy and Public Affairs* 36 (2): 119–56.

———. 2010. ‘The Facts of Inequality’. *Journal of Moral Philosophy* 7 (3): 397–409. https://doi.org/10.1163/174552410X511383.

———. 2017. ‘Survey Article: Philosophy and Public Policy after Piketty’. *Journal of Political Philosophy* 25 (3): 343–75. https://doi.org/10.1111/jopp.12129.

O’Neill, Onora. 1996. *Towards Justice and Virtue: A Constructive Account of Practical Reasoning*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511621239.

Otsuka, Michael. 1991. ‘The Paradox of Group Beneficence’. *Philosophy and Public Affairs* 20 (2): 132–49.

Otsuka, Michael, and Alex Voorhoeve. 2018. ‘Equality versus Priority’. In *Oxford Handbook of Distributive Justice*, edited by Serena Olsaretti, 65–85. Oxford, UK: Oxford University Press. http://ukcatalogue.oup.com/.

Owen, Robert. 1970. *A New View of Society (1813), and Report to the County of Lanark (1821)*. Edited by V. A. C. Gatrell. Harmondsworth: Penguin Classics.

Owens, David. 2012. *Shaping the Normative Landscape*. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199691500.001.0001/acprof-9780199691500.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford University Press.

———. 1991. ‘Equality or Priority’. The University of Kansas, November 21. http://hdl.handle.net/1808/12405.

———. 2011. ‘On What Matters: Volume One’. 2011. http://www.oxfordscholarship.com/view/10.1093/acprof:osobl/9780199572809.001.0001/acprof-9780199572809.

———. 2016. ‘Can We Avoid the Repugnant Conclusion?’ *Theoria* 82 (2): 110–27. https://doi.org/10.1111/theo.12097.

———. 2017. *On What Matters, Volume Three*. Oxford Scholarship Online. Oxford: University Press. http://dx.doi.org/10.1093/oso/9780198778608.001.0001.

Parijs, Philippe van. 1997. *Real Freedom for All: What (If Anything) Can Justify Capitalism?* Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/0198293577.001.0001/acprof-9780198293576.

Paskov, Marii, Klarita Gërxhani, and Herman G. van de Werfhorst. 2013. ‘Income Inequality and Status Anxiety’. *GINI Discussion Paper*, no. 90 (August).

Pettit, Philip. 1993. ‘Consequentialism’. In *A Companion to Ethics*, edited by Peter Singer. Oxford: Blackwell.

———. 2015. *The Robust Demands of the Good: Ethics with Attachment, Virtue, and Respect*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198732600.001.0001/acprof-9780198732600.

Piketty, Thomas. 2020. *Capital and Ideology*. Translated by Arthur Goldhammer. London: The Belknap Press of Harvard University Press.

Pinkert, Felix. 2014. 'What We Together Can (Be Required to) Do'. *Midwest Studies in Philosophy* 38 (1): 187–202. https://doi.org/10.1111/misp.12023.

———. 2015. 'What If I Cannot Make a Difference (and Know It)'. *Ethics* 125 (4): 971–98. https://doi.org/10.1086/680909.

Pogge, Thomas W. 2002. *World Poverty and Human Rights: Cosmopolitan Responsibilities and Reforms*. Cambridge: Polity.

Postow, B. C. 1977. 'Generalized Act Utilitarianism'. *Analysis* 37 (2): 49–52. https://doi.org/10.1093/analys/37.2.49.

Railton, Peter. 1984. 'Alienation, Consequentialism, and the Demands of Morality'. *Philosophy & Public Affairs* 13 (2): 134–71.

———. 1988. 'How Thinking about Character and Utilitarianism Might Lead to Rethinking the Character of Utilitarianism'. *Midwest Studies In Philosophy* 13 (1): 398–416. https://doi.org/10.1111/j.1475-4975.1988.tb00135.x.

———. 2003. 'Facts and Values (1986)'. In *Facts, Values, and Norms: Essays toward a Morality of Consequence*, 43–68. Cambridge Studies in Philosophy. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511613982.003.

———. 2015. 'Innocent Abroad: Rupture, Liberation, and Solidarity'. *3:AM Magazine*, 2 March 2015. https://www.3ammagazine.com/3am/innocent-abroad-rupture-liberation-and-solidarity/.

Rakowski, Eric. 1993. *Equal Justice*. *Equal Justice*. Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198240792.001.0001/acprof-9780198240792.

Rawls, John. 1971. *A Theory of Justice*. Harvard University Press.

Raz, Joseph. 1975. *Practical Reason and Norms*. Hutchinson.

———. 1986. *The Morality of Freedom*. Oxford University Press.

———. 1996. *Ethics in the Public Domain: Essays in the Morality of Law and Politics*. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780198260691.001.0001/acprof-9780198260691.

———. 1999. *Postscript to the Second Edition: Rethinking Exclusionary Reasons*. *Practical Reason and Norms*. Oxford University Press.

https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780198268345.001.0001/acprof-9780198268345-chapter-7.

Regan, Donald H. 1980. *Utilitarianism and Co-Operation*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198246091.001.0001.

Reid, Jimmy. 1972. '"We'Re Not Rats. We're Human Beings."' *Tribune Magazine*, 1972. https://tribunemag.co.uk/2021/07/jimmy-reid-were-not-rats-were-human-beings.

Ross, Jacob. 2006. 'Acceptance and Practical Reason'. New Brunswick, NJ: Rutgers. https://drive.google.com/file/d/1Wj0-zg-TfUhTeHniKVaieNkqf8WFub8H/view?usp=sharing&usp=embed_facebook.

Sandvik, Ed, Ed Diener, and Larry Seidlitz. 1993. 'Subjective Well-Being: The Convergence and Stability of Self-Report and Non-Self-Report Measures'. *Journal of Personality* 61 (3): 317–42. https://doi.org/10.1111/j.1467-6494.1993.tb00283.x.

Śāntideva. 1995. *The Bodhicaryāvatāra*. Translated by Kate Crosby and Andrew Skilton. Oxford World's Classics. Oxford: Oxford University Press.

Sartre, Jean-Paul. 2004. *Critique of Dialectical Reason, Vol. 1*. Verso.

Scanlon, T. M. 1982. 'Contractualism and Utilitarianism'. In *Utilitarianism and Beyond*, edited by Amartya Kumar Sen and Bernard Williams, 103–28. Cambridge University Press.

———. 1998. *What We Owe To Each Other*. Cambridge, Mass. ; London: Belknap Press of Harvard University Press.

Scheffler, Samuel. 1994. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Clarendon Paperbacks. Oxford: Oxford University Press. https://doi.org/10.1093/0198235119.001.0001.

———. 2003. 'What Is Egalitarianism?' *Philosophy and Public Affairs* 31 (1): 5–39.

———. 2010. 'Morality and Reasonable Partiality *'. In *Partiality and Impartiality*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199579952.003.0006.

———. 2015. 'The Practice of Equality'. In *Social Equality*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199331109.003.0002.

———. 2018. *Why Worry About Future Generations?* Uehiro Series in Practical Ethics. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198798989.001.0001.

Schmidt, Andreas T., and Daan Juijn. 2021. 'Economic Inequality and the Long-Term Future'. *Global Priorities Institute Working Papers* (blog). 11 May 2021. https://globalprioritiesinstitute.org/economic-inequality-and-the-long-term-future-andreas-t-schmidt-university-of-groningen-and-daan-juijn-ce-delft/.

Schmitt, Frederick F. 1994. 'The Justification of Group Beliefs'. In *Socializing Epistemology: The Social Dimensions of Knowledge*, edited by Frederick F. Schmitt, 257–88. Rowman & Littlefield.

Schofield, Philip. 2006. *Utility and Democracy: The Political Thought of Jeremy Bentham*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198208563.001.0001.

Schröder, Martin. 2018. 'Income Inequality and Life Satisfaction: Unrelated Between Countries, Associated Within Countries Over Time'. *Journal of Happiness Studies* 19 (4): 1021–43. https://doi.org/10.1007/s10902-017-9860-3.

Schwenkenbecher, Anne. 2020. *Getting Our Act Together: A Theory of Collective Moral Obligations*. 1st ed. Routledge Studies in Ethics and Moral Theory. London: Routledge.

Sen, Amartya Kumar. 1987. 'Equality of What?' In *Liberty, Equality, and Law: Selected Tanner Lectures on Moral Philosophy*, edited by John Rawls and Sterling M. McMurrin. University of Utah Press.

———. 1995. 'Gender Inequality and Theories of Justice'. In *Women, Culture and Development: A Study of Human Capabilities*, edited by Martha C. Nussbaum and Jonathan Glover, 259–73. New York ; Oxford: Oxford University Press.

Sidgwick, Henry. 1884. *The Methods of Ethics*. 3rd ed. London: Macmillan.

———. 1998. *Practical Ethics: A Collection of Addresses and Essays*. Edited by Sissela Bok. Oxford: Oxford University Press.

Sinclair, Upton. 1994. *I, Candidate for Governor: And How I Got Licked (1934)*. University of California Press.

Singer, Peter. 1972. 'Famine, Affluence, and Morality'. *Philosophy and Public Affairs* 1 (3): 229–43.

Smith, Holly M. 1989. 'Two-Tier Moral Codes'. *Social Philosophy and Policy* 7 (1): 112–32. https://doi.org/10.1017/S0265052500001047.

Smith, Thomas H. 2009. 'Non-Distributive Blameworthiness'. *Proceedings of the Aristotelian Society* 109: 31–60. https://doi.org/10.1111/j.1467-9264.2009.00257.x.

Srinivasan, Amia. 2015. 'Stop the Robot Apocalypse'. *London Review of Books*, 23 September 2015. https://www.lrb.co.uk/the-paper/v37/n18/amia-srinivasan/stop-the-robot-apocalypse.

Sugden, Robert, and Natalie Gold. 2007. 'Theories of Team Agency'. In *Rationality and Commitment*, edited by Fabienne Peter and Hans Bernhard Schmid. Oxford University Press.

Syme, Timothy. 2019. 'Charity Vs. Revolution: Effective Altruism and the Systemic Change Objection'. *Ethical Theory and Moral Practice* 22 (1): 93–120. https://doi.org/10.1007/s10677-019-09979-5.

Tadros, Victor. 2021. ‘Overdetermination and Obligation’. In *Principles and Persons*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780192893994.003.0017.

Tanyi, Attila. 2015. ‘Moral Demands and Ethical Theory: The Case of Consequentialism’. In *Bloomsbury Companion to Analytic Philosophy*, edited by Barry Dainton and Howard Robinson, 500–527. Bloomsbury Academic.

Temkin, Larry S. 2001. ‘Inequality: A Complex, Individualistic, and Comparative Notion’. *Noûs* 35 (s1): 327–53.

———. 2003. ‘Equality, Priority, or What?’ *Economics & Philosophy* 19 (1): 61–87. https://doi.org/10.1017/S0266267103001020.

Thompson, E. P. 1980. *The Making of the English Working Class*. London: Penguin Books.

Ting, Shi, Wenbin Zang, Chen Chen, and Dapeng Chen. 2022. ‘Income Distribution and Health: What Do We Know from Chinese Data?’ *PloS One* 17 (1): e0263008. https://doi.org/10.1371/journal.pone.0263008.

Todd, Benjamin. 2012. ‘Which Ethical Careers Make a Difference?: The Replaceability Issue in the Ethics of Career Choice’. Masters Degree in Physics and Philosophy, University of Oxford. https://www.academia.edu/1807196/Which_Ethical_Careers_Make_a_Difference_The_Replaceability_Issue_in_the_Ethics_of_Career_Choice.

———. 2018. ‘Doing Good Together: How to Coordinate Effectively, and Avoid Single-Player Thinking’. 80,000 Hours. 2018. https://80000hours.org/articles/coordination/.

Tollefsen, Deborah. 2002. ‘Collective Intentionality and the Social Sciences’. *Philosophy of the Social Sciences* 32 (1): 25–50. https://doi.org/10.1177/004839310203200102.

———. 2015. *Groups as Agents*. Polity.

Trotsky, Leon. 1938. ‘Their Morals and Ours’. *The New International*, June 1938.

Tuomela, Raimo. 1992. ‘Group Beliefs’. *Synthese* 91 (3): 285–318. https://doi.org/10.1007/bf00413570.

———. 2013. *Social Ontology: Collective Intentionality and Group Agents*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199978267.001.0001.

Twenge, Jean M. 2000. ‘The Age of Anxiety? The Birth Cohort Change in Anxiety and Neuroticism, 1952–1993’. *Journal of Personality and Social Psychology* 79 (6): 1007–21. https://doi.org/10.1037/0022-3514.79.6.1007.

Van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford University Press.

Walker, Margaret Urban. 2008. *Moral Understandings: A Feminist Study in Ethics*. 2nd ed. Studies in Feminist Philosophy. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195315394.001.0001.

Wallace, R. Jay. 2019. *The Moral Nexus*. Princeton University Press.

Walzer, Michael. 1983. *Spheres of Justice: A Defense of Pluralism and Equality*. New York: Basic Books.

Wanniski, Jude. 1978. 'Taxes, Revenues, and the "Laffer Curve"'. *National Affairs*, Winter 1978. https://nationalaffairs.com/public_interest/detail/taxes-revenues-and-the-laffer-curve.

Weimann, Joachim, Andreas Knabe, and Ronnie Schöb. 2015. 'The Economic Determinants of Happiness'. In *Measuring Happiness*, 57–74. The Economics of Well-Being. The MIT Press. https://doi.org/10.2307/j.ctt17kk7zf.8.

Wenar, Leif. 2011. 'Poverty Is No Pond: Challenges For the Affluent'. In *Giving Well: The Ethics of Philanthropy*, edited by Patricia Illingworth, Thomas Pogge, and Leif Wenar, 104–32. New York: Oxford University Press, USA.

Wietmarschen, Han van. 2021. 'What Is Social Hierarchy?' *Noûs* n/a (n/a): 1–20. https://doi.org/10.1111/nous.12387.

Wiggins, David. 2006. *Ethics: Twelve Lectures on the Philosophy of Morality*. Harvard University Press.

———. 2009. 'Solidarity and the Root of the Ethical'. *Tijdschrift Voor Filosofie* 71 (2): 239–69. https://doi.org/10.2143/tvf.71.2.2038077.

Wilkinson, Hayden. 2022. 'Market Harms and Market Benefits'. *Philosophy & Public Affairs* 50 (2): 202–38. https://doi.org/10.1111/papa.12210.

Wilkinson, Richard, and Kate Pickett. 2009. *The Spirit Level: Why More Equal Societies Almost Always Do Better*. London: Allen Lane.

———. 2017. 'The Enemy between Us: The Psychological and Social Costs of Inequality'. *European Journal of Social Psychology* 47 (1): 11–24. https://doi.org/10.1002/ejsp.2275.

Williams, Bernard. 1973. 'A Critique of Utilitarianism'. In *Utilitarianism: For and Against*, 77–150. Cambridge: Cambridge University Press.

———. 1981a. 'Persons, Character, and Morality'. In *Moral Luck: Philosophical Papers 1973–1980*, edited by James Rachels. Cambridge University Press.

———. 1981b. 'Utilitarianism and Moral Self-Indulgence'. Moral Luck: Philosophical Papers 1973–1980. December 1981. https://doi.org/10.1017/CBO9781139165860.004.

———. 1981c. 'Practical Necessity'. In *Moral Luck: Philosophical Papers 1973–1980*, 1st ed., 124–31. Cambridge University Press. https://doi.org/10.1017/CBO9781139165860.

———. 1985a. *Ethics and the Limits of Philosophy*. Harvard University Press.

———. 1985b. 'The Human Prejudice'. *Philosophy as a Humanistic Discipline.*

———. 1988. 'The Structure of Hare's Theory'. In *Hare and Critics: Essays on Moral Thinking*, edited by Douglas Seanor and N. Fotion, 185–98. Oxford: Clarendon Press.

———. 1992. 'Moral Incapacity'. *Proceedings of the Aristotelian Society* 93: 59–70.

———. 1995. ‘Replies’. In *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, edited by J. E. J. Altham and Ross Harrison, 185–224. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511621086.011.

Wolf, Susan. 1982. ‘Moral Saints’. *The Journal of Philosophy* 79 (8): 419–39. https://doi.org/10.2307/2026228.

———. 2012. ‘One Thought Too Many: Love, Morality, and the Ordering of Commitment’. In *Luck, Value, and Commitment: Themes From the Ethics of Bernard Williams*, edited by Ulrike Heuer and Gerald Lang. New York: Oxford University Press, USA.

Wolff, Jonathan. 1998. ‘Fairness, Respect, and the Egalitarian Ethos’. *Philosophy and Public Affairs* 27 (2): 97–122.

Wollner, Gabriel. 2020. ‘Socialist Action’. *Philosophical Topics* 48 (2). https://doi.org/10.5840/philtopics202048224.

Woodard, Christopher. 2019. *Taking Utilitarianism Seriously*. Oxford: Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/oso/9780198732624.001.0001/oso-9780198732624.

Woolf, Virginia. 1943. ‘Professions for Women’. In *The Death of the Moth and Other Essays*, edited by Leonard Woolf. London: Readers Union: the Hogarth Press.

Young, Iris Marion. 1990. *Justice and the Politics of Difference*. Princeton University Press.

———. 1994. ‘Gender as Seriality: Thinking about Women as a Social Collective’. *Signs* 19 (3): 713–38.

———. 2006. ‘Responsibility and Global Justice: A Social Connection Model’. *Social Philosophy and Policy* 23 (1): 102–30. https://doi.org/10.1017/S0265052506060043.