

Utilitarianism for a broken future.

Tim Mulgan, University of St Andrews, tpm6@st-andrews.ac.uk

November 2010 for the Club of Amsterdam

Philosophers have a collaborative relationship with our own history. We work within traditions of thought that emphasise some questions, problems, and methods rather than others. I work in the liberal utilitarian tradition, associated most famously with Jeremy Bentham, John Stuart Mill, and Henry Sidgwick. This tradition places human well-being centre stage. Utilitarians judge everything – actions, moral codes, political and legal institutions, and even beliefs – by its impact on human flourishing. Utilitarianism is also completely impartial. As Bentham put it, ‘each is to count for one, and none for more than one’. Utilitarians count all human happiness equally, wherever and whenever it occurs.¹

My recent work has focused on the demands of utilitarianism, and our obligations to future people.² My current research draws on that earlier work, and applies it to the ethical challenges of climate change. I argue that only an impartial morality like utilitarianism can help us to think clearly about the threat of dangerous climate change. But the result is to transform our moral ideals – and especially our conception of rights.

1. The broken world.

Until very recently, moral philosophers ignored the future, concentrating exclusively on interactions between contemporaries. Future generations were only ever an afterthought. The main reason was the (often unstated) assumption that future people will be better-off than us. (A classic example is John Rawls, widely regarded as the foremost political philosopher of the 20th century, whose classic text *A Theory of Justice* devotes just ten pages to justice between generations.³) This optimistic assumption enables economic cost-benefit analysis to discount the future costs of climate change. It also enables contemporary political theory to ignore the future. We need only look after ourselves, do what is best for present people, and then bequeath our stable liberal democratic institutions, thriving economy, and scientific advances to future people. What is good for us, is also good for them. There is no conflict between present and future.

One reason that climate change matters is because it challenges this optimistic picture. We no longer take it for granted that we *will* leave our descendants better-off, or even that we *can*. And future people are no longer an afterthought. They have moved centre stage, both in popular debate and in moral philosophy.

I am especially interested in the philosophical implications of the possibility that dangerous human-induced climate change may produce a *broken world*. Following Rawls, modern western political philosophy assumes that our society enjoys *favourable conditions* – all basic needs can be met without sacrificing basic liberties. We then debate the distribution of rights

and resources beyond these minimum requirements. By contrast, a broken world lacks favourable conditions. Natural resources are insufficient to meet the basic needs of the population – perhaps due to inadequate supplies of food and drinkable water. A broken world cannot feed itself, and the resources of the earth cannot support all human beings. The climate is very unpredictable, and extreme weather events are common. Each generation is worse-off than the last, and our affluent way of life is no longer an option. Within this harsh global context, some areas are more broken than others. Indeed, some are uninhabitable by human beings.

This is not our world. Humanity currently has the resources to meet everyone's needs. But nor is the broken world merely imaginary. It is one possible future.

My current research has two dimensions. The first reimagines ethics *within* a broken world. To make the thought experiment vivid, I imagine a history of philosophy class *in* the broken world, studying classic texts from a past age of affluence (our present day). This highlights the *contingency* of our moral and political ideals – as when we study *past* political philosophers in *their* historical context.⁴

My second research topic is the impact of a broken future *on us*. If we consider only present people and their needs, then our world is not broken. But – *for utilitarians* – the well-being of future people matters as much as our own. So the needs of ‘our world’ include the needs of future people, and ‘our resources’ include its future resources. If, on this wider definition, our resources are insufficient to meet all our needs, if we must choose between present and future needs, then our world is *already broken*.

Everything about climate change is controversial in public debate. If the internet teaches us nothing else, it does remind us that every fact is denied by someone. So I want to stress the modesty of my empirical assumptions. I claim only that past and present human behaviour may produce something like a broken world at some point in the future. This modest claim is sufficient to motivate the discussion in this paper; and no-one can reasonably be confident that it is false.

2. Why future people are philosophically problematic.

Faced with a possibly broken future, we must take future people in account. Unfortunately, this is surprisingly difficult to do. Metaphysical puzzles upset the foundations of our moral thinking, while changing circumstances call into question our most cherished moral beliefs. Utilitarianism offers a way through this moral maze.

Our present moral interactions are greatly simplified by two remarkable facts: we all exist, and we can interact with one another. We bargain, trade, cooperate, threaten, and so on. Our moral thinking is about our interactions. So it is no sur-

¹ For an overview of utilitarianism – both historical and contemporary – see Mulgan, *Understanding Utilitarianism*.

² Mulgan, *The Demands of Consequentialism*; Mulgan, *Future People*.

³ Rawls, *A Theory of Justice*.

⁴ Mulgan, *Ethics for a broken world*.

prise that moral theorists presuppose existence and interaction. But future people do not yet exist – and their existence depends upon what we decide to do. Nor can we interact with them – at least not with people in the distant future.

As a result, familiar modes of moral thinking often break down when applied to future people. Consider a simple tale from the Oxford philosopher Derek Parfit.⁵ We must choose between two energy policies. The first is completely safe, but very expensive. The second is cheaper, but riskier – burying nuclear waste where there is no earthquake risk for several centuries, but a significant risk in the distant future. Suppose we choose this risky policy. Many centuries later, an earthquake releases radiation, killing thousands of people.

Our choice seems clearly wrong. But why? Intuitively, we do wrong because we harm those who die. But suppose the two energy policies lead to radically different futures – with different patterns of migration and social interaction. Now take any particular individual killed by the catastrophe. Suppose the precise chain of events leading to her existence would not have occurred if we had chosen differently – her parents would not have met, and might not even have existed themselves. But now it appears we have harmed no-one. For how can we harm someone when she would not otherwise have existed? And, if we harm no-one, how can our choice be wrong?

These puzzles of existence and interaction are especially problematic for the great traditional rival to utilitarianism: the social contract theory. This tradition, going back at least to Thomas Hobbes and John Locke in the seventeenth century, pictures justice as a mutually advantageous bargain. Justice *is* what rational self-interested people would agree to under fair conditions.⁶ If you make the bargain sufficiently hypothetical, this may work for contemporaries. Unfortunately, we cannot bargain with people in the far distant future. We hold their quality of life, and their very existence, in our hands; and they have nothing to offer us in return. A social contract with future people makes no sense.

Utilitarianism solves our puzzles of existence and interaction – by ignoring them. Utilitarianism is impersonal. Faced with two possible futures, we should choose the one with happier people. It does not matter whether they are the same people who would have existed otherwise, or different people. Also, because utilitarianism is impartial, it does not care when people live, and it does not regard interaction *with us* as a requirement of moral standing.

Social contract theories *start* with partiality, constructing justice out of individual self-interest. Utilitarians start with impartial value, and *then* make concessions to partiality and self-interest. When each person can represent her own interests, the two approaches often meet in the middle. Things go best for all if each is free to pursue her own interests – to live her life according to her own values. But when some people cannot represent themselves at all, the two approaches come apart, and it really matters which we choose. Utilitarianism gives

future people equal moral status, while any self-interest theory leaves them at our mercy.

3. *Problems for utilitarianism.*

Utilitarianism can generate obligations to future people. But utilitarianism also faces many objections. Here are some familiar ones.⁷

1. *Demandingness.* Given the state of our world, utilitarian impartiality threatens to be extremely demanding. You have ten dollars in your pocket. You could buy a book, see a movie, or give it to a reliable charity who will use it to restore someone's sight. It's pretty clear which produces more happiness. So you make the donation, and go to the cash machine to get money to go to the movies. But now you have ten dollars in your pocket. What should you do? You can see where this is going.... No movies for you.
2. *Injustice.* You are the law-enforcement officer in an isolated frontier town. A murder has been committed. Most people believe that Bob is guilty, but you know he is innocent. Unless you hang Bob now, there will be a riot in town and several people will die. You are powerless to stop the riot by lawful means. Utilitarianism says you must hang Bob.
3. *Parfit's repugnant conclusion.* Suppose you could create any possible world, with any possible population. What would you choose? Utilitarians, who seek to maximise happiness, need a theory of aggregation – taking us from the values of individual lives to the value of a population as a whole. Derek Parfit's *Reasons and Persons* launched a huge literature on this topic.⁸ The simplest account is the *total view*, where the best outcome contains the greatest total happiness. If happiness is valuable, then surely more happiness is better than less. But now consider a choice between two possible futures: A (where ten billion people enjoy wonderful lives) and Z (where a vast population have lives barely worth living). In terms of total happiness, overcrowded Z is *better* than flourishing A. Parfit finds this conclusion 'repugnant', and argues that any acceptable moral theory must avoid it.⁹
4. *Impracticality.* Utilitarianism demands that I do whatever maximises human well-being into the future. Doesn't this require impossible feats of calculation? How could I possibly know how my actions will affect the whole of human history? It is hard enough to predict their immediate impact on me. Some even deny that we know what would be good for future people, as their situation is so different from our own.

These objections reinforce one another. Are we obliged to create a world like Parfit's Z? Will the needs of future people swamp our present rights and freedoms? How can we calculate the impact of our actions on people in the distant future?

⁵ Parfit, *Reasons and Persons*, chapter 16. See also Mulgan, *Future People*, chapters 1 and 2; and Roberts and Wasserman, *Harming Future Persons*.

⁶ I offer a utilitarian critique of contemporary I offer a utilitarian critique of intergenerational social contracts in *Future People*, chapter two. For a variety of defenses of such contracts, see Gosseries and Meyer, *Intergenerational Justice*.

⁷ I discuss these objections more fully in Mulgan, *Understanding Utilitarianism*; and Mulgan, 'Consequentialism'.

⁸ Parfit, *Reasons and Persons*. See also Mulgan, *Future People*, chapter three.

⁹ Parfit, *Reasons and Persons*, p. 388.

My preferred solution to the demandingness and injustice objections draws on the liberal utilitarian tradition of J S Mill, and on recent work on rule utilitarianism.¹⁰ Instead of seeking the single individual act that maximises human happiness, we focus instead on the moral code it would be best to teach to the next generation. The best utilitarian code is *moderate* – because human psychology limits the demands that can be taught to a broad population. It *prohibits injustice* – because otherwise no-one can enjoy the key elements of happiness that Bentham labels ‘security’. More generally, to avoid extreme demands and pervasive injustice, utilitarians erect barriers between ourselves and others – protected moral spheres where we are free to put ourselves first. These barriers promote human happiness overall – by marking our lives more secure and expanding our freedom. A world with protected moral spheres is happier than one without. Utilitarianism supports barriers to its own demands. But the ultimate moral justification for those barriers remains impartial.

Individual rights are one key barrier against the impersonal utilitarian calculus. Liberal utilitarians take rights seriously. We regard our rights as non-negotiable, inalienable, inviolable – not to be traded-off against the common good or economic productivity. My rights mark the boundaries of my protected moral sphere.

Next, consider Parfit’s repugnant conclusion. I believe that our strongest intuitions in this area concern, not the comparative values of possible futures, but our obligations to future people – not what is good, but what is permitted. When, with Parfit, we object to the repugnant conclusion, what we really deny is not that Z is better than A, but that A-people have an obligation to transform their world into Z. Because rule utilitarians deny that we are always obliged to maximise the good, they can agree that Z is better than A, but still deny that we must turn our world into Parfit’s Z.

Many philosophers avoid the *repugnant conclusion* by positing a lexical threshold on our scale of human well-being. If people enjoy lives above that level, then this outweighs any number of lives barely worth living.¹¹ Parfit’s A world is better than Z. The problem is where to put this lexical threshold, as any precise location seems ad hoc. (Parfit asks us to imagine a continuum of possible lives, each slightly less desirable than the last. Proponents of a lexical threshold must find a precise point on this continuum where, suddenly, lives become incomparably more significant.) My proposed solution treats the lexical threshold, not as a fact about objective values, but as something we project onto the world in particular deliberative contexts.¹²

I then use a lexical threshold to unite a broad range of seemingly disparate features of rule utilitarianism. It structures the many commonsense prohibitions and permissions of the ideal code; and marks out the private space where each individual is both morally and practically free to put her own goals ahead of aggregate well-being. When deliberating, I can legitimately

discount options that would take me below my lexical threshold. The lexical threshold sets the limits of my utilitarian rights, and captures the essential components of a flourishing human life.

Finally, recall the charge of impracticality. Utilitarians have two main replies. The first emphasises what we *do* know about future people. We have no idea how happy future people will be. But we do know they will benefit from clean air, drinkable water, edible food, and a stable climate. Depriving future people of these things will make their lives go worse. Beyond basic needs, we know that human lives go better if people choose for themselves from a diverse range of valuable alternatives. We cannot predict what those alternatives will be, nor how future people will choose. But we don’t have to. It is enough to know that some moral principles and some political institutions encourage the emergence of valuable freedoms, while others do not. We don’t need to know what future people will do with flexible liberal democratic institutions to be reasonably confident that those institutions will be more use to them than inflexible despotic ones.

A second utilitarian reply regards our existing moral practices as stores of knowledge about well-being. Rules that have enabled our society to survive and flourish *probably* promote human well-being. We should only adopt new rules if we are very confident that they would be better. Recall the rule utilitarian’s key question: what should we teach the next generation? A tried-and-tested moral code is surely better than a brand-new code designed by philosophers.

We might now worry that rule utilitarianism is *too* conservative. Will it ever sanction departures from the existing moral code? Utilitarians have always been at the forefront of reform – consider Bentham’s famous critiques of the English common law, or Mill’s criticism of Victorian attitudes to women. Sometimes a rule or institution persists despite its negative impact on well-being – especially where that impact falls on disenfranchised or marginalised groups. In the context of climate change, existing rules and institutions fail to protect the interests of future people – the most disenfranchised group of all.

Utilitarianism is not just another complacent, conservative defence of the liberal status quo. It retains its potential for radical critique – its ability to shock. Changing circumstances can render established rules redundant. We see this very clearly when we consider the ethics of the broken world.

4. *Ethics in the broken world.*

The broken world matters because it introduces a hitherto unacknowledged phenomenon: the possibility of a lowering lexical threshold. Future people in a broken world will be worse-off than us. Our way of life is not an option for them. Needs and rights are fluid concepts. Our notions of what we need, and of what we are entitled to, inevitably expand to fit our resources. So any future where people are worse-off will *seem* broken to us. Long before they reach the point where starvation threatens, future people will find themselves unable to enjoy the quality of life that we regard as a bare minimum. The will fall below *our* lexical threshold. Climate chaos may disrupt the global economy, leaving future people with fewer material resources; while the loss of diversity among species and eco-systems, or the destruction of time-honoured human

¹⁰ The classic text is Mill, *On Liberty*. The leading contemporary rule utilitarian is Brad Hooker. See Hooker, *Ideal Code, Real World*; and Hooker, “Rule Consequentialism.” I present my own view in Mulgan, *Future People*, chapters five through nine.

¹¹ On lexical solutions to the Repugnant Conclusion, see Crisp, “Utilitarianism and the Life of Virtue”; Griffin, *Well-being*; and Parfit, “Overpopulation and the Quality of Life”.

¹² I present this solution in Mulgan, *Future People*, chapter three.

habitats and cultures, may reduce the ability of future people to choose from among a range of diverse and valuable ways of life. This would render their lives less meaningful, even if they are not worse-off in purely monetary or material terms.

How might people in a broken future set *their* lexical threshold? As the setting of a lexical threshold depends on debates within the broken world, it is impossible to specify too much in advance; what follows is necessarily speculative.

Some changes are obvious. If our broad range of resource-intensive goals, and our long lifespan, are not widely available, then no-one will insist on them. Rights to health care, and other resources-intensive needs, will be more modest.

Other revisions might be more drastic. In a broken world, we cannot meet all basic needs. So we cannot recognise a universal right to subsistence. If not everyone can survive, and if we seek a common lexical threshold for all, then we must think of it, not as a guarantee of a worthwhile life or even of survival, but as an entitlement to a fair and equal *chance* of those things. Take one simple case: the allocation of water under conditions of scarcity. If an equal share of water is insufficient for survival, it makes no sense to give everyone an equal inadequate share rather than an equal *chance* of an adequate share. Utilitarian political institutions would shift from securing needs and liberties for all, to managing a fair distribution of chances to secure them.

Furthermore, basic needs will inevitably conflict with other rights – especially those involving personal liberty. In a broken world, *any* inefficiency in food production leads to starvation, as does *any* diversion of economic activity from necessities to luxuries. Even if liberty generally tends to promote economic efficiency, it is extremely unlikely that it always maximises food production. Future people in a broken world must rethink the relationship between freedom and survival.

We might be tempted to throw up our hands, and say that justice and morality no longer apply in such a world. But we shouldn't. Utilitarian justice still makes sense, as it compares feasible institutions, not unattainable ideals. Even in a broken world – indeed, especially in a broken world – we can still ask whether one arrangement better promotes human welfare than another. Instead of abandoning justice, we must reconceptualise justice – isolating some essential notion that *does* translate to the broken world.

And, if we take seriously the thought that future people matter as much as present people – then perhaps we need to rethink our lexical threshold now. Take a concrete example. Suppose we discover that, if we insist on seventy years of good health for ourselves, and insist on the necessary investment in medical technology, then our descendants can only hope for a reasonable chance of fifty years of moderate health. Can we still regard a lifespan of seventy years as a right? If so, why is it a right for us and not for future people? Or suppose we discover that, if we guarantee our basic needs, then our descendants will have to institute a survival lottery – a social decision procedure to determine who lives and who dies. Can we insist on guaranteed survival for ourselves; or should we move in their direction, and institute a survival lottery across generations?

Utilitarianism does not easily answer these questions. Rethinking priorities and balancing competing needs are tasks for so-

cial deliberation, not philosophical decree. But utilitarianism does offer a framework within which these questions – vital to the future of happiness – can be addressed.

Utilitarianism is often attacked for its willingness to think the unthinkable. Elizabeth Anscombe went so far as to describe utilitarian thinking as the product of a corrupt mind.¹³ In a broken world, where the unthinkable must be thought, this willingness becomes, not a vice, but a necessary virtue.

REFERENCES

- Anscombe, G. E. M., 'Modern Moral Philosophy', *Philosophy*, 23, 1957, pp. 16-17.
- Crisp, R., 'Utilitarianism and the Life of Virtue', *The Philosophical Quarterly*, 42, 1992, pp. 139-160.
- Gosseries, A., and Meyer, L., *Intergenerational Justice*, Oxford University Press, 2009.
- Griffin, J., *Well-being*, Oxford University Press, 1986.
- Hooker, B., *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*, Oxford University Press, 2000.
- Hooker, B., "Rule Consequentialism." In *The Stanford Encyclopedia of Philosophy (Spring 2004 Edition)*. Edited by Edward N. Zalta. [<http://plato.stanford.edu/ezproxy.auckland.ac.nz/archives/spr2004/entries/consequentialism-rule/>]
- Mill, J. S., *On Liberty* (first published 1861).
- Mulgan, T., *The Demands of Consequentialism*, Oxford University Press, 2001.
- Mulgan, T., *Future People*, Oxford University Press, 2006.
- Mulgan, T., *Understanding Utilitarianism*, Acumen, 2007.
- Mulgan, T., 'Consequentialism', in Oxford Bibliographies Online, forthcoming.
- Mulgan, T., *Ethics for a broken world: reimagining philosophy after catastrophe*, Acumen Publishers (UK), forthcoming 2011.
- Parfit, D., *Reasons and Persons*, Oxford University Press, 1984.
- Parfit, D., 'Overpopulation and the Quality of Life', in P. Singer, ed., *Applied Ethics*, Oxford University Press, 1986, pp. 145-164.
- Rawls, J., *A Theory of Justice*, Harvard University Press, 1971.
- Roberts, M., and Wasserman, D., eds., *Harming Future Persons: ethics, genetics and the nonidentity problem*, Springer, 2009.

¹³ Anscombe, 'Modern Moral Philosophy'.