‘Utilitarianism for a Broken World’

Tim Mulgan
University of Auckland and University of St Andrews
t.mulgan@auckland.ac.nz

Paper to be presented to the International Society of Utilitarian Studies meeting at New York University in August 2012.

Draft only–please do not cite without permission.

In my recent book *Ethics for a Broken World*, I explore the philosophical implications of the fact that climate change–or something like it–might lead to a *broken world*–a place where resources are insufficient to meet everyone's basic needs, where a chaotic climate makes life precarious, where each generation is worse-off than the last, and where our affluent way of life is no longer an option. I argue that the broken world has an impact, not only on applied ethics, but also on moral theory. This paper explores that impact, with a focus on rule utilitarianism.

Section 1 outlines the broken world, and the philosophical role I intend it to play. Section 2 illustrates the impact of the broken world on non-utilitarian moral theories–especially libertarianism and contractualism. Section 3 briefly addresses act utilitarianism. Section 4–the heart of the paper–explores the impact of the broken world on rule utilitarianism. Finally, Section 5 asks how we might re-imagine reflective equilibrium justifications in light of the broken world. (In my presentation to the ISUS conference in August 2012, I will focus on the material presented in Sections 1, 4 [especially 4.2, 4.3, and 4.4.], and 5. I will omit the material in Sections 2 and 4.1.)

1. *Ethics for a broken world.*

My earlier work focused on the demands of consequentialism, and our obligations to future people.[1] My current work draws on that earlier work, and asks how utilitarians might confront the ethical challenges of climate change. Climate change has obvious practical

[1] Mulgan, *The Demands of Consequentialism*; Mulgan, *Future People*.

implications. It will kill millions of people, wipe out thousands of species, and so on. My interest in this paper is much more parochial. How might climate change impact on moral theory–and especially on the debate between utilitarians and their non-utilitarian rivals?

I explore this question using the imaginary device of a *broken world*. This is a place where resources are insufficient to meet everyone's basic needs, where a chaotic climate makes life precarious, where each generation is worse-off than the last, and where our affluent way of life is no longer an option. 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. Indeed, if the needs of 'our world' include the needs of future people, and 'our resources' include its future resources, then our world may be *already broken*.

My broken world is not some post-apocalyptic nightmare. Human society does survive there, and social cooperation is still possible. Questions of morality and justice still arise. But the background to that cooperation, and the answers to those questions, are radically transformed. Morality and justice have a new meaning in the broken world.

Climate change–or something else–might produce a broken future. To simplify my discussion, in this paper I shall stipulate that we know that our world faces a broken future. (Or, at the very least, that it will face one unless we abandon affluent business-as-usual.) This stipulation is unrealistic. However, it is now (sadly) no less realistic than traditional optimistic assumptions that future people will be better-off than ourselves, and that favourable conditions will continue indefinitely. Our epistemic situation regarding climate change is too uncertain to allow us to make confident predictions either way.[2]

I do not assume a world without risk or uncertainty. The precise details of life in the broken world are left open, including its degree of broken-ness. I ask you to imagine that you inhabit a world where, unless something drastic is done–and perhaps even if it is–favourable

[2] For what it is worth, my own (inexpert) reading of the empirical evidence is that we can be confident neither that the future will be as bad as my broken world, nor that it will not be much worse. A particular source of uncertainty is the inability of even the most informed observers to attach meaningful probabilities to outlier possibilities where various feedback loops cause the global climate to spiral out of control once some threshold is passed.

conditions will soon disappear and future people will be worse-off than present people. (If you are a climate-change-sceptic or a technological-optimist, then you may find this hard to imagine. But then your ability to enter into the spirit of my paper will be testament to your powers of imagination.)

I have explored the broken world elsewhere–notably in my 2011 book *Ethics for a Broken World*.[3] In that book, to make the thought experiment vivid, I imagine a philosophy class in the future broken world–where both students and teachers look back in disbelief at the philosophy of a lost age of affluence, and try to make sense of the opulent worldview of affluent philosophers such as Nozick and Rawls.[4] This device of imagining how actual future people might rethink our philosophy is not merely a pedagogical marketing gimmick. It also plays a substantive philosophical role, or so I will argue.

In this paper, I focus on how the broken world might impact on utilitarianism. I must first justify the claim that this imaginary device has any impact on moral *theory* at all. Recognizing a broken future might seem relevant only to applied ethics–a simple matter of adjusting one's preferred moral theories or principles to new circumstances. Unsurprisingly, I think things are not so simple.

The broken world has three systematic impacts. First, it introduces the real prospect of conflicts between the interests of present and future people–thus forcing us to confront our obligations to distant future people. Second, imagining a future that is less prosperous than our affluent present forces us to re-evaluate our notion of what is essential to a flourishing human life. Third, by suspending the assumption of favourable conditions–by imagining a

[3] Mulgan, *Ethics for a broken world*. See also Mulgan, 'The Future of Utilitarianism'; Mulgan, 'Theory and Intuition in a Broken World'; Mulgan, 'The impact of climate change on utilitarianism and Christian ethics'; and Mulgan, 'Contractualism for a broken world'. I have also presented work-in-progress on this research project to audiences in St Andrews, Rennes, Bath, Edinburgh, Auckland, Otago, and Princeton. I am grateful to all these audiences for helpful comments.

[4] The terms 'affluent philosophy' and 'affluent philosopher' are idioms borrowed from these imaginary future students, and refer to our contemporary philosophy and its practitioners.)

world where not all basic needs can be met–the broken world forces us to deal with tragic conflicts where we must decide who lives and who dies.

The broken world thus removes three ubiquitous (and often unacknowledged) presuppositions of contemporary moral philosophy–that the interests of present and future people are largely congruent; that future people will be better-off than ourselves; and that favourable conditions will continue indefinitely. Some moral theories cope better than other with distant future obligations, with declining well-being, or with the loss of favourable conditions. Introducing a broken future thus significantly alters the balance between competing theories.

Some specific impacts of the broken world are predictable, while others are more surprising. Consider those traditions in naturalistic meta-ethics that identify moral facts with the end-points of processes of empirical moral inquiry that may turn out to be inextricably linked to an unsustainable way of life;[5] or the many strands of contemporary moral philosophy built on *intuitions* about simple cases–intuitions that are very closely tied to our affluent present;[6] or consider those modern theories of rights that implicitly presume a world where the central elements of a worthwhile life can be guaranteed to everyone.[7] All these philosophical debates must be re-thought for a world facing a broken future.

Can contemporary moral theory be reinterpreted for a broken world? Can our reflective moral intuitions about that possible future play any useful role? A characteristic of the utilitarian tradition–as opposed to more rationalistic alternatives–is that we treat questions such as these

[5] One prominent example is Frank Jackson's identification of moral facts with the 'mature folk morality' of some future community of rational inquirers. (Jackson, *From metaphysics to ethics.*) I explore the shortcomings of naturalist meta-ethics further in my forthcoming book *Purpose in the Universe*.

[6] Think of Peter Singer's busy commuter who walks past a child drowning in a pond, or Judith Thompson's famous trolley case. (Singer, 'Famine, Affluence and Morality'; and Thomson, 'Killing, Letting Die, and the Trolley Problem'.) I discuss these examples further in Mulgan, 'Theory and Intuition in a Broken World'.

[7] I discuss the problem of rights in a broken world in Mulgan, *Ethics for a broken world*; and, in relation to utilitarianism, in section 4.4 below.

as empirical rather than a priori. To discover whether our ethics can extend to the broken world, we must experiment.

Ethics for a broken world is one such experiment–or rather a series of them. Only by entering into the lives and minds of the inhabitants of one specific possible future can we discover which of our contemporary moral values, priorities, principles, idioms, or theories might earn their keep there. By imagining *different* possible futures–varying the distance from our affluent present–we can explore the limits of current morality. Such experiments are tentative and fallible, of course, but that does not make them worthless.

As the term 'experiment' is used so variously in contemporary ethics, I should say a few words about what I do–and do not–have in mind here. Our experiments must be *thought* experiments–exercises of the philosophical imagination–rather than actual experiments. There are at least four reasons for this. First, future people are not actual people. So there is no question of surveying their opinions. We can only imagine how they might think, feel, or reason. Second, my experiments take place against the background of a reflective equilibrium justification of rule utilitarianism. And reflective equilibrium deals in *considered* judgements, not raw moral beliefs. So what we must imagine is future *reflection*. Third, as in our own world–indeed, perhaps more so–the vast majority of future people will have no interest in moral theory. So there is no point in asking what future people in general will think about, say, the relative merits of Hooker's rule utilitarianism and Nozick's libertarianism. We must imagine future *philosophical* reflection. Finally, future philosophers will no doubt be as divided as we are. Let T be any future philosophical theory. We know, a priori, that most future philosophers will be convinced that T is hopeless. The important question is this: Will the features that render (some ancestor of) T attractive to some moral philosophers *today* suffice to render T attractive to philosophers of a similar temperament who are living in a broken world? To answer this question, one must enter into the worldview of both the broken future and the partisan of T. I have tried to do this elsewhere (and very briefly in the next section of this paper) with regard to libertarianism and contractualism. But things are obviously more straightforward if one is already a partisan of T. One imaginative leap is easier than two. So my focus in this paper is on utilitarianism–and especially on rule utilitarianism. In particular, I will address one specific experimental question: Does the reflective equilibrium defence of rule utilitarianism carry over to the broken world?

2. Non-utilitarianism in a broken world.

To illustrate the impact of the broken world in a little more detail, this section discusses three examples: Nozick's libertarianism, the contractualism of Rawls and Scanlon, and the prospects of a pluralism that offers different ethics for affluent and broken worlds. [This section sketches arguments developed at greater length elsewhere. I will skip-over it in my ISUS presentation.]

2.1. Libertarianism.

The opening chapters of my recent book re-read Nozick's *Anarchy, State, and Utopia* for a broken world.[8] I suggest that broken world dwellers will be amazed that any of Nozick's affluent acolytes looked to *Anarchy, State and Utopia* for a *defence* of their rights–rather than reading the book, as its author obviously intended, as a sustained ironic reductio ad absurdum of its own opening sentence. 'People have rights. Here are the necessary conditions for *any* rights to exist. These conditions have–obviously–never been met. Therefore, no-one has ever owned anything (including themselves).'

To illustrate the difficulties facing libertarianism in a broken world, consider the proviso that Nozick borrows from Locke–where any initial acquisition of property is legitimate only if it leaves 'enough and as good for others'. Nozick reinterprets Locke so that any system of property rules must leave everyone better-off than they would have been in the absence of any property rules.[9] If 'everyone' includes future people–and how could it not?–then a *broken* future spells chaos for Nozick's proviso. No property system that leads to the destruction of

[8] This section abbreviates a longer discussion in Mulgan, *Ethics for a broken world*, chapters 1 to 4.

[9] Nozick, *Anarchy, State and Utopia*, pp. 174-182. I discuss the proviso in Mulgan, *Ethics for a broken world*, chapter 3. Nozick's proviso is subject to many other objections, but my focus is on those that are introduced or exacerbated by the broken world. Not all libertarians adopt Nozick's own proviso. However, any plausible libertarianism must offer some account of the origin of property rights. And it is hard to see how any such account could deliver robust libertarian rights without making some optimistic assumptions about the future course of events in a libertarian 'free society'.

favourable conditions could possibly leave everyone better-off than they would otherwise have been. Indeed, even if the future were rosy, it is still vanishingly unlikely that any property rules remotely similar to Nozick's would leave everyone (including *each* future person) no-worse-off. Once we factor in future people, Nozick's conditions of just acquisition seem impossible to meet. But, if nothing has ever been justly acquired, then no-one has ever owned anything.

Nozick could perhaps accommodate a broken future by accepting property systems that benefit most people rather than everyone. But this would be to abandon one of the central features that distinguishes Nozick's libertarianism from utilitarianism–his commitment to individualism. As we'll now see, the broken world has a similar impact on contractualism.

2.2. Contractualism.

We turn now to contractualism.[10] The broken world brings to the fore familiar difficulties for contractualism, and also pushes the theory in new directions.

The first impact of the broken world is simply that it forces contractualists to confront a challenge that many would rather ignore: the need to offer a coherent account of our obligations to distant future people.[11]

Any contract with distant future people must overcome several barriers. The most obvious is the lack of reciprocal interaction. We cannot bargain, negotiate, or cooperate with those who will live long after us. We can do a great deal to (or for) posterity, but, as the saying goes, what has posterity ever done for us?

[10] This section brings together longer discussions in Mulgan, *Ethics for a broken world*, chapters 12 to 15; and Mulgan, 'Contractualism for a broken world'. The discussion of Scanlon and aggregation draws on discussion following my presentation at Rennes in May 2012, and was inspired by a question from Nick Southwood.

[11] For a variety of recent contractualist accounts, see Gosseries and Meyer, *Intergenerational Justice*. I present my own general critique in Mulgan, *Future People*, chapter two.

To crystallise the problem, imagine a 'time bomb' that devastates people in the distant future but has no direct impact until then.[12] (Real life analogues might involve the storage of nuclear waste or the destruction of the global climate.) Suppose that the people who will be affected are so far in the future that no-one alive today cares for them at all. Intuitively, most people believe it would still be very wrong to gratuitously plant a time bomb. But can any social contract deliver this result?

To make matters worse, puzzles of non-identity and different numbers put pressure on contractualism's person-affecting individualism–its insistence on measuring justice by impacts on individuals. This feature is often seen as a its great moral advantage over utilitarianism. But how can we begin to imagine contracts, bargains, or cooperative schemes involving future people whose existence, identity, and number depend upon what we decide? Contractualists as diverse as Kant, Rawls, Gauthier, and Scanlon all face serious difficulties here.[13]

Many ingenious contractualists have constructed idealised contracts with distant future people. But these often seem contrived and ad hoc. The underlying problem is that the metaphor of *contract*–with its subsidiary picture of a bargain between rational individuals who interact to their mutual benefit–is not a natural or helpful idiom for thinking through our relations with distant future people. Indeed, I doubt that anyone is drawn to the social contract tradition because of its account of distant future people. Rather, they are drawn to it for other reasons. Perhaps they believe that social contract copes comparatively well with other (more pressing) moral problems, or that it offers a better account of other (more significant) foundational issues, or that it is more realistic in what it assumes about (or requires of) human moral agents. (Of course, the utilitarian thinks those alleged advantages of the social contract are illusory. But that is another story.) Having been attracted to social contract for these reasons, its proponents then continue to accept it *despite* its inability to account for the possibility of intergenerational justice. After all, no theory is perfect.

[12] I owe the notion of a time bomb to Gosseries, 'What do we owe the next generation(s)?'

[13] I discuss Kant, Rawls, and Gauthier in *Future People*, chapters 1 and 2; and Scanlon in 'Contractualism for a broken world'.

This is where the broken world begins to bite. Contractualists can side-step future people only because they assume that, if we construct a fair contract between contemporaries, then the future will take care of itself. The wealthy, just, and stable society that we establish for ourselves will also best serve the interests of future people. By introducing conflicts between generations, the broken world upsets this complacent optimism, and forces contractualists to accommodate future people

The possibility that future people will be worse-off than ourselves exacerbates this difficulty. Contractualists typically assume that, under liberal democratic business-as-usual, future people will inevitably be much better-off than present people. If we grant this assumption, then even if conflicts between generations do arise, they can legitimately be settled in favour of present people. Without this optimistic assumption, it is very hard to justify the contractualist tendency to put future people to one side.

The possibility of declining wealth is especially troubling for those contractualists–notably Rawls–whose signature departure from utilitarianism is to give priority to the worst-off. Any contract which incorporates that priority will place extreme demands on present people facing a broken future.

If the prospect of conflict between an affluent present and a broken future is bad news for contractualism, then the loss of favourable conditions is arguably even worse. Favourable conditions are central for Rawls, whose theory of justice famously only applies under such conditions. But they also play a crucial–if unstated–dialectical role for other contractualists. Consider Scanlon, who presents his contractualism as a non-utilitarian respecter of the separateness of persons–safeguarding each individual by ensuring that she is not sacrificed to the greater good.[14] This advantage slips away in imaginary cases where we must choose between different individuals–such as where two groups are stranded on isolated rocks, or trolleys are hurtling out of control all around us. In an affluent world, such dilemmas can be dismissed as a philosopher's gimmick that sensible moralists can ignore.[15] In a broken world, where such situations are ubiquitous, this dismissal is untenable. Contractualism must make trade-offs–and thereby lose its distinctiveness as a moral theory.

[14] Scanlon, *What we owe to each other.*

[15] See, for instance, Wood, 'Humanity as End in Itself'.

This highlights a second impact of the broken world. As well as creating problems for contractualism in general, it also favours some versions of the theory over others. The broken world pushes contractualism in several complementary directions. First, I have argued elsewhere that, to accommodate future people, contractualism must become more hypothetical and idealised.[16] Second, as Derek Parfit has recently argued, to escape puzzles of non-identity and different numbers, contractualism must become more impersonal.[17] Third, as we have just seen, to cope with the loss of favourable conditions, contractualists must move further in a utilitarian direction–and further away from their distinctive individualism–by introducing trade-offs between individuals.

2.3. Pluralism.

Faced with these difficulties, a consistent contractualist might bite the bullet. If morality is based on mutual advantage under favourable conditions, then to remove those conditions is to leave morality behind. Obligations to future people are a contradiction in terms. We may happen to care about future people, we may choose to take their interests into account, but we owe them nothing.[18]

I return to this radical response in the final section. A more modest contractualism might limit itself to the *part* of morality that covers relations between present people under favourable conditions. For instance, Rawls's liberal theory of justice is designed for certain historical conditions. Absent those conditions, justice has no meaning. Rawls is silent about the broken world. Similarly, Scanlon's contractualism covers only what *we* owe to each other. It is best-suited to cases of reciprocal interaction, under favourable conditions, where everyone is

[16] See Mulgan, *Future People,* chapter 2; and Mulgan, 'Contractualism for a broken world'. The reason is that the lack of reciprocal interaction is especially fatal for one specific popular class of *idealised* contracts–namely those that assume contractors who are purely self-interested. Despite various desperate and ingenious attempts, no contract between such agents can deliver justice to future people.

[17] Parfit, *On what matters,* volume 2, chapters 21 and 22.

[18] Among philosophers, this response is most prominently defended by David Heyd, though not on explicitly contractualist grounds. (Heyd, *Genethics*.)

motivated to justify herself to specific individuals. If our relations with future people lack these features, then perhaps we should conclude that Scanlonian contractualism no longer applies.

Unlike the radical bullet-biter, the modest contractualist offers an account of only one part of the normative picture. This leaves open the possibility that other moral notions–characterised in non-contractualist terms–govern our relations with a broken future.

Once we admit distinct moral realms, one urgent moral task is to balance them against one another. Some non-utilitarians regard different normative realms as incommensurable–and thus refuse in principle to countenance balancings or trade-offs. But utilitarians are rightly suspicious of this cuddly pluralism.

Several recent authors–most notably Samuel Scheffler–have highlighted one disturbing feature of our commonsense morality.[19] The collective impact of its various permissions and special obligations is to promote the interests of the better-off at the expense of those less fortunate. Special obligations *oblige* the affluent to look after one another, rather than redistributing their collective wealth where it would do the most good. We buy one another dinner rather than saving the lives of the distant starving–and we feel obliged to do so.

Faced with a broken future, our common-sense rights and liberties have a similarly distorting effect *between generations*–obliging us to safeguard present rights even if the cost is a much greater infringement of future rights. The incommensurabilist refusal to balance competing normative reasons thus hands present people carte blanche to ignore the future. Given our natural tendencies to privilege the present over the future, and ourselves over others, this is the last thing the future needs.

This is a variation on a familiar Benthamite theme. All too often, the alternative to utilitarian impartiality is not principle but 'caprice'–where privileged minorities further their own interests behind a veil of convenient moral fictions. In this case, the privileged minority consists of those who happen to exist now.

[19] See, for instance, Scheffler, 'Relationships and Responsibilities'.

The broken future calls for a moral unified theory to balance present against future, and thus avoid the tyranny of the present. If contractualism cannot provide that balance, then we must look elsewhere.

3. *Act utilitarianism for a broken world.*

Enough about everyone else. How does the broken world impact on utilitarianism?

The utilitarian tradition encompasses a bewildering variety of themes and theories. To get a manageable discussion, I simplify enormously–artificially dividing utilitarians into two main groups: act and rule. While these are familiar names, I use them more broadly, to contrast two composite positions.[20]

Act utilitarians defend a purely impartial moral theory. They evaluate individual actions solely by their impact on aggregate human pleasure; and they accept the resulting verdicts, however extreme or counterintuitive. Having embraced a radical critique of our affluent moral practices, act utilitarians cannot then defend their theory on intuitive grounds. Instead, some offer more abstract arguments, deriving the utility principle from general moral ideals such as *impartiality* or *universalizability*; while others draw analogies with individual rationality–arguing that, just as the *prudent* act maximises the individual's well-being, so the *moral* act maximises aggregate well-being.

Rule utilitarians favour a moderate morality. They picture morality as a collective enterprise, and evaluate moral codes and political institutions by their collective impact on human well-being. Rule utilitarians endorse many non-utilitarian prohibitions, permissions, rights, and freedoms. Rule utilitarians often offer a reflective equilibrium justification–arguing that rule utilitarianism does the best job of tying together (and explaining) our considered moral judgments about both specific cases and general moral ideals.

This artificial division captures two competing strands to the utilitarian tradition–radical

[20] This section draws on Mulgan, *Ethics for a broken world*, chapter 6; and Mulgan, 'The future of utilitarianism.'

iconoclasm and liberal moderation.

The broken world impacts differently on different versions of utilitarianism. The beauty of act utilitarianism is a simple moral principle that applies to any situation. Act utilitarians hold that, whatever her circumstances, every agent should always perform the action that produces the best consequences. In a broken world, act utilitarians face no significant theoretical difficulties, as this simple principle carries over unchanged. However, the broken world does exacerbate one perennial difficulty for act utilitarianism–its counter-intuitiveness. In particular, act utilitarianism is notoriously demanding even when confined to an affluent present. The broken world greatly exacerbates those demands. Think of all those distant future people, worse-off than us, whose well-being is so dependent on our actions.

As with libertarianism and contractualism, the need to accommodate future people itself increases the demands of act utilitarianism–*even if* we can safely assume that those future people are very well-off. (Consider what the act utilitarian principle might demand in matters of individual reproduction, for instance.[21]) And, obviously enough, many of the devices that act utilitarians use to mitigate those demands in the case of an affluent future (such as appeals to diminishing marginal utility, or other reasons to disproportionately favour the worse-off) will backfire when it is future people, and not ourselves, who are worse-off.

Act utilitarianism is so extreme because it pictures morality as a project given to a single utilitarian agent–who must heroically maximize human happiness in a non-utilitarian world. Unsurprisingly, her life is demanding, alienating, and unattractive. But this individual model seems especially out of place against the backdrop of a broken future–where the most pressing moral issues are collective and intergenerational. (Consider the futility of asking what I should do to avert dangerous anthropogenic climate change!) In this new context, the rule utilitarian picture of morality as a task given, not to each individual agent, but to a *community* of *human beings* begins to seem much more apt.

The broken world thus shifts the intuitive balance in favour of rule utilitarianism. But act utilitarians have a radical reply–to which we return in the final section.

[21] I discuss this question in Mulgan, *Future People,* chapter 1.

4. Rule utilitarianism for a broken world.

For rule utilitarians, the fundamental moral questions are: 'What if *we* did that?', and 'How should *we* live?' My version of rule utilitarianism draws on Brad Hooker's recent formulations.[22] We first seek an *ideal moral code*. Acts are then assessed *indirectly*: the right act is the act called for by the ideal code. We imagine ourselves choosing a moral code to govern our community. I operationalize this by asking what would happen if we (the present generation) attempted to teach a given moral code to *the next generation*. This sets aside the cost of changing existing moral beliefs, but factors-in the cost of (for instance) trying to get a new generation to accept a very demanding ethic.

My formulation neatly by-passes debates over what percentage of the population should be assumed to have internalised the ideal code, and to what extent. In my rule utilitarianism, these are empirical questions to be explored, not matters that are built into the foundations of our moral theory. This has the great advantage that, as circumstances change, the theory automatically updates its recommendations. Instead of being stipulated ad hoc in advance, different rates of internalization will emerge naturally to reflect the relative difficulty of *teaching* differently demanding rules in different circumstances. Compliance with such rules may be higher in a broken future than in an affluent present; or it may be lower. But these are empirical issues to be worked through as we ask what would happen if we tried to teach this code to these (possible) future people.

Rule utilitarianism promises an overarching moral theory grounded in the utilitarian tradition–one that bases morality on the promotion of well-being, but avoids the extreme demands and injustices of act utilitarianism. Its proponents argue that it 'does a better job than its rivals of matching and tying together our moral convictions'.[23] For the sake of the present argument, let us suppose this is correct–*on the assumption* that we confine our attention to an affluent present. It is still true once we recognise a broken future?

[22] Hooker, *Ideal Code, Real World*. I develop my version in Mulgan, *The Demands of Consequentialism,* especially chapter 3; Mulgan, *Future People,* chapters 5 and 6; and in Mulgan, *Ethics for a broken world*, chapter 7.

[23] Hooker, *Ideal Code, Real World,* p. 101.

In contrast to act utilitarianism, the transition to a broken world is *not* theoretically straightforward for rule utilitarianism. In particular, it threatens to undermine the theory's *moderate* liberal credentials–a potentially fatal problem for those who defend rule utilitarianism as a moderate *alternative* to act utilitarianism.

I will consider the impact of the broken world on rule utilitarianism in two stages. First, rule utilitarians must accommodate the *future*. Second, they must adapt to a *broken* future where future people are worse-off and favourable conditions are lost. Unsurprisingly, given their desire to defend common-sense non-utilitarian moral intuitions, rule utilitarians faces difficulties analogous to those that plague libertarians and contractualists.

4.1. Well-being at the foundations.

The broken world raises the moral significance of the distant future. Rule utilitarians must look beyond the next generation, develop accounts of the morality of individual reproduction and of intergenerational justice, and engage directly with the vast literature on puzzles of aggregation. In my book *Future People*, I explore the reformulations this change in emphasis requires. In particular, a future-focus forces rule utilitarians to rethink well-being.

Well-being enters the rule utilitarian picture in two places. First, and more obvious, we must know what we are to maximise. In this section, I sketch an argument for a surprising conclusion: by focusing our attention on the future, the broken world shifts the balance between competing accounts of what well-being *is*. [24] The key premise is that, when it comes to accounting for distant future obligations, *subjectivists*–who tie a person's well-being to her own mental states, preferences or desires, and hold that what is good for you is what you think is good–are at a very significant disadvantage relative to *objectivists*–who offer a list of things that are good in themselves irrespective of the agent's attitude to them. (Such a list might include knowledge, achievement, friendship, understanding of the world and one's place in it, and so on.)

[24] This section draws on Mulgan, 'The impact of climate change on utilitarianism and Christian ethics'; and Mulgan, 'Contractualism for a broken world'.

Indeed, as even its most prominent contemporary defender–Peter Singer–has himself recently acknowledged, preference utilitarianism cannot capture what is wrong with avoiding our obligations to future people simply by manipulating their psychology–or their environment–so that they never want the good things we are now destroying.[25] Consider Singer's own example. What if future people prefer virtual reality to encounters with the real natural world?[26] What if they would *not* thank us for preserving virgin rainforest rather than pouring all our energies into developing ever-more-intricate computer games? If distant future people have no experience of diverse ecosystems or a stable climate–if they cannot even imagine such things–how can they prefer them? But then how can we fault any present decision to destroy those good things?

By contrast, proponents of an objective list theory of well-being can easily explain what is wrong with such wanton destruction.[27] We must ensure that distant future people have access to desirable things–drinkable water, breathable air, stable climate, ecological diversity–whatever their actual preferences. If a connection to the natural world is intrinsically valuable, then human lives go better (and perhaps can only go well) when they instantiate that value.

The further we look into the future, and the more that future might differ from our affluent present, the harder it is to believe that predictions about what people will (or might or could) desire have any real moral significance–or to believe that such predictions provide a solid foundation for morality.

The need to accommodate the future doesn't refute preference utilitarianism. Still less does it refute subjectivist accounts tout court. But it does render such accounts less plausible–both as guides to practical action and as foundations for an ideal code.

[25] Singer, *Practical Ethics*, third edition, 2011. Singer's own change of heart nicely illustrates my general theme in this paper. As a *practical* ethicist, Singer focuses on first-order moral issues, such as abortion, our treatment of animals, or our obligations to the distant poor. His shift away from a preference theory is driven by the failure of his own attempts to apply his subjective utilitarianism to the newly urgent practical questions posed by climate change.

[26] Singer, *Practical Ethics*, p. 244.

[27] The name 'Objective List Theory' is from Parfit, *Reasons and Persons*, Appendix I.

The shift from subjectivism to objectivism alters the foundations of rule utilitarianism–changing what the ideal code must maximise. This shift also supports rule utilitarianism. This is because, as we'll now see, rule utilitarianism is most plausible when combined with an objective list theory of well-being.

4.2. Well-being within the ideal code.

As well as influencing our story about what we should strive to maximise, the need to accommodate the future also impacts on how well-being is understood *within* the rule utilitarian ideal code.

In my book *Future People*, I argue that, to avoid the relentless demands of act utilitarianism, rule utilitarians *must* posit lexical thresholds–whereby some possible human lives are lexically superior to others.[28] This is the only way to escape Derek Parfit's repugnant conclusion without extreme demands. However, any appeal to lexicality must dissolve a seemingly devastating objection presented by Parfit himself. Parfit asks his readers to imagine a continuum of possible lives, going from the best to the worst. Where on that continuum would we locate our lexical gap? Any location seems arbitrary and ad hoc. Why draw the line *there*? His original example is worth quoting in full.[29]

'The good things in life do not come in quite different categories…Mozart and Muzak..seem to be in quite different categories. But there is a fairly smooth continuum between these two. Though Haydn is not as good as Mozart, he is very good. And there is other music which is not far below Haydn's, other music not far below this, and so on. Similar claims apply to the other best experiences, activities, and personal relationships, and to the other things which give most to the value of life. Most of these things are on fairly smooth continua, ranging from the best to the least good. Since this is so, it may be hard to defend the view that what is best has more value than any amount of what is nearly as good.'

[28] This section draws on Mulgan, *Future People*, chapter three; and Mulgan, 'Two Parfit Puzzles' [originally presented at the ISUS conference in Lisbon in 2003].

[29] Parfit, 'Overpopulation and the Quality of Life', p. 164.

My solution is to treat the lexical threshold, not as part of the theory of comparative value that is built into the foundations of rule utilitarianism, but rather as one component of the ideal code that emerges from the theory. Lexicality is not an objective feature of the world. Instead, it represents a stance to be adopted in particular deliberative contexts.

This solution arose from my diagnosis of the impasse over Parfit's repugnant conclusion. I suggest that the vast literature on Parfit's repugnant conclusion addresses the wrong question. Our strongest intuitions in this area concern, not the comparative values of possible futures, but our obligations to future people. We are most concerned, not with what is good, but with what we are obliged (or permitted) *to do*.

Act utilitarians cannot exploit this separation between foundational values and moral obligations. For them, evaluation and obligation must go together. Consider Parfit's original example, where A contains ten billion people with very flourishing lives, while Z contains a vast population whose lives are barely worth living.[30] If Z is better than A, then anyone who has a choice between these two possible futures *must* opt for Z over A. Act utilitarians who wish to avoid the obligation must deny the value claim. They must then explain *why* Z is worse than A. But this leads to an impenetrable thicket of mere addition paradoxes and impossibility theorems.

Rule utilitarianism severs this tight connection between value and obligation. This allows a more nuanced response to Parfit's repugnant conclusion–where we *agree* that Z is better than A, but then *deny* that anyone has an obligation to turn an A-world into a Z-world. We can incorporate lexical thresholds into our moral deliberations–and thus include them in our ideal code–even if our foundational value theory is still resolutely classical total utilitarian.

This is one place where the shift from subjectivism to objective list assists rule utilitarianism. Lexical gaps are much more plausible if well-being is modelled as a complex basket of diverse goods, than if the lexical threshold is a line drawn arbitrarily on a numerical scale measuring pleasure or preference-satisfaction. Rule utilitarians need an objective list because they need lexicality.

[30] Parfit, *Reasons and Persons*, p. 388.

4.3. Rule utilitarianism and the declining future.

I argue that, by shifting from preference to objective list, and by incorporating lexicality within the ideal code, rule utilitarians can offer an intuitively plausible account of future people. Once again, for the sake of argument, suppose that this is correct *on the optimistic assumption that the future is bright*. Is it still true for a broken future?

An essential feature of my rule utilitarianism is that the lexical threshold is context-dependent. This immediately raises a very significant set of moral questions. When we look to the past, to the future, to other lands, or even to less fortunate areas of our own society–we find different lexical thresholds. When my actions impact on other people, and I am deliberating about what I should do, should I use *my* interpretation of the lexical threshold or *theirs*?

Consider a simple example. Like all moderate moral theorists, rule utilitarians believe both that I am obliged to ensure that others have worthwhile lives, but also that I am not thereby obliged to sacrifice a worthwhile life for myself. The vast literature on the demands of beneficence boils down to a simple question: Does 'worthwhile life' have the same referent in both the other-regarding obligation and the self-regarding permission? Am I obliged to bring myself down to the level to which I raise others? Can I insist for myself on what cannot be guaranteed to all? In my vocabulary, can I legitimately employ two distinct lexical thresholds in the same deliberative context?

We are familiar with *rising* lexical thresholds. Over time, as our notion of the worthwhile life has evolved, our lexical threshold has risen. We now take for granted a broader range of goals, a much longer life span, and a far greater level of security than previous generations would have thought possible. But can rule utilitarianism makes sense of *decreasing* lexical levels, where each generation's aspirations fall below those of their parents? In other words, can rule utilitarianism adapt to a broken future?

In *Future People*, I argued that, to preserve its moderate credentials, rule utilitarianism must allow us to privilege our higher lexical level when confronted by the needs of people who adopt lower lexical levels due to their straightened circumstances.[31] In doing so, I explored a range of familiar utilitarian defences of such differential treatment. However, I also noted that these familiar defences presuppose the familiar case where those with lower lexical levels are strangers now living in distant lands. And, as I have argued more recently, it is unlikely that the same arguments can be deployed equally well when those who are worse-off are distant *future* people.[32] Without going into the details, the underlying problem is obvious. Given the extent to which future welfare depends on present choices, and the enormous numerical superiority of future people, it seems very unlikely that the ideal code that maximises well-being across time will allow any one generation to privilege its own lexical threshold if that means condemning future generations to a lower threshold. (By contrast, if future people will be better-off, and if their interpretation of the lexical threshold is more expansive and generous than ours, then it seems much less problematic to privilege our own *lower* threshold.) Unfortunately, unless its ideal code includes some permission to favour our own *higher* threshold, rule utilitarianism cannot remain moderate in the face of a broken future.

4.4. *Rule utilitarianism without favourable conditions.*

As with libertarianism and contractualism, the loss of favourable conditions is perhaps the most unsettling aspect of the broken world for rule utilitarians.

Drawing on arguments made famous by J. S. Mill, rule utilitarians champion their ability to accommodate a wide range of common-sense rights and freedoms, to favour democratic government over despotism, liberal society over its rivals, free markets over command-and-control economics, and so on.

In *Future People,* I suggested that we should use an expanded lexical threshold to structure the many commonsense prohibitions and permissions of the rule utilitarian ideal code. The lexical threshold marks out a protected moral sphere–a private space where each individual is both morally and practically free to concentrate on her own projects, goals, and relationships even at the expense of aggregate well-being. The lexical threshold thus evolves from a single

[31] Mulgan, *Future People,* chapter seven.

[32] Mulgan, ‘The future of utilitarianism’.

point on a scale of wellbeing to a richer notion of the essential components and background conditions of a flourishing human life. The lexical level defines my utilitarian *rights*.

My lexical idiom is idiosyncratic. But any moderate utilitarian needs to accommodate rights and freedoms, liberty and democracy. This accommodation is central to the reflective equilibrium case for rule utilitarianism against its non-utilitarian rivals. Unfortunately, like those rivals, rule utilitarianism presupposes favourable conditions. While this presupposition is often unstated, I believe it is essential to any successful utilitarian defence of traditional *rights*.

In my rule utilitarian ideal code, the lexical threshold represents a worthwhile life that is *guaranteed* to everyone. In a broken world, where favourable conditions have been lost, no-one can reasonably insist on the broad range of resource-intensive goals which, over the past few centuries, we have built into our interpretation of the lexical threshold. More drastically, if it is not possible for everyone to survive, then there is *nothing* that can meaningfully be guaranteed to everyone. Rule utilitarians must re-imagine the lexical threshold. They might begin with the notion of a fair and equal *chance* of surviving (or living a worthwhile life)–and then insist that *this fair chance* is what must be guaranteed to all. (Consider one simple case. If an equal share of water is insufficient for survival, then it makes no sense to give everyone an equal inadequate share rather than an equal *chance* of an adequate share.)

Today, perhaps with good utilitarian reasons, we regard the violation of basic human rights as unthinkable. (In my language: we build inviolable rights into our lexical threshold.) A broken world alters both the content and the strength of rights. The best utilitarian political institutions may reluctantly have to shift from securing everyone's survival to managing a fair distribution of chances to survive.

To cope with the loss of favourable conditions, utilitarians who dwell in a broken world may draw inspiration from other debates within our affluent philosophy. Perhaps future people will think of land the way we think of fishing reserves–as a fluid common resource, rather than something to be individually owned; or they may see water as we see expensive medical treatment–as something to be rationed, rather than given freely to all. As this last analogy suggests, survival lotteries–institutions that determine who lives and who dies–may emerge as the paradigm of utilitarian justice for a broken world.[33]

Just as rule utilitarians must rethink affluent rights, so must they rethink affluent freedoms. I argued earlier that the loss of favourable conditions is fatal for the contractualist. But it is also deeply worrying for any liberal utilitarian. This is hardly surprising–as rule utilitarians have sought to mirror the claims of contractualist liberals. The standard rule utilitarian case for liberty borrows the contractualist picture of contemporaries who reciprocally interact to their mutual advantage. The liberal utilitarian *then* makes optimistic assumptions about the human response to freedom, the productivity of liberal capitalism, and so on. In *that* context, on *those* assumptions, perhaps freedom does reliably promote wellbeing. But these optimistic assumptions are out-of-step with a broken world. Reciprocity is impossible when dealing with *future* people. For future people, liberal democracy has all the defects of despotism, as it leaves present people free to arbitrarily impose their will on future people. And, of course, once we face up to the possibility of a *broken* future, pessimistic assumptions are surely more reasonable than optimistic ones. Given the ways we might exercise it–given that it might produce a broken future–how could we reasonably believe that *our* freedom will benefit future people? Perhaps J. S. Mill could reasonably be confident that utilitarianism and liberalism would coincide. Faced with a broken future, we have no grounds for similar confidence. Yet without a coincidence between utilitarianism and liberalism, rule utilitarianism loses its claim to capture our considered moral judgements.

It is important not to get carried away. Many rule utilitarian arguments will translate to any world–however broken–that contains recognisable human agents. Some logically possible codes of rules will remain forever too demanding, alienating, or intricate to be effectively taught to any human population–whatever the benefits of doing so. The broken world doesn't collapse rule utilitarianism back into act utilitarianism. But it does remove our confidence that the ideal code that maximises human wellbeing into the future will resemble commonsense morality.

[33] I explore survival lotteries throughout *Ethics for a broken world*. Any such institution strikes affluent readers as both morally repellent and absurdly impractical. But, in a broken world, it might emerge as the fairest alternative–one that inspires loyalty, even from those who are unsuccessful.

The broken world thus provides a decisive intuitive test for rule utilitarians. Is our commitment to moderation an external constraint on utilitarian reasoning, or is it a contingent output of empirical reasoning? Is moderation something we retain no matter what, or is it rather something that we might jettison in the harsh light of a broken future?

5. *Justification in the broken world.*

Is this gap between theory and intuition an objection to rule utilitarianism? More broadly, suppose one's preferred moral theory, when translated to a broken world, conflicts with one's moral intuitions. Does this matter?

One radical response is to deny that this is any problem at all. Act utilitarians, for instance, might conclude that intuition–even 'considered reflective judgement'–is irrelevant to moral theory. Our moral intuitions have evolved to fit an affluent world. They thus have no normative force in a broken one, and it is no objection to a moral principle that it is 'counterintuitive'. Rather than exacerbating familiar problems of demandingness, alienation, or injustice, the broken future dissolves them. As many act utilitarians are already wary of intuitions, the broken world is further grist to their mill.

Perhaps even more radically, a thoroughgoing contractualist might accept that contractualism applies only to present people under favourable conditions–and conclude that morality (or justice) as we understand it is similarly limited. No doubt those who actually dwell in the broken world will develop their own new normative concepts–but it is not our place, as theorists of morality or justice, to speculate about that. We are sober theorists of the real-world, not purveyors of fanciful speculative fiction.

Act utilitarians apply their theory unchanged to the broken future, while their radical non-utilitarian opponents deny that any such application is required. As ever, rule utilitarians seek a middle-road, where both theory and intuition play key roles. Against the act utilitarian, rule utilitarians will ask where we might possibly find any solid foundation for our moral theory if not in an intuitive judgement of *some* kind. Against the contractualist, they will appeal to the compelling thought that, despite their inability to interact with us, and despite the many puzzles thrown-up by non-identity and different number cases, the interests of future people surely do have some *non-optional* normative force *for us*. If morality cannot capture that moral force, then so much the worse for morality.

However, if rule utilitarians want to retain a role for intuition, they must address our opening question. Does the reflective equilibrium defence of rule utilitarianism carry over to the broken world?[34]

On some readings, the answer is clearly: No. When rule utilitarianism is applied to a broken world, the result is much further from our current moral intuitions than when it is either confined to the present or applied to an affluent world with a bright future. The ideal code for a broken world–the code whose widespread internalisation would maximise human well-being throughout time–does not mirror what we think morality now demands of us. If that is what reflective equilibrium means, then it cannot be found.

But, surely, this is not the right equilibrium to seek. Reflective equilibrium deals, not in actual moral opinions, but in considered moral judgements–what we *would* believe if we reflected in light of all the morally relevant facts. And, unless we are rationalist fanatics, we surely agree that the discovery that our world faces (or even might face) a broken future *is* a morally relevant fact. Such a discovery *should* impact on our rights, permissions, and obligations. Rule utilitarians will argue that the very factors that lead their theory to offer different verdicts in a broken world should also lead us–as reflective moral thinkers–to change our moral beliefs. The question is not: (1) Does the rule utilitarian ideal code for a broken world accord with our current beliefs about what morality demands in an affluent world with a broken future? We should instead ask two subtly different questions. (2) Does this ideal code fit what we–on reflection–think morality requires *in a broken world*? (3) Does it fit what we reflectively think morality requires in an affluent world *with a broken future*?

And, to guide our reflection here–to push us reluctantly in the right direction–I throw two further questions into the mix.

[34] The remainder of this section is a response to a paper presented by Brad Hooker to a workshop on *Ethics for a broken world* in St Andrews in April 2012.

(4) What will future people–who have grown up in a broken world–believe that morality did actually demand of their ancestors in the affluent past (i.e.: of *us*) to avoid or mitigate their broken world?

(5) What will those same future people believe that morality demands *of them*–both in relation to one another, and in relation to their own descendents?

If these four new questions deliver different answers, that should give us pause for thought. Consider first the gap between (2) and (3). Could morality demand less of affluent people facing a broken future than it does of people already living in a broken world? Or now consider a gap between our reflective intuitions and those of future people–between (2) and (3), on the one hand, and (4) and (5) on the other. Can we justifiably believe that morality demands less of us than we have reason to believe that future people will *think* it demanded of us? Could we justify this discrepancy by citing our prerogative to favour our own interpretation of the lexical threshold? Or should we ensure that our moral principles are justifiable to future people–by imagining ourselves into their perspective, and asking how they might remember us? Finally, turning to our last pair of questions, can we reasonably suppose that future people–accustomed to a broken world–might hold us to a less demanding standard than they set for themselves. If so, on what basis? Can we coherently picture the grim reality of their moral lives, and still refuse to make similar sacrifices for ourselves?

By keeping the inhabitants of the broken future in mind–by asking how our collective behaviour affects their well-being–we bring our considered moral judgements into line with theirs. If we can do this, then the reflective equilibrium defence of rule utilitarianism still stands. Our ideal code has been radically transformed, but so too have our considered moral judgements. And, because the two transformations respond to the same underlying facts, there is good reason to expect them to move in tandem. The ideal code that maximises well-being into a broken future will fit our best judgements of what morality demands in such a world.

REFERENCES.

Gosseries, A., 'What do we owe the next generation(s)?', *Loyola of Los Angeles Law Review*, 2001, 35, pp. 293-354.

Gosseries, A., and Meyer, L., (eds.) *Intergenerational Justice*, Oxford University Press, 2009.

Heyd, D., *Genethics: Moral Issues in the Creation of People*, University of California Press, 1992.

Hooker, B., *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*, Oxford University Press, 2000.

Jackson, F., *From metaphysics to ethics*, Oxford University Press, 1999.

Mulgan, T., *The Demands of Consequentialism*, Oxford University Press, 2001.

Mulgan, T., 'Two Parfit Puzzles', in Ryberg, J., and Tannsjo, T., (eds.), *The Repugnant Conclusion: Essays on Population Ethics*, Springer, 2005, pp. 23-45.

Mulgan, T., *Future People*, Oxford University Press, 2006.

Mulgan, T., *Ethics for a broken world: reimagining philosophy after catastrophe*, Acumen Publishers [UK] and McGill-Queen's University Press [North America], 2011.

Mulgan, T., 2011, 'The Future of Utilitarianism', *The Tocqueville Review/La Revue Tocqueville*, 32, pp. 143-168. [Available on conference website.]

Mulgan, T., 'Theory and Intuition in a Broken World'. [English version of paper published as 'Teoria etica e intuizioni in un mondo in frantumi', *La società degli individui*, volume 39, 2010. Available on conference website.]

Mulgan, T., 'The impact of climate change on utilitarianism and Christian ethics'. [Revised draft of paper presented to the conference 'Peter Singer meets Christian ethics' at the University of Oxford in May 2011. Available on conference website.]

Mulgan, T., 'Contractualism for a broken world'. [Paper presented to workshop on contractualism, Universite de Rennes, May 2012. Available on conference website.]

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.

Parfit, D., *On What Matters*, Oxford University Press, 2011.

Roberts, M., and Wasserman, D. (eds.), *Harming Future Persons: ethics, genetics and the nonidentity problem*, Springer, 2009.

Scanlon, T., *What We Owe to Each Other*, Harvard University Press, 1999.

Scheffler, S.,‘‘Relationships and Responsibilities’’, in his *Boundaries and Allegiances*, Oxford University Press, 2001, pp. 97-110.

Singer, P., ‘Famine, Affluence and Morality’, *Philosophy and Public Affairs*, 1, 1972, pp. 229-243.

Thomson, J., ‘Killing, Letting Die, and the Trolley Problem’, *The Monist*, 1976.

Wood, A., ‘Humanity as End in Itself’, in Parfit, D., *On What Matters*, Oxford University Press, 2011, volume 2, pp. 58-82.