SYMPOSIUM

MORAL AND POLITICAL PHILOSOPHY FOR A BROKEN WORLD?

EDITORIAL PREFACE

BY

SEBASTIANO MAFFETTONE – LEAD EDITOR

GIANFRANCO PELLEGRINO – EXECUTIVE EDITOR

MICHELE BOCCHIOLA – MANAGING EDITOR

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): i-ii
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Editorial Preface

The impact of global climate change and the ensuing environmental crisis raise theoretical as well as practical challenges in contemporary moral theory and political philosophy. Philosophers cannot rely any longer on the well-settled and optimistic assumptions that dominated the debate in the last century. In a world afflicted by climate change—what Tim Mulgan calls "a broken world"—it is not longer true that future generations will be better than their ancestors, an equitable division of resources is possible, and more than an equal chance to survive can be ensured to all. Moreover, climate change makes scientific uncertainty directly relevant for moral and political issues. Likewise, it seems that any merely ideal theory will turn out to be empty, or simply unfeasible, because of the radically change in the circumstances of human life in such a broken world. Finally, reliance on intuitions—as many philosophers did by default in their justificatory strategies—seems doomed, since intuitions fitting to actual world could be completely false or misplaced in the broken world.

This Symposium of *Philosophy and Public Issues* addresses these problems. In the first part of the volume, Tim Mulgan presents his recent *Ethics for a Broken World*: *Imagining Philosophy After*

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): i-ii
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

Catastrophe (Ithaca: McGill-Queen's University Press 2011), addressing questions by Timothy Chappell, Ben Saunders and Jesse Tomalty. In the second part, we host three papers critically engaging with some attempts to conceptualize the impact of climate change and similar catastrophic challenges to contemporary moral and political theorizing.

Sebastiano Maffettone

Gianfranco Pellegrino

Michele Bocchiola

Editors of *Philosophy and Public Issues*

SYMPOSIUM

MORAL AND POLITICAL PHILOSOPHY FOR A BROKEN WORLD?

A PRÉCIS TO *ETHICS FOR A BROKEN WORLD*

BY

TIM MULGAN

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 3-10
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

A Précis to *Ethics For a Broken World*

Tim Mulgan

I

The Broken World

In my recent book *Ethics for a Broken World*, I imagine a future *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.[1] In a philosophy class in that broken world, students and teachers look back in disbelief at the philosophy of a lost age of affluence (our own time), and try to make sense of the opulent worldview of affluent philosophers such as Nozick and Rawls.

The starkest contrast between the broken future and our affluent present is the lack of what John Rawls calls "favourable conditions."[2] A society enjoys favourable conditions if it has reached a level of sophistication and prosperity such that its members can establish liberal democratic institutions to ensure that all basic needs are met without sacrificing any basic liberties. Modern liberal democracies in North America, Western Europe,

[1] Tim Mulgan, *Ethics for a Broken World: Imagining Philosophy After Catastrophe* (Montreal & Kingston-Itacha: McGill-Queen's University Press, 2011).
[2] J. Rawls, *A Theory of Justice* (Cambridge (MA): Harvard University Press 1999).

 Philosophy and Public Issues (New Series), Vol. 4, No. 2 (2014): 3-10
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

and Australasia clearly enjoy favourable conditions. And the material threshold is very modest. Liberal institutions do not require continuous economic growth, and a free society can remain comparatively poor. According to Rawls, virtually all modern societies enjoy favourable conditions.

To produce a clear, stark contrast with our affluent world, I stipulate that societies in the broken world cannot meet all basic needs. So they cannot possibly establish Rawlsian liberal institutions that both meet basic needs and protect basic liberties. In the broken world, favourable conditions are gone. Even if we do not regard the broken future as very likely, focusing on this worst-case scenario is a useful way to explore the limits and resources of our moral and political thinking.

I picture this scarcity, not as a one-off catastrophe, but as an ongoing fact of life. (A parallel might be the seasonal fluctuations in food supply experienced by traditional Inuit communities—Rawls's own example of a society lacking favourable conditions.) The scarcity of material resources (especially water) and an unpredictable climate mean that broken world societies periodically face population bottlenecks where not all can survive.

On the other hand, my broken world is not apocalyptic. Human societies do exist there. But each such society must institute a *survival lottery*—some institution that determines, in times of crisis or scarcity, who lives and who dies. For broken world political philosophers, the design of a just survival lottery is *the* central topic.

This device of imagining how actual future people might react to our contemporary philosophy is not merely a pedagogical marketing gimmick. It also plays a substantive philosophical role. The broken world affects moral *theory*, in three systematic ways, by removing three ubiquitous (and often unacknowledged)

presuppositions of contemporary moral philosophy. First, it introduces real conflicts between the interests of present and future people. This forces us to confront our obligations to distant future people. Second, if future people are worse off than present people, then we must ask what is truly *essential* to a flourishing human life. Third, a world where not all basic needs can be met raises tragic conflicts not found under favourable conditions.

Some moral theories cope better than others 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 moral and political theories.

Ethics for a broken world is an extended thought experiment. It asks how the inhabitants of one specific possible future might re-imagine our contemporary moral values, priorities, principles, idioms, or theories. By imagining different possible futures, we can explore the contingent limits of current morality. Such experiments are tentative and fallible, but they still offer valuable moral lessons. In particular, imaging the reactions of particular individuals in some specific future forces us to ask how (if at all) we might justify ourselves to them. This introduces a second-personal urgency into the otherwise abstract topic of intergenerational justice.

II

Outline of The Book

The book consists of lectures delivered by a philosophy teacher in the future broken world. The topics addressed are

those found in standard contemporary introductory courses in moral and political philosophy. (The book is thus designed to serve as a text for such a course, especially for students with a particular interest in environmental or development issues.)

An introductory lecture highlights those features of our present-day affluent world that would seem strange to a visitor from the broken future. These include natural abundance, social affluence, and stable climate; in addition to the distinctive features of contemporary philosophy.

The rest of the book is divided into four parts: libertarianism, utilitarianism, social contract, and democracy. Each part begins with several chapters outlining the relevant strand of affluent philosophy, and ends with a chapter applying the theory to the broken world. These final chapters sometimes take the form of dialogues between students. These dialogues are designed to represent the variety of different ways that future people might respond to affluent philosophy.

Part One deals with rights, focusing on a close reading of Robert Nozick's *Anarchy, State, and Utopia*.[3] 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)."[4]

To illustrate the difficulties facing libertarianism in a broken world, consider the proviso that Nozick borrows from Locke—

[3] R. Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974).
[4] T. Mulgan, *Ethics for a Broken World*, 22-23.

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.[5] 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 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.

Part One ends with a discussion of nationalism. To provide a stark contrast with the contemporary world, I stipulate that national boundaries have collapsed in the broken future. I then ask how future philosophers might interpret present attempts to justify national entitlements.

Part Two covers the central topics in contemporary utilitarianism: act utilitarianism, rule utilitarianism, theories of well-being, and the connection between liberty and utility (exemplified by J. S. Mill). It also includes a chapter on obligations to future people, focusing especially on Derek Parfit's seminal *Reasons and Persons*.[6] While this topic is not always covered in introductory ethics classes, it has been the subject of considerable recent debate. And one theme of the book is that future people, living with the consequences of our failure to

[5] R. Nozick, *Anarchy, State, and Utopia*, 174-182.
[6] D. Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984).

protect their interests, will take intergenerational obligations much more seriously than we do.

A theme of Part Two is that, although utilitarianism has the flexibility to adapt itself to new circumstances, affluent utilitarian defences of liberty and moderation will be hard to sustain in a broken future. Utilitarians may be forced to choose between their utilitarianism and their liberal moderation.

Part Three begins with the founders of social contract theory—Thomas Hobbes and John Locke. The future philosophy teacher suggests that, in a broken world, these pre-affluent thinkers may seem more relevant than their more affluent successors. The central focus is on Rawls, who is the most significant recent contract theorist. Following the pattern of Part Two, Part Three includes a separate chapter on Rawlsian intergenerational justice. Rawls is explicit that his liberal theory of justice applies only under favourable conditions. Survival bottlenecks and survival lotteries have no place in Rawls's justice as fairness. However, the final chapter in Part Three explores several ways that a suitably modified liberal contract-based theory could make sense in a broken world.

The book concludes with two brief chapters on democracy. Most contemporary political philosophers (apart from libertarians) assume that democracy is the best way to resolve political disagreements. To interrogate this presumption, I stipulate that broken world societies are not democratic, and that their inhabitants are suspicious of democracy precisely because it lead to their broken world. Part Four thus asks whether affluent arguments for democracy are strong enough to persuade a sceptical outsider, especially one committed to intergenerational justice.

III

General Themes of The Book

Several overarching themes emerge from the future teacher's survey of affluent philosophy. First, some central affluent moral concepts are very hard to translate to the broken world. The most obvious is our affluent notion of *rights*. Rights are not much use if you cannot stay alive. We are rightly suspicious that the rights of the rich will trump the needs of the poor. So many rights theorists defend rights to subsistence, to have your basic needs met, to be provided with a decent standard of living, or basic education, or adequate health care, and so on. But in a broken world, we cannot meet all basic needs. So we cannot honour all these positive rights. And basic needs also inevitably conflict with other rights. In a broken world, *any* inefficiency in food production leads to starvation, as does *any* diversion of economic activity to produce luxuries rather than necessities. A broken world may require restrictions on personal liberty on a scale that people have only previously accepted in times of war, or other temporary crisis. Private land might be requisitioned to grow food, as might individual labour; the use of fossil fuels for private purposes might be severely curtailed; and individual lifestyle choices—especially reproductive decisions—might be much more tightly regulated and constrained.

A related theme is that the liberalism and moderation characteristic of much affluent philosophy is hard to defend in a broken world. This is most obvious in Part Two, where the broken world students interrogate the moderate liberal utilitarian tradition of J. S. Mill. But moderate forms of libertarianism, nationalism, and contractualism face similar difficulties. Foundational moral theories that yield comfortable results in an affluent world can become extreme and illiberal in a broken one.

Finally, when it comes to intergenerational justice itself, the book suggests that all affluent theories rely, in different ways, on the complacent assumption that obligations to future people can be ignored because future people will be better-off than present people. Once this presumption is removed, intergenerational issues become both much more urgent, and much harder to address.

University of St Andrews

&

The University of Auckland

If you need to cite this article, please use the following format:

Mulgan, Tim, "A précis to *Ethics for a Broken World*," *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014), 3-10, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

MORAL AND POLITICAL PHILOSOPHY FOR A BROKEN WORLD?

DEMOCRACY AND FUTURE GENERATIONS

BY

BEN SAUNDERS

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 11-28
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Democracy and Future Generations

Ben Saunders

It is sometimes suggested that democratic principles require that everyone whose interests are affected by a decision should be included in the decision-making demos. For instance, Kristian Skagen Ekeli writes that "everyone whose living conditions and life prospects are seriously affected by a collectively binding decision, should also have the opportunity to influence the decision process and participate or be represented in the making of that decision."[1] Lecture 17 of Mulgan's *Ethics for a Broken World* begins with an observation and two important questions: "present decisions always *impact* (often very seriously) on future people. Not everyone *affected* by affluent "democratic" decisions was able to vote. But was this really democratic? Did affluent democracies treat future people justly?"[2]

These two questions need to be taken separately, since it is possible, firstly, for democracies to act unjustly and, secondly, for

[1] K. Ekeli, "Constitutional Experiments: Representing Future Generations Through Submajority Rules," *The Journal of Political Philosophy* 17 (2009): 440-61, at p. 445.

[2] T. Mulgan, *Ethics for a Broken World: Imagining Philosophy After Catastrophe* (Montreal & Kingston-Itacha: McGill-Queen's University Press, 2011), p. 211. Subsequent unattributed page references are to this book.

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 11-28
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

non-democratic arrangements to treat people justly.[3] I shall argue that there is nothing *undemocratic* about excluding future generations. Whether we treat them justly is a more difficult question, which I do not answer. If we are unjust to future generations though, this is not because of our democratic institutions. While these institutions do not guarantee justice, there is little reason to think that alternative institutions would perform better on this score, and some reason to think that they may do worse; a (liberal) democracy may offer better prospects of justice, for both current and future generations, than likely alternatives.

I

The All Affected Principle

Democracy means the rule of the people. *Which* people, however, is far from obvious. Until comparatively recently, it was commonly assumed that 'the people' were defined by national boundaries, but this is problematic. First, there are various questions about who 'the British people' are. But, even assuming that these are resolved, there are deeper questions about what the British people have the right to decide. If their decisions affect only themselves, and not outsiders, then there seems no reason to object. Matters are less clear when the decisions of one 'people' profoundly affect other peoples, such as neighbouring states. For instance, if the British build a coal-fired power station it will

[3] It has been argued that democracy is a requirement of justice; for example T. Christiano, *The Constitution of Equality* (Oxford: Oxford University Press, 2008). If this is so, then it would appear that future people cannot be treated *fully* justly. However, I assume we can still distinguish treatment that is just and unjust in other respects.

contribute to atmospheric pollution affecting Sweden. What gives the British the right to make this decision?

The all affected principle says that all of those affected by a decision should be included in the decision-making body (demos). This has some intuitive appeal.[4] One reason for extending the franchise to 'the people' is so that the people can protect their own interests from predatory rulers. Thus, we might think it natural that all whose interests are affected should be given the opportunity to protect their interests. Nonetheless, the all affected interests principle faces a number of difficulties which, I believe, should lead us to reject it.

Note that I do *not* reject the starting assumption, *viz.* that there is something wrong with the British people unilaterally imposing pollution on the Swedish. I can accept that this is unjust, but I think that the all affected principle fails to explain what it wrong here. The wrong is that the Swedes are unjustly harmed by the decision, not that they were not enfranchised in the making of it. Enfranchising them would not necessarily have legitimised the decision, assuming there are losses that a majority cannot permissibly impose on a minority even democratically. It may be suggested that excluding the Swedes represents an additional wrong; for the UK to pollute without enfranchising the Swedes would be more wrong than to do so after a vote in which they were included. But this further wrong is less obvious. I think our intuition of wrongdoing can be explained simply by the illegitimacy of the harm.

Before criticizing the all affected principle, however, it is necessary to state it more precisely. As Goodin observes, any

[4] This appeal may be partly due to the principle's lack of determinate content. I assume that it is most plausible when interpreted as 'all affected interests' (as opposed, for instance, to all those causally affected). Even so, its implications rest upon a theory of interests, which will not be developed here.

application of this principle must resolve a number of indeterminacies.[5] For instance, we cannot enfranchise only those *actually* affected by a decision, since who is affected by a decision will depend on what is decided which, of course, will depend on who is included in making it. Since any decision may have far-reaching consequences, Goodin argues for an expansive reading of the all affected principle, according to which we should enfranchise all whose interests are *possibly* affected by any possible decision. The easiest way to do this, he adds, is to enfranchise everyone on every decision.

We might add one other possibility that Goodin does not explicitly address. We might ask whether only actual (present or future) persons should be enfranchised or whether all *possible* persons ought to be enfranchised. I do not see why we ought to enfranchise an actual person whose interests are only possibly affected, but not a merely possible person. One obvious reason to exclude the latter would be that their interests are less likely to be affected, since it is only possible that they will even exist and thus have interests to be affected. But there may be possible people who will certainly be affected *if* they exist, so the possibility that these people will actually be affected may be no less than the possibility of other, actual people being affected. Moreover, in Goodin's view the likelihood of being affected does not matter. People should not be disenfranchised simply because the chance of their being affected is small. If we want the demos to be maximally inclusive, then it seems that not only all actual people (present and future) ought to be included, but even *possible* people, who may never exist, should also be included, since their interests may be affected if they exist. It may be responded that only actual people are of moral concern, so the interests of merely

[5] R. E. Goodin, "Enfranchising All Affected Interests, and its Alternatives," *Philosophy & Public Affairs* 35 (2007): 40-68.

possible people do not matter. My argument below does not assume the inclusion of possible people; I merely wish to highlight the difficulties in identifying who is affected.

II

Against the All Affected Principle

Advocates of the all affected principle claim that it allows those affected by a decision to protect their own interests. Democracy is not simply about voting though; it also involves deliberative justifications. If Swedes are enfranchised, then they can call upon Britons to justify their decision. The British might simply say 'it's in our interests, and we don't care about you', but this is unlikely. Public deliberation usually forecloses certain lines of justification, including such appeals to naked self-interest. The aim, then, is that deliberative inclusion will lead to better (more just) outcomes, since voters will be 'forced' to consider the interests of others. This is a worthy aim but, I think, the wrong way to go about achieving it.

Having to justify ourselves to others makes it more likely that we will take their interests into account. It does not, however, require that we think of them as entitled to inclusion in the decision-making body. The all affected principle may fit with some of our intuitions, but it is radically out of keeping with other intuitively acceptable practices, as I have argued elsewhere.[6]

[6] See my "Defining the Demos" *Politics, Philosophy and Economics* 11 (2012): 280-301 and related arguments developed in my "Democracy, Rights and Immigration," *Theoria: A Journal of Social and Political Theory* 58 (2011): 58-77 and "Scottish Independence and the All-affected Interests Principle," *Politics* 33 (2013): 47-55.

We ordinarily assume that individuals have the right to make certain decisions even if they affect others. Consider a woman contemplating a divorce; this decision will undoubtedly affect her husband, any children, and perhaps many others (the husband's mistress, for instance). We do not ordinarily think that these others have a right to be included in the decision-making. Perhaps she ought to consider their interests in deciding, but this is very different from saying that they ought to be joint makers of the decision. The decision is hers – and hers alone – to make, but we hold her responsible for the effects of her choice on others.

This decision is not a democratic one, but that does not mean that it is irrelevant here. If we think that the woman is entitled to decide unilaterally, considering but not including affected others, then it shows that decisions need not include all affected parties in order to be legitimate. Thus, either i) democracy is not necessary for legitimacy or ii) including all affected persons is not necessary for democracy. Note that this applies even if others are left worse-off as a result of her decision.

This lesson can be extrapolated from the individual case to group cases. Members of one group might unilaterally decide to stop purchasing goods that they were previously buying from another, even though that decision may have profound affects upon the former supplier. Similarly, I suggest, whether the British people wish to build a coal-fired power station is their decision. If this decision imposes harms upon others, such as Swedes, then we might hold the British people responsible, and perhaps even require them to pay compensation, but this does not require us to say that they ought to have included the Swedes in the decision-making body.

I have said little about future generations, but I hope it is obvious where the argument leads. The mere fact that some people are affected, even negatively, by a decision does not in

itself show that they must be included in making it. Either there is nothing undemocratic in this or we sometimes have the right to make certain decisions non-democratically. If we accept this, then it presumably applies to intergenerational decisions. The present generation has the right to make certain decisions, even if those decisions may impact negatively on future generations. It does not follow that the present generation may do anything that they like, since we may still criticize their decisions as unjust. Nonetheless, these decisions are not illegitimate simply in virtue of excluding future generations.

Whether a given decision is just depends on a theory of intergenerational justice. *Ethics for a Broken World* critically examines several leading contenders, including moral theories based on natural rights, utility, and a hypothetical social contract. Perhaps none of these are satisfactory, but let us imagine that an acceptable theory has been found, perhaps a rule-consequentialist theory of the kind that Mulgan has developed elsewhere.[7] We can now judge whether or not the democratically-authorised decisions of the present generation treat future generations justly. Perhaps the answer is negative. What should we do in the face of democratic injustice? Can we modify democratic procedures to reconcile them with what justice requires?

III

Democracy and Trade-offs

Democracy can take a variety of forms; for instance, we may or may not have a second legislative chamber, an entrenched

[7] T. Mulgan, *Future People: A Moderate Consequentialist Account of our Obligations to Future Generations* (Oxford: Oxford University Press, 2006).

constitution, proportional representation, etc. Some of these mechanisms may be more democratic than others, but even if judicial review is less democratic than unconstrained majoritarianism (for instance), this does not mean that it should be rejected. I assume that democracy is the only justifiable form of government in our circumstances, but this does not tell us *how* democracy ought to be weighed against other values.

Perhaps there is a sense in which it would be more democratic if decisions over my private life, such as what religion I should practice, were taken out of my hands and decided by a vote with universal franchise. But, if this is so, then giving me rights over my own private life is a justified departure from democracy. Similarly, if democracy requires a universal franchise (at least among those affected by a decision), then democracy might require us to enfranchise even young children. Restricting the vote to those over, say, sixteen would, on this view, be less democratic, but might still be justified by other considerations. In other words, we need not assume that democracy is something that must be maximized in order for political decisions to be legitimate.

Denying that we must *maximize* democracy is not to say that some non-democratic regime is preferable. But, within the class of reasonably democratic regimes, we may justifiably opt for one that is less democratic than alternatives if it better realizes other values. Thus, when considering questions of institutional design, such as whether to adopt entrenched rights, our argument need not be confined to whether or not such a measure is more or less democratic than unconstrained majoritarianism. We may prefer less democratic institutions if they better realize justice.

IV

Democracy and Constitutionalism

One of the most obvious choices faced in designing democratic institutions is whether to include entrenched constitutional rights, as in the USA, or whether to trust in unchecked democratic decision-making, as in the UK.[8] It is often assumed that constitutional rights, enforced by unelected judges, are a departure from democracy. But, even if constitutional rights represent departures from democracy, they may be justifiable where they better secure just outcomes. Both constitutional and non-constitutional regimes are within the range of democratic possibilities identified in the previous section, so our choice between them should depend on which better serves other substantive values.

We might expect a constitutional regime to better protect rights than a non-constitutional one, since the point of taking these rights out of ordinary democratic decision-making is to protect them. If the majority can do whatever they like, then nothing stops them tyrannizing over a minority. To give the minority rights is to give them a 'trump' card,[9] with which they can override the ordinary process of democratic decision-making. This allows the minority to protect themselves from what the majority might otherwise do. But this overlooks the fact that there is likely to be disagreement not only over what to do, but also over rights.

[8] This position is often described as unchecked majoritarianism, but democratic decision-making mechanisms need not be majoritarian. See my "Democracy, Political Equality, and Majority Rule," *Ethics* 121 (2010): 148-77. Nonetheless, I shall speak of majoritarianism for simplicity.

[9] R. Dworkin, *Taking Rights Seriously* (London: Duckworth, 1977).

It is all very well saying that rights ought not to be determined by the majority, if their purpose is to check that very same majority, but they must be determined by someone. Constitutional regimes give the power to determine rights to judges, but this means that a panel of unelected, unaccountable, and (usually) unrepresentative individuals have the power to frustrate the will of the people. Judicial activism has sometimes been a force for good, as in the case of civil rights in the USA, but note that judicial review did nothing, for a long time, to prevent racist segregation and even slavery. Furthermore, it may have been easier to abolish unjust practices had a simple majority been empowered to do so, without the need for a constitutional amendment. Constitutional checks, by privileging the status quo, can be impediments not only to injustice but also to just reforms. While placing the rights of minorities in the hands of a simple majority carries obvious risks, there are also risks attendant to judicial review. We need some reason to believe that judges will do a better job of protecting the rights that minorities should have than majorities will.

There is a considerable literature on the merits of constitutional rights, but not all of these arguments apply to future generations. For instance, Anthony McGann argues that minorities are better protected by simple majority rule than supermajority rules.[10] This is because winning majorities are usually loose coalitions of minorities, so even a small minority may hold the balance of power. If a given minority feel harshly treated by the present majority, then they need only join enough others to become part of a new winning coalition. If they are presently treated badly enough, then they will be willing to 'sell' their support cheaply, so should have little trouble finding

[10] A. McGann, *The Logic of Democracy: Reconciling Equality, Deliberation, and Minority Protection* (Ann Arbor: University of Michigan Press, 2006), chapter 5.

coalition partners. According to McGann, minorities are better protected if only a simple majority is needed to form a winning coalition, since this makes it easier to do so. Conversely, if a supermajority is required, then the minority need to find more willing coalition partners.

This argument suggests that the 'instability' arising from Condorcet cycles may be a good thing, as it allows minorities to protect their fundamental interests by joining new winning coalitions against previous oppressors. The extent to which this happens depends on contextual features of the political culture. It must be the case that others are willing to enter into coalition with the minority in question, so this offers no protection to a despised minority that others refuse to bargain with, even on favourable terms. Furthermore, only groups with a vote stand to gain from the potential to join coalitions in this way. This argument gives us no reason to think that outside groups, such as foreigners or future generations, are better protected by simple majority rule. Future people are unable to join a present coalition because they do not yet exist.

This example highlights the presentism of contemporary debates about constitutionalism. Most of the arguments focus exclusively on what institutional arrangements protect the rights of present people. But, even if there was universal consensus on *this* question, it would still be an open question whether these arrangements also treat future people justly. We might think that, since future people do not have votes, they need some other extra-democratic protection, of the sort that might be afforded by judges. But, again, this would be too quick. There is no reason to assume that unelected judges would better protect the rights of future people than the electorate as a whole would.

Jeremy Waldron has argued that rights are only necessary as checks on majoritarianism if we assume that the majority will vote

in a self-interested fashion that ignores the interests of others.[11] If we suppose, instead, that voters will aim at something like the common good then they will already take the interests, and rights, of others into account in their deliberations. This, on its own, seems to show that constitutional rights may be *unnecessary*, but not that they would be positively bad. We could supplement Waldron's argument, however, by speculating that, in a society where judges are expected to protect rights, citizens may feel it less incumbent on them to attend to the interests of others. They may assume that they have the license to vote self-interestedly, confident that judges will protect justice. This possibility suffices to show that constitutional rights do not necessarily better protect people's fundamental moral rights.[12]

So long as ordinary voters are sensitive to the rights of future people, there is no reason to suppose that these rights will necessarily be neglected. However, one might equally say that 'so long as men are sensitive to the rights (or interests) of women, women do not themselves need the vote.' While we expect (or at least hope) that voters will consider the interests of others, we think that those others ought not to have to rely upon the goodwill of other voters. Women can and should be enfranchised, but this is impossible for future generations. Before concluding I wish to consider several (non-judicial) institutional

[11] J. Waldron, "Rights and Majorities: Rousseau Revisited" in J. Chapman and A. Wertheimer (eds.) *NOMOS XXXII: Majorities and Minorities* (New York: New York University Press, 1990), pp. 44-75.

[12] See Mulgan, *Future People*, pp. 253-4: "It is tempting to assume that, for any conceivable future threat, it would be possible to design a political system that perfectly avoids that threat, by embedding certain measures in its constitution […]. [But] constitutional entrenchment, just like [majoritarian] democracy, is not infallible […]. There is no good reason to believe that the present generation could design a constitutional system which would do a better job of finding the appropriate balance of responses to future threats than open public deliberation at a later date."

mechanisms that have been proposed to protect future generations.

V

Institutional Innovations

Future generations are a problem for democracy because they cannot be enfranchised. Perhaps we can tackle the problem the other way, by disenfranchising at least some of the present generation. Some have called for the disenfranchisement of the elderly.[13] When we are making decisions with profound, long-term effects, it is unsurprising that the elderly may take a different, shorter-term perspective, than younger voters.[14] Obviously, any disenfranchisement of the elderly, even over only a subset of decisions, would surely be controversial. Moreover, as Van Parijs notes, such a move may backfire. While the elderly have less to lose from present environmental degradation, they also have less to gain from present over-consumption.[15] Perhaps, with less personally at stake, elderly voters are more likely to be swayed by ethical concern for future generations, whereas younger voters are more likely to privilege their own lifetimes.

Van Parijs' article surveys several other proposals, stopping short of disenfranchising the elderly, but intended to have much the same effect. One possibility is age-stratified plural voting. We

[13] See P. Van Parijs, "The Disfranchisement of the Elderly, and Other Attempts to Secure Intergenerational Justice," *Philosophy and Public Affairs* 27 (1998): 292-333.

[14] This can also be supported by other arguments, such as one from symmetrical treatment of the young and old; see J. Lau, "Two Arguments for Child Enfranchisement," *Political Studies* 60 (2012): 860-876.

[15] P. Van Parijs, "The Disenfranchisement of the Elderly," p. 323.

might allow that all adults should have *at least* one vote, but think that those with longer to live should have a greater say over long-term decisions that will affect them more.[16] Perhaps those aged between 18 and 27 should get three votes, those between 28 and 55 two votes, and those over 55 only one vote.[17] Relative to 'one person, one vote' this over-represents the young, but the distinction reflects remaining life expectancy: each person should have one vote for each remaining 25 years of life expectancy. Another, more moderate, proposal is asymmetric compulsory voting, where the young but not the old are required to vote.[18] In both cases, political power is shifted from the old to the young, but, again, this only serves intergenerational justice if the young are more sensitive to the needs of future generations.

Assuming we reject these reforms, we may seek to improve democratic performance by altering the decision rule, rather than the electorate. Ekeli suggests that future generations can be protected by sub-majority rules. He proposes that, where a prima facie case can be made that a proposed law can inflict risk of serious harm on future people, then a minority of legislators (one-third) should have the power to block the final enactment of that law until there is either a new election or a referendum on the proposed law.[19] This falls short of an absolute block on proposed courses of action,[20] so is arguably more democratic than constitutionally entrenched rights. Nonetheless, while a majority

16 That the more affected should have more power is suggested by H. Brighouse and M. Fleurbaey, "'Democracy and Proportionality," *Journal of Political Philosophy* 18 (2010): 137-55.

17 P. Van Parijs, "The Disenfranchisement of the Elderly", p. 305.

18 Ibid., p. 306. I have proposed something similar, though with a different rationale, in my "Tasting Democracy: A Targeted Approach to Compulsory Voting" *Public Policy Research* 17 (2010): 147-51.

19 K. Ekeli, "Constitutional Experiments", p. 449.

20 Ibid., p. 456.

may still proceed, if they wish, they are required to pause long enough for concerns about future generations to be heard. The temporary check prevents a majority from acting *hastily*.[21] If, after due consideration, they still wish to press ahead then at least it is clear that this really is the considered preference of a wide section of the population.

This proposal is no guarantee that the rights of future generations will be taken into account but makes it more likely. Moreover, it has the advantage, over proposals such as those considered by Van Parijs, that it does not privilege any particular members of the electorate.[22] While some would prefer to trust the young, or those appointed by environmental pressure groups, to protect future generations, Ekeli's proposal allows *any* significant minority of legislators to act on behalf of future generations. Furthermore, while entrenched constitutions or attempts to disenfranchise some sections of the population may justifiably be regarded as departures from democracy, Ekeli's proposals can be construed as offering *more democracy* as a solution to the problem of future generations. A minority may temporarily frustrate the wishes of a majority, but only long enough for democratic deliberation.

What mechanisms best protect future generations is an empirical question and the answer is likely to differ in different circumstances. My point, however, is that there is no need to abandon democracy in order to protect future generations, since democracy has the necessary resources to take future generations into account. I shall conclude by considering the alternative.

[21] Ibid., p. 456.
[22] Ibid., p. 453.

VI

If Not Democracy, Then […] What?

The people of Mulgan's broken world regard democracy as "dangerous, future-destroying anarchy" and "dare not even ask whether democracy has any place in [their] broken world" (p. 199). But, if we reject democracy, what is the alternative? We do not know what form of government the broken world operates under, though presumably power is concentrated in the hands of relatively few people.

One traditional argument for monarchy was that it would serve the long-term interests of the nation. Elected politicians rarely look beyond the next election, but a lifetime monarch – assuming that s/he wishes to rule over a prospering nation – has reason to look to the longer term when making decisions. Further, in a hereditary system, s/he will presumably want to bequeath a flourishing kingdom to his/her descendants. Thus, monarchs may better serve the people's long-term interests than elected politicians. But this is an idealizing argument. We may agree that an *ideal* monarch would be better than actual democracy, but we must either compare two well-functioning ideals or, perhaps more relevantly, two realistic likelihoods. Maybe the best that we can say for democracy is that, even when it operates imperfectly, it is generally less bad than one person's dictatorship.

Let me close by noting that whatever political system operates in Mulgan's broken world is also biased towards their present. As noted in the introduction, "whatever we do, our descendants cannot hope to enjoy even the quality of life that we ourselves take for granted […]. Although we show far greater concern for our descendants than affluent people did, we still tend to keep a disproportionate share of resources for ourselves, sacrificing our

descendants to save our contemporaries" (p. 11). Societies in the broken world must, through necessity, employ 'survival lotteries' to determine who lives and who dies when times are hard (pp. 10-11).[23] Though the details are unclear, it sounds as if as many present people as possible are saved, even if this predictably leads to more deaths among future generations. The people of the broken world get by the best they can, but they, like us, leave their descendants a worse world than the one that they inherited.

Those in the broken world might reply that we are the ones responsible for their dire circumstances, but that one generation has been wronged by its predecessors does not mean that they are entitled to rob their own descendants to make up for the shortfall in their inheritance. The non-democratic arrangements in place in the broken world are, like our affluent democratic institutions, regimes to manage the world *for present people*. The harsh realities of a broken world may make non-democratic government necessary, but this does not mean that such government is any better at respecting the rights of *future* people.

What is true of future people seems true for us too. Suppose that some of us are concerned that our societies are acting unjustly towards future generations. Assuming that we cannot stage a *coup d'état* and impose ourselves as philosopher-rulers for the greater good, what options do we have? Either power is held by the masses or it is held by a few. Neither group is accountable to the future, so neither can be assured to treat future generations justly, but at least a (liberal) democratic society allows us to speak

[23] It seems that these are not *literally* lotteries: "most societies distribute food partly on the basis of age or health, so that people are not kept alive once they can no longer make a productive contribution" (p. 10). On the nature and justification of lotteries, see G. Sher, "What Makes a Lottery Fair?", *Noûs* 14 (1980): 203-216, my "The Equality of Lotteries," *Philosophy* 83 (2008): 359-72, and P. Stone, *The Luck of the Draw: The Role of Lotteries in Decision Making* (New York: Oxford University Press, 2011).

on behalf of future generations. As the history of non-democratic regimes has shown, once we abandon democracy, there is no guarantee of the right to dissent, much less of being heard. Democracy may not be perfect, but it is arguably less dangerous to the future than any feasible alternative.

University of Southampton

If you need to cite this article, please use the following format:

Saunders, Ben, "Democracy and Future Generations," *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014), 11-28, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

MORAL AND POLITICAL PHILOSOPHY FOR A BROKEN WORLD?

THE FUTURE-PERSON STANDPOINT

BY

TIMOTHY CHAPPELL

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 29-46
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

The Future-Person Standpoint

Timothy Chappell

> The more the separateness and differentness of other people is realised, and the fact seen that another man has needs and wishes as demanding as one's own, the harder it becomes to treat a person as a thing[*]

The following, I take it, is an exceedingly unpromising form of moral argument:

> *In desperate circumstances* B@ [to be pronounced 'the broken world'], X would be right;
>
> Therefore X is right.

It's unpromising, obviously, because the argument says nothing about why we should care what *would* be right *if* B@ obtained, if B@ doesn't obtain. A second form of moral argument is little better:

> In desperate circumstances B@, X would be right;
>
> *B@ is possible*;
>
> Therefore X is right.

[*] Iris Murdoch, *The Sovereignty of Goodness* (London: Routledge, 1970), 66.

 Philosophy and Public Issues (New Series), Vol. 4, No. 2 (2014): 29-46
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

This argument is unpromising because the premises only imply that X is *possibly* right. (Though perhaps, if B@ is possible, the best response to these premises is not to infer this conclusion, but to try and seal off B@'s possibility.) A parallel criticism holes a third kind of bad argument:

In desperate circumstances B@, X would be right;

B@ is future;

Therefore X is right.

Here the premises seem to imply only the conclusion that X *will be* right. (Though perhaps, if B@ is desperate, the most sensible response to these premises is not to infer this conclusion, but to try and avert B@ from being future.)

I'm afraid I suspect arguments of these three unpromising forms do, in practice, influence us more than they should. That can be the rhetorical or emotional effect on us of watching too many disaster movies, or of contemplating too many doomsday scenarios. It may perhaps also be the effect of thinking about too many far-fetched philosophical examples, quite a lot of which seem—in practice at least—to be deployed to smuggle past us unpromising arguments of these or similar forms.[1]

[1] In Anscombe's famous words: "the point of considering hypothetical situations, perhaps very improbable ones, *seems to* be *to* elicit from yourself or someone else a hypothetical decision to do something of a bad kind. I don't doubt this has the effect of predisposing people—who will never get into the situations for which they have made hypothetical choices—to consent to similar bad actions, or to praise and flatter those who do them, so long as their crowd does so too, when the desperate circumstances imagined don't hold at all." G.E.M. Anscombe, "Modern Moral Philosophy," *Philosophy* 33 (1958), 13.

Central to Tim Mulgan's *Ethics for a Broken World* are two forms of argument which (you'll be relieved to hear) I don't think are any of the above.[2] One is this:

> In desperate circumstances B@, X would be right;
>
> but X looks *obviously* wrong to us, or to most of us (suppose e.g. X is one of the survival lottery arrangements that Mulgan describes at EBW 10-11);
>
> So maybe we should consider the possibility that our objections to X are less absolute than we take them to be.

And the other is this:

> In affluent circumstances A@ [the affluent world], X looks *obviously* right to us;
>
> *but X would be wrong in desperate circumstances B@*;
>
> So maybe we should consider the possibility that our acceptance of X in A@ is less defensible than we take it to be.

As Mulgan himself begins by (in effect) telling us, one guiding thought behind both these arguments is a thought about *contingency*, parochialism. As Harold Macmillan famously put it, we've never had it so good: we rich westerners occupy—and it is one of the great merits of Mulgan's book to remind us of this fact so forcefully—a highly unusual and no doubt strictly temporary position in history, the position of affluence. Our perspective is, historically speaking, a most unusual perspective. What reason is there to expect it not to be also a *warping* perspective? What might

[2] T. Mulgan, *Ethics for a Broken World: Imagining Philosophy After Catastrophe* (Montreal & Kingston-Itacha: McGill-Queen's University Press, 2011). Henceforth EBW.

political and moral philosophy look like, if we tried to get free of the distortions of that perspective? If we tried, say, to adopt the perspective of a broken world's inhabitants instead, and think about how moral assumptions that seem entirely natural to us might strike them?

This line of thought suggests that either or both of two interesting general theses about moral argument may also structure Mulgan's project in EBW—two master-theses, as I shall rather grandiloquently call them. The first master-thesis is the less ambitious of the two. It has a rather contractualist air. We may hear echoes in it of both Rawls and Scanlon, not to mention echoes of the Golden Rule:

> MT1 For an arrangement to be justified, it must be justifiable to the person who does worst out of it.

The second and more ambitious master-thesis is a ringing denial of parochialism:

> MT2 For an arrangement to be justified, it must be absolutely justifiable—justifiable irrespective of any kind of contingency (such as the contingent circumstances of our own affluent society).

One interesting way to take MT1 is to take the 'arrangements' of which it (with studied vagueness) speaks as being dispositions. So understood, MT1 gives us a framework within which to address the question of how the virtues are to be justified. For justice, say, or courage to be a virtue, it must be true that justice or courage makes sense not only for those who benefit from justice and courage, but even for the person who does worst out of it, or in the situation in which we in general do worst—or more broadly in situations where humans do very badly.

One reason why it would be interesting to pursue this line is because it gets us thinking about dystopias, a very interesting moral topic indeed. Take the *Melancholia* dystopia, the situation in which everyone is depressed and a collision with a rogue planet is about to incinerate all terrestrial life. Do the virtues, as traditionally conceived, make sense there?

Or take the *Never Let Me Go* dystopia, the situation in which you and your friends turn out to have been cloned to provide non-cloned members of society at large with a source of donor-organs. Do the traditional virtues make sense there?

Or the *Hunger Games* or *1984* or *Brazil* or *Matrix* or (no fictionalising italics, alas) 1930s-Germany dystopias: same question. (And so on for various dystopias.)

In all these cases, as it happens, I think the answer is yes: despite everything that is so horribly lost in these various awful cases, the traditional virtues do make sense even in these scenarios, simply because living according to the virtues goes on being the best way to live *no matter what may come*. The reason why this is so, or part of it, is that the virtues are not means to a further end called flourishing: if they were, then in these dystopias, where by definition flourishing of a further-end sort is unattainable, the virtues could not possibly achieve their end. Rather, living according to the virtues, even in situations where misery or annihiliation is inevitable, is *itself* a fulfilment of flourishing; only living according to the virtues has that particular shine of genuine admirability or beauty that Aristotle calls *to kalon*. The striking thing about the virtues on the traditional lists is that it is they and things like them, and only such things, that pass the test that is set by thinking about such awful dystopias. That, of course, is why these are the *virtues*.

This line of argument about the virtues is an interesting one and a venerable one, but it is not the main thing I want to pursue here. What I most want to talk about is, as my title suggests, a different line of thought suggested by MT1. This has to do with what, as a small kind of homage to Stephen Darwall, I'll call *the future-person standpoint*. I turn to this now.

Take MT1 to be about our present political and moral arrangement: the institutional, economic, social, and personal set-up of our society. And take the future people of B@ to be, collectively, the people who do (will do) worst out of that arrangement. Then the basic form of Mulgan's critique of our present arrangement is that it is not justifiable to the imaginable B@ people. So insofar as our present arrangement has the predictable consequence of bringing those B@ people—and B@ itself—into actual existence, it is not justifiable at all.

This, I think, is a very promising form of argument. Very plausibly we might make it a quite general necessary condition of the acceptability of any political or moral proposal, that it should be justifiable to whoever is least benefited/ most harmed by it. Informally, the gist of this line of argument is given by rhetorical questions like 'If we don't look after the world we pass on to our descendants, how could we look them in the eye if we met them?' In a very different application, something like this argument is stated, to the king in disguise the night before Agincourt, by the pessimistic soldier Williams in *Henry V* Act 4 Sc.1:

> But if the cause be not good, the king himself hath a heavy reckoning to make, when all those legs and arms and heads, chopped off in battle, shall join together at the latter day and cry all 'We died at such a place;' some swearing, some crying for a surgeon, some upon their wives left poor behind them, some upon the debts they owe, some upon their children rawly left. I am afeard there are few die well that die in a battle; for how can they charitably dispose of any thing, when blood is their argument?

> Now, if these men do not die well, it will be a black matter for the king that led them to it.

Williams is concerned too about the harrowing prospect of dying at the king's behest in a state of mortal sin—and this is the side of his argument that the disguised Henry himself chooses to take up, possibly because it is the easiest bit to deal with, in a long and highly legalistic response that is bound to strike the reader as, in more than one sense, windy. But it is also at least part of what Williams is getting at here that it is the PBI who are at the sharp end of Henry's wars, so that if those wars are to be justified at all, they must be justified first to the foot-soldiers. (So far as I recall, neither Henry nor Williams takes any account of the French civilians, who are, *sans doute*, even more at the sharp end. Nor come to that do any of the French warriors.)

As for MT2, the prospects for this look less good at first sight than they do for MT1. 'Justifiable irrespective of *any* kind of contingency' is a vertigo-inducing phrase, and it sets a tall order. But MT2 emerges naturally enough from things that Mulgan either actually says or clearly implies. As above, the problem with the kind of justifiability that appeals to us in our affluence is that our affluence is such a historically unusual perspective. No doubt it is too much to hope that we might actually get an account of justifiability that is *completely* free of *every* kind of irrelevant contingency. Still, such an account is not an unreasonable target or ideal. To dismiss MT2 out of hand with the usual glib schtick about 'the impossibility of a view from nowhere' is just boring standard-issue academic relativism.

Looking in a slightly different direction, it may seem tempting to rephrase MT2's talk about 'absolute justifiability' in a way that moves MT2 towards MT1: as justifiability to anyone, hence justifiability even to the person who does worst under whatever arrangement is to be justified. I won't go in that direction here,

for what I take to be a good reason. The reason is that as they stand, MT1 and MT2 usefully pick out different aspects of the task of justification: its *audience* (MT1), and its *conditions* (MT2). In what follows I would like to keep these aspects apart from each other.

A different line of argument begins from what is sometimes called 'the idiot's veto objection.' MT1, recall, says that an arrangement is not justified unless it is 'justifiable to the person who does worst out of it.' But suppose that under some otherwise highly attractive arrangement, the least-benefited person is offered a cogent and convincing justification for it, which he is not impressed by. According to MT1, we might then say, the arrangement, no matter how attractive, and no matter how cogent the justification, can't be justified; for its chief victim refuses to count it as justified. MT1 will then have the unappealing consequence that it takes the side of the obstinate and stupid against arrangements that, though inconvenient for them, are clearly best overall. MT1, it seems, gives a veto to idiots, or a charter to dogs in the manger.

This can be countered in a way familiar from discussions of Scanlon's famous clause about *reasonable* rejection.[3] According to this counter, the question is not whether the person least-benefited *does* reject the attractive arrangement. It is whether he *can reasonably* reject it: whether he is *entitled* to say that the justification he has been offered is inadequate.

The trouble with this counter (it is also fairly standard to say) is that now all the questions that we started off hoping to answer using Scanlon's formula, or the present formula, reappear as questions about what rejections are *reasonable*, or about what

[3] T.M. Scanlon, *What We Owe To Each Others* (Cambridge (MA): Harvard University Press, 1998).

justifications *ought to be* accepted. The formulae promised to give us some informative and useful grip on the normative. But we have no account, so far, of the nature of the reasonable, or of what ought to be accepted. And to give an account of either seems pretty well to start from scratch on the task of building an account of the normative. So it seems that the earlier promise of the reasonable-rejection formula, or of the present justifiable-to-the-worst-off formula, now turns out to be illusory, and that being informative about the normative is no closer now than it ever was. In the end—so runs this line of response—MT1 leads us no less into hopeless vertigo than MT2 does.

I have already criticised certain familiar responses to MT2 as giving up too easily, and I think the same criticism applies to the present line of response to MT1. That this line of response goes wrong *somewhere* should be clear from the fact that it ends up in an obviously wrong place. For it ends up saying that the trouble with Scanlon's reasonable-rejection formula, or with the present formula MT1 about justifiability, is that it has no account of what counts as reasonable, or as a justification that ought to be accepted. But this is an unfair demand. There is no ground at all to expect either formula, in and of itself, to provide any such account. (This is one reason why so much of Scanlon's book is, as students reading it often complain, 'about other things': that is, it is not about Scanlon attempting to spin an account of what rejections are reasonable out of the mere idea of reasonable rejection. He doesn't do that for the good reason that he sees that it can't be done.)

What is worth keeping hold of in MT1—and no doubt in Scanlon's formula too—is something different. It is not that either of these formulae gives an *a priori* handle on the notion of 'the reasonable' or 'the justifiable.' Very differently from that, the main point of MT1 is to set *a condition of second-personality* on moral

and political justification. To justify some arrangement is not a disengaged exercise in *a priori* cerebration that I might as well conduct all on my own. For some 'you' who is affected by that arrangement and some 'me' who is the defender of the arrangement, my task is to use whatever resources I have in the way of an account of what is reasonable and why, not to demonstrate the justifiability of the arrangement in the abstract, but to justify it to *you*.

What difference does it make to apply this condition of second-personality to political justification? As Stephen Darwall has brought out in his wonderful recent book *The Second-Person Standpoint*,[4] it makes all sorts of differences to see our political, moral, and other public decisions, not just as decisions of a consequentialist sort to promote or honour the impersonal good or goods, nor again just as decisions of a non-consequentialist sort numbly 'to do our duty' against the vague sepia background of The World In General, but also as needing to be backed by justifications that are addressed, second-personally, to particular other people.

Here is a trivial example of the kinds of difference second-personality can make to moral/ political discourse. (At least at outset, Darwall himself motivates much of his discussion by reference to a different trivial example—the difference between standing on a stranger's foot accidentally, and standing on the foot as it were *at* the stranger, as an insult or a message directed to the stranger as a person.) Compare these three scenes, common enough in the life of a humble long-distance railway traveller like myself:

[4] Stephen Darwall, *The Second-Person Standpoint. Morality, Respect, and Accountability* (Cambridge (MA): Harvard University Press, 2006).

1. You're alone in a railway carriage. It's too hot. The window-blind on the sunny side of the coach could be drawn down to put the interior of the carriage into the shade. Since pulling the blind down will make things better, you pull it down.

2. You're with other people in a railway carriage. It's too hot. The blind on the sunny side of the coach could be drawn down to put the interior of the carriage into the shade. Since pulling the blind down will make things better, you pull it down.

3. You're with other people in a railway carriage. It's too hot. The blind on the sunny side of the coach could be drawn down to put the interior of the carriage into the shade. You point this out to the other passengers, and agree with them that pulling the blind down will make things better. So you pull it down.

The key thing is that in Scenario 2 *you don't ask the other passengers*. Disregarding what they might have to say about it, you just pull the blind down 'because that will make things better.' As I've set up Scenario 2, your action will indeed 'make things better.' But that's not enough to make your action all right. The other passengers in Scenario 2 have a legitimate ground of complaint against you, namely that *you've acted as if they didn't exist.* You decided on and performed an action which affected every person in the carriage as if you were the only person there, and hence the only one who could, or had the right to, decide on and perform such actions. Your action is criticisable on this interestingly second-personal ground: because it evinces a lack of respect for them.

The example is, as I say, trivial, and exceptions or counter-examples to or complications of the moral I want to draw from it are readily imaginable. One type: on occasion I've seen people

take unilateral action with railway-carriage window-blinds, and been, on balance, simply grateful to them for it. In trivial cases like this, one doesn't always *care* if one is simply disregarded. Another type: I've certainly seen people take unilateral action against loud muzak in a bar and been *very* grateful—though this latter case is complicated by the fact that disconnecting the wires to the speakers is likely to be a furtive act. And a third type: another thing that unfortunately happens on public transport sometimes is racist ranting. (There is Youtube footage of one well-known recent case of this in London which led to a prosecution.) In a less well-known case that I witnessed myself, it was striking how one person took it on himself to tell the racist ranter to shut up. Interestingly, the ranter's response to his challenger was 'And who might you be?'[5]—in other words, 'What is your public authority to tell me to shut up?.' Equally interestingly, the other passengers, including me, cheered the challenger for what he had taken it on himself to do—that is to say, we *gave* him a sort of public authority. Effectively, he became the spokesman for us other passengers by *ex post acto* acclamation. That shows that second-personality in public decision-making can often come in by assumption; it is not always a matter of explicit *ante-rem* deliberation. But it doesn't undermine my claim that second-personality is key to public deliberation (and the further claim that I would also want to make—that therefore, second-personality is key to political legitimacy and authority); if anything, the opposite is true.

I hope none of these complications distract us from the valuable point that I think my trivial little example makes, which is that when non-trivial political, moral, or otherwise public actions fail to be appropriately second-personal, this can be a *deep* failing in those actions. In many, perhaps most, significant public

[5] That at any rate—to quote Alan Bennett—was the gist of his response.

actions, it is simply *wrong* to act as if it were 'just the world and me;' as if there were nothing to consider except my own agency on the world, and the impact on the overall goodness of states of affairs that my agency can bring about—as if I were, as Bernard Williams once famously put it, simply "the janitor of the impersonal utility system."[6] What is missing from such pictures of (much or most) public action is the important place in generating it of public or shared deliberation, the consultation of others based on the recognition that the decisions I am proposing to take are not just *my* decisions but *our* decisions. They are decisions in which those others have just as much stake and say as I do, and on which they have an equal right to my own to be recognised as deliberators.

Here now are two interesting misunderstandings of this picture of shared deliberation, and one interestingly correct understanding of it. First, the picture is misunderstood if it is taken as simply registering an instrumental claim rather than a constitutive one. It isn't just that we should consult others on public decisions because, if we don't, they will protest that they have a right to be consulted, and their protests will lead to inefficiency. Rather, the point is simply that they *do* have a right to be consulted. Hence if I make a choice not to consult them, or to disregard what they say when consulted, usually I am not incurring a cost that can be offset in the familiar utilitarian-calculus way against the possible benefits that it generates, e.g. the avoidance of dog-in-the-manger or idiot's-veto problems. It is—usually—more like there is something wrong in the whole way I frame my choosing. The question of the costs and benefits arising from a public choice only comes up once we have already acknowledged that we need to make the choice together: that we

[6] B.A.O. Williams, *A critique of Utilitarianism* (Cambridge: Cambridge University Press, 1973), p.118.

need to be dialogical, second-personal, in our approach to our public decision-problem. Failing to see this is going wrong all right, both morally and rationally. But it is not the same kind of going wrong as going wrong in our cost-benefit analysis.

This does not mean—as a second misunderstanding has it—that a commitment to second-personality in public decision-making has to be *absolute*, in the sense that I can never ever refuse to go on trying to deliberate together with someone else because my interlocutor, the other person, is being manifestly unreasonable. *Of course* I can do that. (So can my interlocutor, if *I* am being manifestly unreasonable.) But it does mean that such refusals should not be parsed as utility-based. My reason for refusing to go on deliberating together with this particular idiot or that particular dog-in-the-manger is not that my interlocutor's unreasonableness is a threat to utility. Rather, the reason is simply my interlocutor's unreasonableness. (After all, that unreasonableness could be objectionable even if it was obviously no threat whatever to utility.) Once again the point is constitutive, not instrumental.

Finally, then, the interestingly correct understanding of this picture of shared deliberation is this: it is non-utilitarian and non-maximising. Certainly such deliberation will aim for optimality in *some* respects, in particular those identified by MT2: as far as possible it will try to throw off every kind of irrelevant contingency or appeal to special interests, and will insist on repeating and repeating the challenge, to all those involved as deliberators, that they should keep inspecting their own motives and intuitions, to make sure that these are not polluted by what Iris Murdoch beautifully calls 'the fat relentless ego,' by special pleading or other forms of covert self-interest. Still, the deliberation will not try to bring about that mythical thing 'the best possible state(s) of affairs.' The reason why not is obvious

from what I've just said: *even if* our shared public deliberation is entirely cost-benefit in form once it begins (a condition which is highly unlikely to be satisfied in practice), still there are key non-utilitarian conditions about respecting our interlocutors as interlocutors which need to be in place before our deliberation can so much as begin.

Another way to put this is to go back to the railway carriage, and see that the right thing to do about the window-blinds is not necessarily what—as the usual highly misleading phrase has it—'is optimal.' Maybe it would 'be best' if we cooled and darkened the carriage by putting the blinds down. But that lady over there tells us, when asked, that she is working on her tan; or that gent in the other corner tells us, when consulted, that he is a bird-watcher hoping to see a hoopoe in the trackside undergrowth. *Despite* the fact that their contributions do not help to get us to what is clearly the optimal solution, in fact directly block it, we had reason to include them in our deliberations, simply because they're passengers too with the same rights as all the other passengers. And now that we have consulted them, we have reason to listen to and try to accommodate what they've said, even if on balance we rather regret their contributions. Despite the heat and the glare, they don't want the blinds down, and they have good reasons for this preference—reasons that we may not share, but nonetheless find intelligible enough as *their* reasons. It is not 'best' that we all sit here in the heat with the blinds up. But there are cases—not *all* cases are like this, but *significantly many* are—where getting a solution that everyone accepts is more important than getting 'what's best.'

It is tempting, but it is a distortion, to insist that what this must really mean is just that we have changed our conception of the best. We can equally say that, for any one of us, his/ her conception of what's best has undergone no change whatever,

but that the point is that what we should collectively do does not depend on any individual's conception of the best, but on the result of our shared deliberation.

Shared deliberation of the kind that I've described is genuinely epistemically humble, genuinely open to the thought 'Well, I may think, after the most careful and intelligent reflection that I can manage, that *this* is obviously right; but these other people think that it's *not* obviously right, or perhaps even obviously *not* right, and they're no less careful and intelligent than I am. So I must be at least open to the possibility that I am wrong, or have missed something important.' This epistemic humility makes us dependent, in our shared deliberations, on those we share them with, on our interlocutors. It also makes us vulnerable to the danger of falling well short of what are, or would otherwise be, optimal outcomes because of its insistence on respecting all those involved. But that, in my view, is not only not a decisive objection to such deliberation. It's not an objection at all.

Suppose I am right to think that this line of argument that I've sketched out is a fair extrapolation from what Mulgan actually says. (It certainly isn't a careful or scholarly exposition of his text *ad litteram*, and I don't claim it is. Mulgan is his own best expositor, and my aim here is not to be a faithful expositor of his text but to be at least slightly interesting.) Then the key charge that EBW brings against many of our current assumptions and arrangements is not just the familiar charge that they are unjustifiable (though I think it is pretty clear that Mulgan thinks that too). Rather, the key charge in EBW is the less obvious point that we get by applying MT1 to the case of B@. It is the charge that these arrangements could not be justifiable to the group who, on certain plausible assumptions, will be the group most gravely disadvantaged by them: the inhabitants of B@. (Whoever these inhabitants may be. I am aware of the supposed problems raised

elsewhere by Parfit and indeed by Mulgan himself about 'person-affecting choices,' but to be honest I am unimpressed by these problems. All they seem to me to show is that we can't define 'harm to x' as 'making x worse off than x would otherwise be.') And this fact can be made most apparent by looking at our arrangements as Mulgan looks at them in EBW: from B@'s inhabitants' perspective. From that standpoint much of what we do in our society, as Mulgan repeatedly brings out, seems *just absurd:* motor racing, for example. And are we well-placed to argue back, dog-in-the-manger style, that our choices may not be maximising, but they are at any rate our choices, and that after all, as above, an acceptable public policy can be sub-optimal? We are not. The permission that we derived above to do *less than the optimal* in public policy is not a permission to do *whatever the hell we like*–where 'whatever the hell we like' seems an unhappily apt description of far too many of our own society's arrangements. In any case we are subject, as before, to MT2's stringent requirement to place everything we want to justify to others under the harsh spotlight of the absolute demand that *every* kind of bias and contingency be stripped away from our justifications.

This requirement is another factor, alongside the strange perspective of B@, that generates a kind of Martianism in Mulgan's approach. Martianism, the deliberate adoption of an amazed outsider's view on human arrangements, is another theme that we could fruitfully pursue through his book, and indeed through many other books and articles. Martianism in western philosophy goes back at least to Heracleitus' weirdly estranged mode of observation of his fellow men and their ways. (In western literature it goes back even further: there is one kind of Martianism in the four-legs-two-legs-three-legs riddle that the Sphinx sets Oedipus, and another kind in Sophocles' famous line "Many are wonders, but nothing more wonderful than Man" (Πολλὰ τὰ δεινὰ κοὐδὲν ἀνθρώπου δεινότερον πέλει, *Antigone*

332).) Martianism is of course a central weapon in Socrates' and Plato's philosophical panoplies, in Hobbes', Descartes', Voltaire's, Paine's, and Bentham's, and today in a whole army of writers (especially the utilitarian ones) on applied ethics; the great opponents of Martianism are Aristotle in ancient Greek philosophy, John Locke, Thomas Reid, Edmund Burke, and John Stuart Mill in the Enlightenment, Heidegger and Wittgenstein in the twentieth century, conservatives like John Cottingham, Alasdair MacIntyre, Charles Taylor, Rai Gaita and Roger Scruton in the twenty-first.

It is a striking achievement of Mulgan's book to have found out what so far as I know is a genuinely new variation on this ancient theme of Martianism. However, looking at the arrangements that he wants to criticise from the Martian perspective of B@ is not, as it is in some other authors, a mere rhetorical or heuristic device to bring home to us emotionally the conclusions for which Mulgan wants to argue anyway. On the contrary, this deployment of the second-person standpoint is an autonomous mode of argument in ethics in its own right, quite distinct from other modes of argument such as, in particular, the appeal to consequences. In fact, it seems to me that in presenting this argument Mulgan is presenting a basically non-consequentialist argument for a particular attitude to the future. And I, of course, am the very last person that he will hear complain about that.

The Open University

If you need to cite this article, please use the following format:

Chappell, Timothy, "The Future-Person Standpoint," *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014), 29-46, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

MORAL AND POLITICAL PHILOSOPHY FOR A BROKEN WORLD?

HUMAN RIGHTS AND THE BROKEN WORLD

BY

JESSE TOMALTY

 Philosophy and Public Issues (New Series), Vol. 4, No. 2 (2014): 47-57
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Human Rights and the Broken World

Jesse Tomalty

I

In *Ethics for a Broken World*,[1] Tim Mulgan invites us to partake in a series of lectures delivered in a fictional future on some of the political philosophies that dominate our current tradition. The future he asks us to imagine is one in which the world is 'broken'. In the broken world, climate change has lead to intermittent and unpredictable periods of radical scarcity in which there are insufficient resources to guarantee the survival of all existing persons (8-12). We are also to imagine, rather plausibly in light of recent scientific discoveries, that we bear causal responsibility for this situation (9). Mulgan suggests that the device of the broken world 'serves to highlight the contingency of our moral and political ideals, asking us to see our society and its ideals from the outside' (ix). In this paper, I employ the device of the broken world to reflect on one of the most prominent ideals in contemporary affluent societies, namely that of human rights. In particular, I am interested in what sense the device of the broken world shows our ideal of human rights to be contingent,

[1] T. Mulgan, *Ethics for a Broken World: Imagining Philosophy After Catastrophe* (Montreal & Kingston-Itacha: McGill-Queen's University Press, 2011). All page references in the text refer to this volume.

 Philosophy and Public Issues (New Series), Vol. 4, No. 2 (2014): 47-57
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

and what implications this might have for how we should understand and evaluate this ideal.

My focus will be on human rights as a political ideal.[2] According to this ideal, all political institutions should be structured so as to respect and promote a set of particularly important rights, namely human rights, and it is incumbent on all moral agents to ensure that political institutions fulfill this requirement. There is substantial disagreement over how the content of human rights is determined,[3] and yet there is significant convergence on at least a core set of human rights, which finds expression notably in the United Nations' Universal Declaration of Human Rights (UDHR).[4] These rights include

[2] The fulfillment of human rights is widely accepted to be a political ideal, although it is also taken by many to be a moral ideal. Some theorists reject this view in favour of the view that human rights are essentially political. See, for example, Andrea Sangiovanni, 'Justice and the Priority of Politics to Morality', *Journal of Political Philosophy*, 16 (2008), 137-164; and Charles Beitz, *The Idea of Human Rights* (Oxford: Oxford University Press, 2009). I hope to avoid this debate here by focusing on human rights as a political ideal.

[3] We can identify two main factions: On the one hand are those who think that the content of the rights of international human rights doctrine should be derived from a more abstract set of universal moral rights. See, for example, James Griffin, *On Human Rights* (Oxford: Oxford Unviersity Press, 2008); and John Tasioulas, 'The Moral Reality of Human Rights', in *Freedom from Poverty as a Human Right: Who Owes What to the Very Poor?*, ed. by Thomas Pogge (Oxford: Oxford University Press, UNESCO, 2007), pp. 75-100. On the other hand are those who think that the content of human rights is justified with reference to the point and purpose of the practices governed by human rights norms, and the role these norms play within these practices. See, for example, Sangiovanni and Beitz. These two factions fragment further in light of disagreements respectively over the grounding values of universal moral rights, and how to interpret human rights practice.

[4] The complete text of the UDHR can be accessed at http://www.un.org/en/documents/udhr/. I focus on the UDHR because it is the centrepiece of international human rights doctrine, which now constitutes

rights to life, liberty, and security of person, rights against slavery, torture, and inhumane punishment, rights to due process and equality before the law, rights to nationality and freedom of movement, rights to freedom of conscience, religion, association, and expression, the right to own property, the right to marry and found a family, rights to political participation, rights to work and to decent working conditions, and rights to basic resources, education, and healthcare. In what follows, when I refer to human rights, I mean the set of rights articulated in the UDHR and other official human rights documents unless otherwise specified.

Human rights have come to dominate the discourse on global political morality. They are used to criticize the actions and policies of states and their officials, and they are appealed to as grounds for undermining the sovereignty of states through various forms of intervention. The rhetoric of human rights is used to characterize serious injustices and constitutes a call to action on the part of those who are in a position to agitate for change. The ideal of human rights finds widespread support in our world. But what does it look like from the perspective of the broken world?

II

The claim of Mulgan's fictional lecturer that '[a]ffluent people were obsessed with rights' (18) suggests that people in the broken world are not. This highlights one sense in which the ideal of human rights is contingent, namely in that we have come to hold

an extensive body of mainly legal documents that seek, in general, to specify and in some cases extend the rights of the UDHR.

it as an ideal as a result of the social circumstances in which we find ourselves and the history that has lead to them. The fact that we hold the ideals we do, including the ideal of human rights, is surely contingent in this way. We know that people have not always recognized human rights as an ideal, and we have reason to think that we might never have come to recognize this ideal, had history taken a different course. And although we do not know what the future will bring, there is no reason to rule out the possibility that it will take a course that leads to human rights no longer being recognized as an ideal, as is the case in the broken world Mulgan envisions.

We should not find this 'historical' contingency either surprising or troubling for our ideal of human rights. The fact that we might not have come to hold this ideal or that people in the future might come to reject it does not suggest that we lack reason to hold it now. Ideals can be justified non-contingently even if we only came to recognize them as result of contingent circumstances and histories. We find a parallel in the way we think about scientific principles: We may have discovered them because of contingent circumstances (some were even discovered by accident), but this does not make their truth contingent on the circumstances that lead to their discovery. Something similar might be the case for at least some of our moral and political ideals. We might think, for example, that slavery was wrong even when it was considered socially acceptable because slavery is always and non-contingently wrong. The fact that slavery was considered socially acceptable might serve to partially excuse those who engaged in the practice, but it does not do away with the wrongness of it.

Historical contingency can be distinguished from a second sense in which an ideal might be contingent. In this second sense, an ideal is contingent if the *justification* for holding it as an ideal

depends on particular circumstances obtaining. We can refer to this kind of contingency as 'justificatory' contingency. An important insight gained from Mulgan's discussions of our contemporary ideals from the perspective of the broken world is that many of our ideals are contingent in this way, from the libertarian ideal of minimal government and the nationalist ideal of territorial rights to the more moderate liberal ideals endorsed by rule utilitarians and Rawlsians. These and other ideals are undermined by the conditions of radical scarcity characteristic of the broken world.

This looks to be true of many of the human rights whose fulfillment we take to be a political ideal. Perhaps most obviously problematic are the socio-economic rights expressed in the UDHR and repeated (and sometimes expanded) in other human rights documents. How, for example, could political institutions be structured in such a way as to ensure that everyone enjoys 'a standard of living adequate for the health and well-being of himself and of his family' (Article 25.1) when resources are insufficient to ensure that everyone has enough to survive? How could guarantees of free education (Article 26.1), free choice of employment (Article 23.1), and 'periodic holidays with pay' (Article 24) be possible in conditions of radical scarcity? And how could inhabitants of the broken world justify a right to found a family (Article 16.1) when this would only mean more mouths to feed that cannot ultimately be fed? We should note that the conditions of radical scarcity would also make a number of other rights difficult to justify and secure. For example, the legal institutions necessary for the fulfillment of due process rights (Articles 10 and 11) are extremely costly to operate. Whereas in affluent conditions, the fulfillment of all human rights is possible, in conditions of radical scarcity, we would be faced with the question of whether to forego due process rights in favour of devoting more resources to food production or medicine.

Our ideal of human rights, then, looks to be contingent on conditions of affluence. But contingency poses a special challenge for human rights, which are meant to be universal. The universality of human rights is expressed in the very title of the *Universal* Declaration of Human Rights, but we also find in it the assertion contained therein, that 'Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind' (Article 2). Reflecting on human rights from the perspective of the broken world encourages us to clarify the sense in which human rights are universal. How can the alleged universality of human rights be squared with their apparent contingency?

III

One possibility is to deny that the infeasibility of the realization of human rights for all persons in the broken world makes human rights a contingent ideal. On this view, each inhabitant of the broken world has the same set of human rights as we do, but these rights simply cannot be fulfilled, at least not all at once and for everyone. On this kind of view, human rights might be thought to provide guidance for institutions to approximate the minimal conditions for justice, despite the fact that institutions in the broken world could never fully meet these conditions.

It might be objected that there is little value in pursuing strictly unattainable ideals. But ideals are meant to orient our actions, and there is no reason to think that unattainable ideals are always ineffective in this respect. I can know that I will never be able to play the piano as well as Glenn Gould, but I can look to his performances as exemplars and use them to orient my training.

Why, then, not think that the inhabitants of the broken world could model their institutions on our ideal of human rights despite knowing that they will never succeed in attaining it?

The problem with this idea, I argue, is that it is not clear how inhabitants of a broken world should orient themselves towards our ideal of human rights. Recall that according to our ideal of human rights, political institutions must be structured in such a way as to ensure the fulfillment of human rights for all their members. In the broken world, there are insufficient resources to achieve this. A natural thought is that political institutions in the broken world should approximate our ideal of human rights by seeking to maximize the fulfillment of human rights of their members. But what would the maximization of human rights entail? Suppose that resources in the broken world are adequate to fulfill *some* human rights for *everyone*. Would this be the closest approximation of our ideal? Why not instead use the available resources to fulfill *all* human rights for *some* people, and none for others? Or why not fulfill all human rights for a very few, a very limited number of human rights for some, and none for the rest? Each of these options would involve maximizing the fulfillment of human rights in some sense, but it is not clear which one best approximates our ideal of human rights in the broken world.

Indeed there is reason to doubt that maximizing the fulfillment of human rights in any sense is the right way to avoid injustice in the broken world at all. In the affluent world, our ideal of human rights is a kind of sufficientarian principle that limits just political institutions to those under which everyone's human rights are fulfilled. Not only does this leave an enormous amount of variation in the possible institutional arrangements and the distributions that they allow, it also allows for individuals to attain standards of living far higher than the mere fulfillment of human rights. But this is not the case in the broken world. Maximizing

the fulfillment of human rights would leave no resources for the pursuit of other ends. Maximizing human rights in the broken world would mean that the mere fulfillment of human rights would be the highest standard of living anyone could hope to attain.

The fulfillment of human rights as a minimal standard for the justice of political institutions loses its appeal under these conditions of radical scarcity. Some people might reasonably prefer to have a lower chance of survival in exchange for a higher standard of living should they survive.[5] Some might, in other words, reasonably prefer a smaller chance at a standard of living higher than that of the mere fulfillment of human rights than they would have if as many people as possible were guaranteed human rights fulfillment. It seems that a reasonable standard of justice for political institutions in the broken world should accommodate this kind of preference. This would mean that the maximization of human rights could actually end up undermining justice in the broken world. Our ideal of human rights thus looks to be contingent after all.

IV

But what, then, of the supposed universality of human rights? When we say that human rights are universal, we cannot mean that they are held by all humans and that their fulfillment is a requirement of political institutions in all times and places. The universality of human rights must be bounded. The idea of bounded universality is not unfamiliar. For example, when a state

[5] This point draws on Mulgan's discussions of libertarian and Rawlsian survival lotteries, pp. 62-64 and pp. 189-192.

offers *universal* health care, this just means that health care is provided to all those who qualify as members of that state in the relevant sense (of example, as citizens or residents). The universality of human rights, however, is often thought to imply that the relevant boundaries are those of humanity: What makes human rights *human* rights is that they are held universally by all humans, and not confined only to those living in particular societies. I have shown that the infeasibility of fulfilling human rights in the broken world challenges this view, and suggests that the universality of human rights must be bounded by conditions of affluence.[6]

I argue, however, that this need not make human rights obsolete in the broken world for at least two reasons. One is that inhabitants of the broken world could articulate the injustice of their situation in terms of human rights. Assuming that their plight would be at least partly the foreseeable result of the avoidable actions of current and past generations, inhabitants of the broken world could argue that the failure of our political institutions to prevent these harmful actions constituted a violation of their human rights. This is because where feasibility is the only allowable constraint on human rights, as long as it is feasible to ensure the fulfillment of human rights of future generations, the failure to do contravenes our ideal of human rights.[7] This is true even if it turns out that inhabitants of the

[6] John Tasioulas outlines a similar version of the bounded universality of human rights in 'Human Rights, Universality, and the Values of Personhood: Retracing Griffin's Steps', *European Journal of Philosophy*, 10 (2002), pp. 79-100.

[7] For a detailed discussion of how our failure to mitigate dangerous climate change might constitute a violation of the human rights of future people, see Simon Caney, 'Climate Change, Human Rights, and Moral Thresholds', in *Climate Ethics: Essential Reading*, ed. by Stephen M. Gardiner, Simon Caney, Dale Jamieson, and Henry Shue (Oxford: Oxford University Press, 2002), pp. 163-177.

broken world cannot reasonably pursue our ideal of human rights for themselves.

A second way in which human rights could be relevant for people in the broken world is in a more abstracted form. I have so far been using 'human rights' to refer to the specific set of rights included in international human rights doctrine. We can distinguish this substantive sense of 'human rights' from a more formal sense that focuses instead on the normative role that human rights play. In this sense, we might refer to human rights as the set of rights that all political institutions must fulfill in order to be minimally just, and that ensuring they do so is in some sense everyone's business. Although I have provided reasons for thinking that the particular set of human rights that form our ideal of human rights cannot fulfill this role in the broken world, there is reason to think that some other set of rights will.

Furthermore, there is reason to think that the justification of this set of rights will be similar to the justification for the human rights of our ideal. The human rights of our ideal are justified by the idea that failing to respect them constitutes a failure to appropriately respect the moral status of persons.[8] What counts as a failure to respect a person's moral status will depend in part of the circumstances. For example, to deny someone access to the means for subsistence might be construed as a failure to respect her moral status under affluent conditions, but this could not plausibly constitute such a failure of respect in the broken world. Nevertheless, there will be other ways in which people's moral status can fail to be respected in the broken world. Denying someone access to the means for subsistence might not

[8] I take it that this is a unifying thought that is compatible with the disagreements, which I mentioned in section 1, over the specific values that ultimately ground human rights.

constitute a failure to respect a person's moral status in the broken world, but denying him access to the means for subsistence simply because of his ethnicity plausibly would count as such a failure of respect.

V

Reflecting on our ideal of human rights from the perspective of the broken world shows it to be contingent on conditions of affluence. I have argued that we can, nevertheless, make sense of the universality of human rights in at least two ways. One involves the idea that our ideal of human rights is only bounded by feasibility constraints. This, I have argued, has important implications for how we think about our obligations to future generations. Second, we find that a formal sense of 'human rights' is not burdened by the same contingency as the substantive sense. This leaves room for an ideal of human rights for the broken world that is, in some ways, continuous with ours.

There is a great deal more that can be said about human rights in the broken world, and how we should think about our obligations towards future people with respect to their human rights. My purpose in this brief paper has been to point to some of the ways in which the device of the broken world can help to orient our thinking about human rights as a political ideal both now and in the future.

Nuffield College, University of Oxford

If you need to cite this article, please use the following format:

Tomalty, Jesse, "Human Rights and the Broken World," *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014), 47-57, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

MORAL AND POLITICAL PHILOSOPHY FOR A BROKEN WORLD?

REPLIES TO CRITICS

BY

TIM MULGAN

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 58-92
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Replies to Critics

Tim Mulgan

I am grateful to Ben Saunders, Tim Chappell, and Jesse Tomalty for their careful readings of my book, their generous comments about it, and their engagement with its themes and ideas.[1] In my reply, I will address each commentary in turn, also drawing on my own more recent work, where I apply lessons from the broken world to contemporary moral philosophy.

[1] I am also very grateful to Gianfranco Pellegrino for editing this issue, and to Sarah Broadie for organising a workshop on my book at the University of St Andrews in April 2012, where earlier versions of the three commentaries were presented.

 Philosophy and Public Issues (New Series), Vol. 4, No. 2 (2014): 59-92
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

I
Ethical Lessons From the Broken World

I begin with some general comments about the moral significance of thinking about a broken future.[2] In my book, I focus on presenting ethics *within* a broken world—asking how philosophers in that possible future might respond to, and perhaps rethink, our current moral and political theories. However, the broken future also has lessons for contemporary philosophy.

If we encountered an isolated population, perhaps on some distant planet, living without favourable conditions and operating survival lotteries, that would be unsettling enough. But because the broken world may be *our future*, it also has a significant impact on our *current* ethical thinking.

The broken world is a *credible* future. No-one can reasonably be confident that it won't happen. It involves no outlandish claims, scientific impossibilities, or implausible expectations about human behaviour. Climate change—or some other disaster—might produce a broken future. This is not to say, of course, that the broken future will happen. Many other futures are also credible. Some are much better, others are much worse. Our epistemic situation does not allow us to make confident predictions either way. But the broken world is one very real possibility.[3] As all three commentators accept the credibility of

[2] This sections draws especially on T. Mulgan, "Ethics for Possible Futures," *Proceedings of the Aristotelian Society*, volume 114 Part 1 (2014), 1-17 and "Utilitarianism for a Broken World," *Utilitas*, forthcoming.

[3] 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

the broken future, I will not discus this issue further here. Instead, I shall take the credibility of this future as given, and explore its implications for moral philosophy.

A credible broken future teaches us four main ethical lessons. First, it undermines our tendency to ignore our obligations to distant future people. Philosophers have traditionally marginalised intergenerational issues, because they were confident that they could set the future aside. If we create a stable liberal democratic society in our own generation, then our descendents will inevitably be better-off than us, and therefore their interests do not conflict with ours. In John Rawls's liberal society, for instance, the only intergenerational question is the 'just savings problem': how much better-off we should leave our descendants?[4] The prospect of a broken future undermines this optimistic presentism. We no longer take it for granted that we *will* leave our descendants better-off, or even that we *can*. And many of our most urgent moral dilemmas involve intergenerational conflict. We now realize that future people might be worse-off *because* we have looked after ourselves.

A second lesson is that the broken future alters the *comparative* plausibility of competing moral and political theories, simply because some theories cope better than others with obligations to future people *in general*. One example, which will become very relevant in my reply to Saunders in section 2, is the perennial debate between contractualist and utilitarian accounts of the foundations of morality. I have argued elsewhere that utilitarianism accommodates the future more easily than

probabilities to outlier possibilities where various feedback loops cause the global climate to spiral out of control once some threshold is passed.

[4] J. Rawls, *A Theory of Justice*, rev. ed. (Cambridge, MA: The Belknap Press of Harvard University Press, 1999), 251-259.

contractualism, which is its main rival in the contemporary intergenerational literature.[5]

Utilitarianism bases all our obligations on the fact that our actions impact on the well-being of sentient beings. Obligations to future people are thus theoretically on a par with obligations to present people. While utilitarians endlessly debate the precise details of our intergenerational obligations, they have no difficulty making sense of them. By contrast, contractualists have great difficulty accommodating any obligations to future people at all.

Contractualist accounts of intergenerational obligation face two barriers. The first is Derek Parfit's non-identity problem.[6] Contractualists model morality or justice on a bargain or agreement among rational individuals. But how can we begin to imagine contracts, bargains, or cooperative schemes involving future people whose existence and identity depend upon what we decide? Contractualists as diverse as Immanuel Kant, John Rawls, David Gauthier, and T. M. Scanlon all face serious difficulties here. The second barrier is the lack of reciprocal interaction between present people and distant future people. We can do a great deal to (or for) posterity, but, as the saying goes, what has posterity ever done for us? We cannot bargain, negotiate, or cooperate with those who will live long after us. A *contract* with distant future people seems incoherent.

To crystallise the problem, imagine a 'time bomb' that devastates people in the distant future but has no direct impact

[5] T. Mulgan, *Future People* (Oxford: Oxford University Press 2006), chapter 1 and 2; "Utilitarianism for a Broken World"; "Contractualism for a Broken World," Paper presented to workshop on contractualism, Universite de Rennes (May 2012). (Available from author).
[6] D. Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), Part 4.

until then.[7] (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?

Of course, many contractualists do try to accommodate intergenerational justice. They cite the motivations of present people, exploit contracts between overlapping generations, appoint trustees or ombudsmen for the future, or construct imaginary intergenerational bargaining situations where the parties know neither when nor whether they exist.[8] But these intergenerational contracts all seem troublingly ad hoc. It is hard to escape the conclusion that, for the consistent social contract theorist, intergenerational justice is (at best) an afterthought—an optional extension of a theory of justice designed for contemporaries. Contractualists cannot accommodate the future as easily or as naturally as utilitarians do.

If conflicts between generations were rare, or if we could be confident that future people would be better off, then this comparative weakness of contractualism might not matter. (After all, no theory is perfect, and utilitarianism certainly has problems of its own.) But, if we face a broken future, then our need for a credible account of our obligations to future people is much greater. This doesn't prove that utilitarianism is superior all-things-considered, but it does significantly enhance its comparative appeal.

[7] I owe the notion of a time bomb to A. Gosseries, "What do we owe the next generation(s)?" *Loyola of Los Angeles Law Review* 35 (2001), 293-354.

[8] See, e.g., the essay collected in A. Gosseries and L. Meyer (eds.), *Intergenerational Justice* (Oxford: Oxford University Press, 2009).

This brings us to our third general lesson. Some moral theories handle a *broken* future better than others. Working through the ethical implications of the broken future, both while I was writing my book and in my subsequent work, I have been struck by the number of different ways that philosophers help themselves to optimistic assumptions about the future. Consider four disparate examples: those strands of 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[9]; the many strands of contemporary moral philosophy built on *intuitions* about simple cases—intuitions that are very closely tied to our affluent present[10]; libertarians who presuppose that initial acquirers can leave 'enough and as good' *for all future people*;[11] or Rawlsian liberals who insist that 'justice' only applies while favourable conditions persist.[12] The recognition of a credible broken future thus counts against naturalist meta-ethics, intuition-based ethics, libertarianism, and (Rawlsian) liberalism. As philosophy is a comparative business, the broken future thus supports non-naturalism, theoretical ethics, and alternative political philosophies such as utilitarianism.

The fourth impact of the broken future is that it raises troubling practical questions about how we should live *now*. Can we reasonably justify a refusal to adopt the ethical outlook of the broken world *for ourselves*? If future people will be worse-off, partly as a result of our actions, should we reduce our aspirations,

[9] T. Mulgan, *Purpose in the Universe: the Moral and Metaphysical Case for Ananthropocentric Purposivism* (Oxford: Oxford University Press 2015).
[10] T. Mulgan, "Theory and intuition in a broken world," in M. Di Paola and G. Pellegrino (ed.), *Canned Heat. Ethics and Politics of Global Climate Change* (Delhi/London: Routledge 2014).
[11] T. Mulgan, *Ethics for a Broken World Imagining Philosophy After Catastrophe* (Montreal & Kingston-Itacha: McGill-Queen's University Press, 2011), 18-68.
[12] *Ibid*., 160-196.

and bring our notion of what is necessary for a worthwhile human life into line with theirs? Can we insist for ourselves on goods and opportunities that will not, as result of that very insistence, be available to future people?

I am exploring these four lessons in my own current work, and I will draw on them in my replies to Saunders, Chappell, and Tomalty.

II

Ben Saunders on Broken World Democracy

Ben Saunders makes a number of important points about the difficulties that surround any attempt to extend present-day democracy to include future people. He also helpfully separates questions of democracy from questions of justice. Here is how Saunders summarises his own position: "I shall argue that there is nothing *undemocratic* about excluding future generations. Whether we treat them unjustly is a more difficult question, which I do not answer." (p. 12) In my reply to Saunders, I shall address issues of both intergenerational democracy and intergenerational justice.

Saunders begins with the 'All Affected Principle'—the common thought that everyone who is affected by a decision has a right to participate in making that decision. This principle is often used to defend extensions of the franchise. (To take one example that Saunder himself discusses: If people in distant lands are affected by pollution originating in our country, then perhaps they have a right to influence our deliberations about industrial and environmental policy.)

As Saunders notes, it is not possible for future people to participate directly in our deliberations. (To avoid unnecessary

complications, I shall assume that 'future people' refers to people who exist sufficiently far in the future that they will not overlap with any present person.) When we consider extending the franchise to include additional present people, the barriers are typically practical or political. We could enfranchise people in other countries, if we really wanted to. With future people, by contrast, the barriers to enfranchisement are metaphysical. Only those who exist now can participate in current decisions. Future people do not yet exist. Therefore, we cannot directly enfranchise them. (If democratic institutions endure into the future, then perhaps future people will get to vote when their time comes. But that is not the same as enfranchising them now. And, of course, if we make the wrong decision now, we may prevent future people from enjoying *any* democratic rights.)

At least some of our current decisions will inevitably impact on future people. We will affect them. But we cannot directly enfranchise them. Therefore, we cannot avoid violating the All Affected Principle. If we insist that respect for the All Affected Principle is essential for democratic legitimacy, then it follows that democratic legitimacy is an impossible dream. No government is ever legitimate.

This theoretical puzzle for democratic legitimacy is easily overlooked in practice. As I noted in section 1, political philosophers have typically assumed that future people will be better off than present people, and that there are no significant conflicts of interest between present and future people. If we focus on building stable democratic institutions for ourselves, then the future will take care of itself. Our only obligation to future people is to bequeath our democracy. This is one place where the credible threat of a broken future undermines the optimism of contemporary political philosophy. The

intergenerational adequacy of democracy now becomes a very pressing concern.

Historically, there is a strong connection between social contract theory and democracy. Both are often motivated by the following desires: to respect the separateness of persons, to enable individuals to participate in decisions that affect them, and to avoid sacrificing individual interests on the altar of aggregate well-being. However, thanks to these similarities, democracy is also vulnerable to the objections to social contract theory that I sketched in section 1. The barriers facing any intergenerational contract also threaten intergenerational democracy. As Saunders himself forcefully argues non-identity and lack of reciprocity mean that techniques that successfully extend the franchise to include new present people are unlikely to work for future people.

Like social contract theorists, democrats have tended to side-line these difficulties by making optimistic assumptions about the benefits that present democracy can offer to future people. It doesn't matter whether future people are enfranchised now, because they will inevitably be so much better-off than us. The credibility of the broken future threatens this complacent defence of democracy. This raises the possibility that democracy is another moral ideal that does not translate to a broken world. And, if it might lead to a broken world, perhaps democracy is another indefensible affluent luxury that we ourselves should abandon.

In my book, I imagine a broken future where people have abandoned democracy, where the very idea of democratic government is greeted with derision, and where affluent democracy is identified as one principal cause of the broken world. While contemporary philosophers are almost universal in their support for democracy, my broken world philosophers are

universally suspicious of it. I made this rhetorical decision to emphasise the potentially radical impact of the broken world. (My imaginary philosophers' outright rejection of democracy contrasts with their more even-handed treatment of competing theories of justice—libertarianism, egalitarianism, utilitarianism. In my own recent work, I have used the broken world to argue that utilitarianism is superior to its rivals. But the book aims to be neutral about justice.)

Saunders raises a number of very telling questions about my approach to democracy in *Ethics for a Broken World*. He rightly observes that I do not offer any detailed description of the non-democratic society that operates in the broken world. This lack of detail is a general policy throughout the book. I tried to leave the background facts of everyday life in the broken world as open as possible, so that readers would not be distracted by unnecessary and arbitrary details. However, the lack of detail also limits the power of my implicit critique of democracy. Without a credible alternative, we cannot assume that democracy is not the best available system. Even if democracy might yield a broken future, we should only reject it if some identifiable alternative offers a better prospect.

In my book, I stipulate that people living in the broken world have abandoned democracy. This may suggest that I believe democracy *cannot* survive in such a world. However, I am not so pessimistic. In fact, I am inclined to agree with Saunders's suggestion that, despite its problems, democracy may be the best response to the threat of a broken world. As he says: "Democracy may not be perfect, but it is arguably less dangerous to the future than any feasible alternative (p. 29)."

One significant impact of the broken future is that we must reconsider the comparative plausibility of competing arguments for democracy. We can divide arguments for democracy into two

broad types: intrinsic and instrumental.[13] An *intrinsic* argument claims that democracy is the only legitimate form of government, because democracy alone respects the equal rights, dignity, and autonomy of moral agents. For instance, if everyone has an equal right to participate in decisions that affect them, then people have a right to democracy—even if they would be better-off under a benevolent dictatorship. Democracy is thus the only legitimate system, whether or not it is the most efficient way to promote people's interests. By contrast, an *instrumental* argument claims that, despite the superficial appeal of benevolent dictatorship, democracy is the most efficient way to promote some independently valuable end.

The divide between intrinsic and instrumental arguments often tracks the divide between contractualist and utilitarian theories of justice. Arguments for democracy within the social contract tradition typically focus on democracy's intrinsic merits. Democracy is legitimate because it is what rational agents will choose in a fair and equal bargaining position. On the other hand, when utilitarians defend democracy they invariably offer instrumental arguments. For instance, John Stuart Mill famously argued that, under the right circumstances, representative democracy is the system of government that best promotes human well-being.

I argued in section 1 that, whatever their respective merits in relation to justice between contemporaries, utilitarianism is superior to contractualism in the intergenerational case. For similar reasons, while intrinsic and instrumental arguments for democracy both have their defenders, only instrumental considerations are relevant for future people. Future people cannot themselves participate in our present democracy. Therefore, we cannot plausibly claim that our democracy respects

[13] *Ibid.*, chapter 16.

their right to participate. We can only offer the instrumental argument that democracy best protects the interests of future people. (This might include their interest in having their own democratic institutions.) If democracy is *not* instrumentally justified—if some alternative would do a better job for future people—then perhaps we should endorse that alternative instead.

As Saunders notes, even if democracy is "the only justifiable form of government in our circumstances," we must will weight it against other values (p. 18). Saunders provides a nice example of the potential for conflict between democracy and other ideals. "Perhaps there is a sense in which it would be more democratic if decisions over my private life, such as what religion I should practice, were taken out of my hands and decided by a vote with universal franchise. But, if this is so, then giving me rights over my own private life is a justified departure from democracy" (p. 18). Here, the clash is between democracy and (other) individual rights. In a similar way, we might reasonably prefer a less democratic option today, if that would make life better for future people. (One intriguing possibility is that a non-democratic government today might be the best way to safeguard the rights of future people. I return to the potential for intergenerational conflicts of rights at the end of my reply to Tomalty in section 4.)

If we take seriously the possibility of a broken future, then we must focus, not on the intrinsic merits of democracy, but on its effects. In particular, we must ask whether democracy best serves the interests of future people. This is not a question that has received sufficient attention from political philosophers. Saunders addresses it perceptively in his commentary. He argues that the broken world does not provide sufficient reason to reject democracy. I am inclined to agree.

I begin with a debate *within* democratic theory that Saunders also discusses. This is the familiar debate between majoritarianism

and constitutionalism. Majoritarians favour an unrestricted democratic legislature with the power to remake all laws; while constitutionalists favour an entrenched bill of rights that places some decisions beyond the scope of the current majority. (These two theories roughly correspond to the constitutional systems of the UK and the USA respectively.) Some commentators have argued that, while majoritarianism is more democratic, an entrenched constitution will better protect the rights of future people. As Saunders notes, I have myself defended majoritarianism against this objection. In my earlier book, *Future People*, I argued on utilitarian grounds that the best protection for future people lies in the *motivations* of present people.[14] If present people care for future people, then we will do what we can to promote their interests and safeguard their rights. And there is no good reason to believe that any minority—rulers, judges, experts, or whoever—will reliably perform better as guardians of future people than the majority. Saunders puts it well: "Either power is held by the masses or it is held by a few. Neither group is accountable to the future, so neither can be assured to treat future generations justly" (p. 27). The best hope for future people lies in the possibility that *all* present people will come to see themselves as guardians of the future.

The threat of a broken future raises the stakes in the debate between majoritarians and constitutionalists, and focuses our attention on instrumental arguments based on the impact of present decisions on future people. But I agree with Saunders that the broken world, in itself, does nothing to support constitutionalism over majoritarianism. There is no evidence that judges are more likely than legislators to defend the rights of future people, or to take a long-term view. As my broken world

[14] T. Mulgan, *Future People*, chapter 8. This chapter also provides references to the wider debate between majoritatianism and constitutionalism.

thought experiment makes plain, there is no guarantee that democracy will avoid a broken future. But no alternative political system offers any guarantee either. In chapter 14 of *Ethics for a Broken World*, I ask the reader to imagine how a Rawlsian Supreme Court judge might act in response to the threat of a broken future. I suggest that such a judge, in such a situation, might take very radical (and very undemocratic) action. But the rhetorical point of that imaginary tale is to highlight the fact that, in the real world, we cannot rely on actual Supreme Courts to save us from the threat posed by climate change.

Saunders also explores various proposals to indirectly enfranchise future people. (These include sub-majority rules, the appointment of special representatives of future people, and so on.) I briefly discussed such measures in chapter 17 of *Ethics for a Broken World*. As with constitutionalism, I do not believe that these specific departures from majoritarianism offer any guarantee that we can avoid a broken future. And the same is true of even less democratic alternatives, such as benevolent dictatorship.

Current liberal democracy presupposes affluence and Rawlsian favourable conditions. Our democratic institutions cannot survive into a broken world. But it does not follow that democracy *per se* cannot thrive there. Broken world philosophers will need to re-imagine democracy, just as they must re-imagine many other central themes of affluent philosophy. In my précis of *Ethics for a Broken World* in this volume, I said that the central focus of broken world philosophy is the development of the just survival lottery. Perhaps the just lottery is also a democratic one. I will return to this possibility at the end of my reply, as it brings together themes raised by all three commentators.

III

Timothy Chappell on The Future-Person Standpoint

Timothy Chappell offers a very thought-provoking reading of my project in *Ethics for a Broken World*. His discussion of the 'future-person standpoint' is certainly a 'fair extrapolation' of the ideas presented in my book. The idea that we need to justify ourselves to future people is present in the book, but I do not think it is as explicit or systematic as Chappell suggests. Earlier versions of Chappell's commentary have inspired my own recent work on possible futures, and his extrapolation is most welcome.

Chappell opens with two forms of moral argument that he suggests are central to *Ethics for a Broken World*. The first is as follows:

1. In the desperate circumstances of a broken world, X would be right.

2. But X looks *obviously* wrong to us.

3. Therefore: "maybe we should consider the possibility that our objections to X are less absolute than we taken them to be" (p. 31).

Chappell's second form of argument goes like this:

1. In the affluent circumstances of our actual world, X looks *obviously* right to us.

2. But X would be wrong in the desperate circumstances of the broken world.

3. Therefore: "maybe we should consider the possibility that our acceptance of X in the actual world is less defensible than we take it to be" (p. 31).

I agree that both forms of argument do operate (below the surface) in *Ethics for a Broken World*. The qualifications that Chappell includes in both his conclusions are important. The fact that people in the broken world may have ethical intuitions or moral practices that differ from our own does not *prove* that *all* our moral beliefs are false or unreliable. I am not suggesting that readers of my book should simply abandon their affluent ethics in its entirety! However, thinking about a credible future where things that strike us as abhorrent are everyday facts of life, and where central elements of our own lifestyle are considered abhorrent, should give us pause for thought. Consider two examples. I argue in my book that future people will accept survival lotteries as the paradigmatic example of justice, and that they will regard our gratuitous consumption of fossil fuels as morally reprehensible. Suppose we accept these speculations about the future. Can we still be confident that we are justified in insisting on guaranteed rights for ourselves, or in defending our affluent lifestyle?

Chappell puts the point well: "Our perspective is, historically speaking, a most unusual perspective. What reason is there to expect it not to be also a *warping* perspective?" (p. 31) We naturally think that favourable conditions are the stable end-point of human history, and therefore that ethics should be adapted to such conditions. We think of our affluent moral sensibility as a sign of moral progress. But what if favourable conditions are merely a transient blip, and future philosophers will come to regard affluent ethics as a temporary aberration?

Chappell then outlines two 'master-theses' about moral argument that 'may also structure Mulgan's project' in *Ethics for a Broken World*.

> MT1: For an arrangement to be justified, it must be justifiable to the person who does worst out of it (p. 32).

> MT2: For an arrangement to be justified, it must be absolutely justifiable—justifiable irrespective of any kind of contingency (such as the contingent circumstances of our own affluent society (p. 32).

These are interesting theses, and I have some sympathy for both of them. I also agree that several discussions in my book do suggest these principles. However, I am not convinced that Chappell's two master-theses, as he interprets them, reflect either the project I undertake in my book, or the way that I currently think about the broken world. I will briefly elaborate our disagreement at the end of this section.

Chappell links my project in *Ethics for a Broken World* to two other ethical traditions. The first is what he calls *Martianism*: the "deliberate adoption of an amazed outsider's view on human arrangements" (p. 45). Chappell rightly notes that, in this regard, my book stands in a long philosophical tradition dating back to the Ancient Greeks. Chappell kindly describes my project as "a genuinely new variation on this ancient theme of Martianism" (p. 46). Unfortunately, while it may be unusual within academic philosophy, I do not think my use of imaginary futures is as original as Chappell suggests. After all, similar rhetorical devices are very familiar in science fiction. However, I certainly agree that one aim of my book is to use the imaginary reactions of an amazed outsider to draw attention to questionable and arbitrary features of our ethical practices.

As I suggested in section 1 of this reply, my use of the broken future goes beyond conventional Martianism, because its 'amazed outsider' is neither imaginary nor detached. My imaginary future philosophers are not Martians. They are not visitors from a distant world. Nor can we view their social arrangements with the detachment we might reserve for a broken world unconnected to ourselves. My broken world *is* our future. (Or, at least, it is one

credible future.) This brings us to Chappell's second ethical tradition.

Chappell draws a parallel between my use of the broken world and Stephen Darwall's recent emphasis on the importance of a *second-personal standpoint* in ethics. Our moral decisions often need to be justified, not merely *against* some impartial standard, but *to* the particular individuals who are affected. If I sacrifice your life for the common good, then I must offer a moral justification *that is addressed to you*.

I am sympathetic to Darwall's emphasis on the second-personal dimension of ethics. However, I also worry that this emphasis can lead us to under-estimate the comparative significance of our obligations to future people. It is much easier to offer second-personal justifications to present people than to future people. Our duties to contemporaries naturally engage our moral sentiments, because we must justify ourselves to the actual people whose real-life interests are affected by our actions. By contrast, distant future people are very remote from our everyday concerns. How can I justify myself to some particular individual living five hundred years in the future?

This worry connects with my earlier discussions of contractualism and democracy in sections 1 and 2. Especially in their debates with utilitarians, contractualists and democrats often cite the importance of justification to particular individuals. But this further highlights the barriers to extending contractualism and democracy to future people. Non-identity and the absence of reciprocity seem to make it impossible to offer the necessary second-personal justifications.

I now believe that one important function of my broken world project is to provide these future-directed second-personal justifications. Asking how our actions might impact on actual

future people can help redress the imbalance between present and future people, by giving our obligations to future people the same felt urgency as our obligations to one another. *Imagining* a credible broken future thus raises the *motivational* significance of our intergenerational obligations, as well as their theoretical significance.

Second-personal justification unites the concerns of all three commentators. Saunders argues that the need to justify ourselves to future people goes beyond the question of extending the franchise (p. 15); and the affluent ideal of human rights that Tomalty explores often emphasises a right to have one's complaints heard *and to receive a satisfactory justification*.

I am grateful to Tim Chappell for helping me to frame my own project in this way. One limitation of my book, of course, is that it imagines only one particular credible future. In one sense, this limitation is inevitable. If we are to construct second-personal justifications to future people, then we need to focus on particular futures that contain particular individuals. (We want to ask whether we could justify ourselves *to this particular person*, not merely to some abstract generality of possible future persons.) However, our ethical reasoning is incomplete if we consider only one possibility. In my recent work, I seek to address this limitation by imagining a range of other possible futures, including a virtual future where human beings spend their entire lives in Nozick's experience machine; a digital future where humans have been replaced by unconscious digital beings; and a theological future where the existence of God has been proved.[15] I argue that, like the broken future, these other credible futures also affect our current ethical thinking in surprising ways. While some of the implications of alternative futures are distinctive,

[15] For an overview of my current work here, see T. Mulgan, "Ethics for Possible Futures."

they often point in the same direction as the broken future. In particular, other futures also raise the importance of intergenerational ethics and alter the balance between competing moral theories.[16]

There is thus much in Chappell's commentary that I agree with, and much else that I find fruitful. I will close with some points of disagreement. These relate, perhaps unsurprisingly, to our different commitments in normative ethics.

In *Ethics for a Broken World*, I do not argue either for or against any particular moral theory. I believe this neutrality is appropriate in a textbook written for philosophical beginners. However, in my own more recent work, I do use the broken world to argue for utilitarianism and against its rivals, notably contractualism and libertarianism.[17] I also argue for particular theses within utilitarianism, such as the objective list theory of well-being, and rule utilitarianism.

Chappell is not a utilitarian, and he wants to use his two master-theses to defend non-utilitarian conclusions. By contrast, insofar as I am willing to endorse those two master-theses, I read them in a manner that is consistent with utilitarianism. Indeed, I would argue that utilitarianism can help us to unify Chappell's two theses.

Chappell's two theses deploy two distinct notions of justification. In MT1, we justify ourselves *to an individual*; while MT2 seeks an absolute non-contingent justification. MT1 thus suggest a focus on particular individual perspectives, while MT2

[16] A third common feature is that all my possible futures, including the broken world, push morality in a more objective direction. (T. Mulgan, "Ethics for Possible Futures," section 2.)

[17] T. Mulgan, "Utilitarianism for a broken world"; "Contractualism for a broken world"; "Ethics for Possible Futures."

suggests an impartial standard that is independent of individual differences. Chappell himself uses the concept of virtue to unite the two theses. Virtues respond to individuals, but virtues are also justified absolutely because "the traditional virtues do make sense even in [broken world] scenarios" (p. 33). I will now argue that the utilitarian tradition provides an alternative unification.

Utilitarianism need not be antagonistic to second-personal justification. The caricature of the utilitarian is a calculating machine who treats individuals merely as anonymous utility-containers. While there is some justice in this caricature, it does not represent all utilitarians. In particular, the liberal utilitarian tradition of John Stuart Mill emphasises the importance of individuality, freedom, autonomy, and individual moral judgement.[18]

The utilitarian tradition encompasses a bewildering variety of themes and theories. To get a manageable discussion in *Ethics for a Broken World*, I artificially divided utilitarians into two main groups: act utilitarians and rule utilitarians. (While these are familiar names, I use them more broadly, to contrast two composite positions.) This artificial division captures two competing strands to the utilitarian tradition—radical iconoclasm and liberal moderation.

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. *Rule utilitarians* favour a moderate morality. They picture morality as a collective enterprise, and evaluate moral codes and political institutions by their collective

[18] I explore Mill's utilitarianism further in Mulgan, "Mill and the broken world," *Revue International de Philosophie* (2014).

impact on human well-being. Rule utilitarians endorse many non-utilitarian prohibitions, permissions, rights, and freedoms.

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!)

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.

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.[19]

[19] B. Hooker, *Ideal Code, Real World* (Oxford: Oxford University Press, 2000). I develop my rule utilitarianism in T. Mulgan, *The Demands of Consequentialism*

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.

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. Drawing on arguments made famous by Mill, rule utilitarians champion their ability to accommodate a wide range of common-sense rights and freedoms, and to favour democratic government over despotism, liberal society over its rivals, free markets over command-and-control economics, and so on. One pressing question is whether this liberal moderation can survive into a broken world. I return to that question in my discussion of survival lotteries in section 4.

Act utilitarianism offers a single mode of justification. The right action is whatever, in the circumstances, happens to maximise aggregate well-being. Suppose one individual's interests are sacrificed for the greater good. If this sacrifice maximises aggregate well-being, then it is justified. This is the only justification that is available, and the *same* justification will be offered both to the injured person herself and to any interested bystander. There is no room in act utilitarianism for distinctively second-personal justification. The fact that *my* interests are at

(Oxford: Oxford University Press, 2001), especially chapter 3; *Future People*, chapters 5 and 6; *Ethics for a Broken World*, chapter 7; "Utilitarianism for a Broken World'; "Mill and the Broken World."

stake has no moral significance, and so there is no need to justify anything *to me*.

Those, like Chappell and me, who are sympathetic to Darwall's second-personal emphasis thus have good reason to reject *act* utilitarianism. But it does not follow that we must abandon utilitarianism altogether. Rule utilitarianism has resources that its simpler sister-theory lacks. Given the importance of the second-personal standpoint *to human beings*, it is quite possible that the ideal moral code will recognise the importance of justifications that are addressed to particular individuals. Within rule utilitarianism, these justifications are ultimately grounded in the promotion of aggregate well-being. But this does not undermine their second-personal sincerity. (In the same way, the rule utilitarian's commitment to keep her promises is not rendered insincere by the fact that its ultimate justification rests on aggregate well-being.)

Of course, there are limits to the flexibility of rule utilitarianism. If circumstances change *too much*, then even the most familiar moral rules may need to give way. In a broken world, many of our most cherished affluent privileges, permissions, and prohibitions may make no sense. (I explore several examples in chapter 7 of *Ethics for a Broken World*, and in my reply to Tomalty below.) This marks a theoretical divide between Chappell and me. Chappell interprets MT2 as the claim that particular virtues can be given some absolute justification that is unrelated to contingency and circumstance. For the rule utilitarian, by contrast, what is absolute is only the *standard* of justification. The measure of aggregate well-being is unvarying, but the particular moral code that it justifies may vary with the circumstances.

On the other hand, my theoretical disagreement with Chappell may not generate any great practical disagreement. Rule

utilitarianism is open to the possibility that there are some very general virtues that carry over into the broken world, and into any other credible future. Indeed, there is good utilitarian reason to expect this. For the rule utilitarian, each virtue is a useful collective response to some central aspect of human well-being. If this is so, then we might well expect to find that, in any actual future, the ideal code includes analogues of generosity, benevolence, truthfulness, kindness, courage, prudence, and the like.

Rule utilitarianism offers another link between Chappell's two master-theses. Following Mill, modern rule utilitarians often defend liberal democratic institutions. Within those institutions, decisions are made by following rules that respect individual rights and liberties, not by the calculation of aggregate utility. Real-life second-personal justifications often cite these intermediate institutions. If someone is affected by my actions, then the following is often a perfectly good justification to offer her: 'I have treated you in accordance with the just rules that govern our established institutions'. In a broken world, rule utilitarians will continue to offer such institution-based justifications, even though the precise details of their utility-maximising institutions may differ.

IV

Jesse Tomalty on human rights

Jesse Tomalty offers a very thoughtful discussion of the place of human rights in a broken world. She also asks how the prospect of a broken future might impact on our thinking about rights in our (comparatively) affluent present. I have been

thinking about rights in the broken world a great deal since I wrote my book, and I have found Tomalty's comments extremely helpful.

As Tomalty notes, *Ethics for a Broken World* presents rights as an *affluent* obsession. This suggests that people in the broken world have outgrown this obsession, and they are no longer interested in rights. As with democracy, this is a deliberate rhetorical decision, designed to generate what Chappell might call a Martian outsider's perspective on our own affluent rights-talk. As I noted in my précis, affluent conceptions of rights do not translate easily to the broken world. However, I personally believe that broken world dwellers may still find a use for the general concept of rights.

In my book, I do not address *human* rights per se. I focus instead on general theories of rights (libertarian, utilitarian, Rawlsian), and on rights to self-ownership, property, liberty, or democratic participation. This omission was partly owing to constraints of space, and partly because, unlike Tomalty, I have no particular expertise in human rights theory. However, this omission does not mean that human rights have no place in the broken world. On the contrary, I agree with Tomalty that, if they find any use for the concept of rights, people in the broken world will be interested in human rights.

Tomalty notes that some approaches to human rights that work for us would make no sense in a broken world. Without favourable conditions, it is not possible to respect all human rights (especially if we retain our modern conception of what those rights entail). Nor will it make sense for future people to maximise human rights, as there would be no room for any other human activity. Every resource would be devoted to mere survival.

Tomalty offers an interesting discussion of the different categories of rights. Scarcity has an obvious impact on 'socio-economic' rights, such as rights to food, shelter, education, health care, and so on. In a broken world, where resources are not sufficient to meet all basic needs, these rights cannot possibly be guaranteed to everyone. This difficulty is especially significant for utilitarians. As I noted in my précis of *Ethics for a Broken World*, utilitarians insist on socio-economic rights. Without these positive rights, utilitarians cannot reasonably give rights the priority traditionally associated with them.

This might suggest that a broken world is only problematic for defenders of socio-economic rights. However, as Tomalty insightfully points out, other rights may also be under threat in a broken world. In particular, many civil and political rights rely on legal institutions. The practical protection of these rights thus depends on the continued existence of those institutions. The maintenance of just legal institutions is very costly. If resources are scarce, then legal institutions may be stretched to breaking-point, and even the most negative libertarian rights may be under threat. (In Part One of *Ethics for a Broken World*, I discuss a number of other ways that libertarian rights might collapse under the prospect of a broken future.)

There is nothing in Tomalty's discussion of human rights that I disagree with. And I have learnt much from studying her account. In the remainder of this reply, I want to explore in more detail the possible roles that human rights might play in a broken world, drawing on several of Tomalty's own observations.

Tomalty notes that some human rights translate relatively easily. Rights to non-discrimination, equal treatment, and due process are as compelling in a broken world as they are in our affluent present. There is no guarantee that these human rights *will* survive into a broken world. One very real threat is that, as

the world breaks, illiberal anti-human-rights regimes may emerge. But the *justification* for anti-discrimination rights remains intact, as it does not depend on favourable conditions.

Tomalty highlights one possible future role for human rights: "inhabitants of the broken world could articulate the injustice of their situation in terms of human rights" (p. 55). Future people may express their own moral outrage at the conduct of their affluent ancestors in terms of the violation of their own human rights. Drawing on influential work by Simon Caney, Tomalty points out that, on most accounts of human rights, our failure to avoid a broken world will be a major human rights violation unless we can prove that we could not avoid this result without violating some present human rights.

I agree that future people may find the current philosophical debate over climate change and human rights helpful. I would also suggest that they will re-interpret some key terms in that debate. For instance, one popular theme is the distinction between *luxury* emissions and *subsistence* emissions.[20] While all present people share a common responsibility to reduce their CO2 emissions to protect the human rights of future people, each present individual also has a basic human right to engage in essential subsistence emissions. (There is a clear difference between the British hobby farmer who raises cattle to take advantage of lucrative EU subsidies and the sub-Saharan subsistence farmer who must raise cattle to survive.) The obligation to fight climate change by reducing emissions does not extend to subsistence emissions. Our common responsibility is thus highly differentiated in practice, because anyone whose emissions are entirely subsistence-based has no responsibility to do anything.

[20] See, e.g., H. Shue, "Subsistence emissions and luxury emissions," *Law and Policy* Vol. 15 (1993), 39-60.

In a broken world, where universal survival is impossible, there can be no guaranteed human right to subsistence. Broken world philosophers have two choices. They could regard this human right as something that should have guided affluent decisions, but no longer applies in their broken world. They would then abandon human-rights-talk altogether. Alternatively, as with other aspects of affluent rights-talk, broken world philosophers might try to adapt this human right to their own situation.

To illustrate the ways that rights might evolve in a broken world, let us imagine how a future broken world rule utilitarian, inspired by J. S. Mill, might design a just survival lottery.[21]

Any survival lottery is a very striking departure from contemporary liberal democratic ideals, as it must violate the basic rights and freedoms that affluent liberals claim *for all*.[22] Many liberals will refuse to countenance any survival lottery, and thus deny that justice has any place in a broken world. At the other extreme, act utilitarians will embrace the survival lottery as the obvious and morally unproblematic solution in a dire situation. The Mill-inspired utilitarian liberal is both (as a utilitarian) willing to think the unthinkable, but also (as a liberal) anxious to preserve our rights and freedoms as far as possible.

Act utilitarians will favour a simple hedonist survival lottery—an impersonal aggregative procedure that maximises total pleasure across generations. But my Mill-inspired rule utilitarian, like Mill himself, is not a simple hedonist. Well-being is not simply a matter of counting pleasures. A good human life also

[21] This section draws on T. Mulgan, "Mill and the Broken World"; and my various discussion of survival lotteries throughout *Ethics for a Broken World*.

[22] This is why Rawlsian liberals have such difficulty with the allocation of health care, where the scarcity of resources relative to needs mirrors a broken world.

requires the pursuit of freely-chosen objectively worthwhile goals. For Mill, individuality, self-realization, and freedom are central to well-being. While freedom without survival makes no sense, survival without freedom has no value. Freedom is one of the things that makes survival valuable. Our rule utilitarian might thus rationally accept a lower chance of individual survival in exchange for greater freedom for those who do survive.

Once we admit values other than pleasure or survival, we are confronted by a potentially infinite array of possible survival lotteries—offering different chances of survival, and different opportunities and goods for survivors. The traditional hedonist ranks these lotteries using aggregate pleasure. Mill offers no simple impersonal metric. He would rely instead on the verdicts of his famous *competent judges*.[23] If everyone would prefer a lottery with more freedom and less survival, then this is very good evidence that such a lottery best promotes human well-being. Indeed, it is the only possible evidence.

How will Mill's competent judges, dwelling in their broken world, design a just survival lottery? We know they must depart from our current notion of rights. Scarcity of resources requires a shift from guaranteeing everyone's survival to managing a fair distribution of chances to survive. Instead of vainly trying to guarantee a worthwhile life for all, a just lottery will guarantee everyone a fair and equal *chance* of living a worthwhile life. This is one place where human rights associated with procedural fairness and non-discrimination come to the fore. Indeed, in a broken world, these rights become even more important. Consider the contrast between a survival lottery that selects individuals on the basis of membership in a privileged group, and one that selects its

[23] J. S. Mill, *Utilitarianism* , in *Essays on Ethics, Religion, and Society*, J.M. Robson (ed.) (Toronto/London: University of Toronto Press/Routledge & Kegan Paul, 1969), chapter 2.

survivors at random. Even if the former lottery seems more 'efficient', it clearly violates our sense of procedural fairness. Under extreme scarcity, only a random procedure can respect people's procedural human rights. The survival lottery must literally be *a lottery*. Tomalty puts the point well: "Denying someone access to the means for subsistence might not constitute a failure to respect a person's moral status in the broken world, but denying him access to the means for subsistence simply because of his ethnicity plausibly would could as such a failure of respect" (p. 57).

Act utilitarians cannot accommodate these procedural human rights. But, for reasons similar to those outlined in my reply to Chappell in section 3, I believe that rule utilitarians can respect them. Given the intimate connections between self-respect and well-being, the moral code that maximises human well-being will be one that respects the moral status of persons. And, as Tomalty herself argues, this respect lies at the heart of our ideal of human rights (p. 56).

Another key difference between act and rule utilitarians is their attitude to disagreement. For the act utilitarian, moral disagreement is a sign of error, prejudice, or caprice. The moral truth is identified with the impersonal unchanging utilitarian calculus. For the Mill-inspired rule utilitarian, by contrast, moral disagreement is a positive and dynamic force. Given the complexities of human well-being, and the many different ways that individuals can reasonably respond to different values, it is natural to expect rational individuals to disagree.

Mill's attitude to disagreement is linked to his faith in moral progress. It may seem odd to speak of progress in relation to the broken world. But our concern here is with improvements in moral knowledge. Those who dwell in the broken world are best-placed to adjudicate between competing survival lotteries. Their

verdicts are more reliable than ours. Each new generation in a changing world is better-placed than its parents to make moral judgements pertaining to its own situation.

These rule utilitarian commitments to reasonable disagreement and moral progress have several implications. One implication is that, even within a given generation, different individuals may reasonably reach different verdicts—depending on their personal evaluation of competing values and risks. Perhaps no possible *uniform* lottery will suit all competent judges. In *Ethics for a Broken World*, I explore the possibility that broken world philosophers may develop *flexible* lotteries, where each participant selects her own mix of risk and reward.[24] (Tomalty picks up on this suggestion in relation to the range of reasonable future interpretations of human rights.)

We cannot separate the content of the rights allocated within the lottery from the procedural right to participate in choosing the lottery. This is why no central bureaucracy or code of rules could hope to either design the best lotteries, or choose between them. The best way to 'design' a lottery for future people is to enable them to design it for themselves. Mill offers inspiration here. His own ideal of the utilitarian reformer is not a single-minded bureaucrat, nor a slavish follower of some simple utilitarian principle or code, but rather a creative individual engaged in a moral 'experiment in living' who invites her fellow citizens to follow her example. In the broken world, moral entrepreneurs might imagine new ways of life, new moral ideals, and new survival lotteries.

Rule utilitarian attitudes to disagreement also support democracy and majoritarianism. Once several candidate lotteries are on the table, the choice between them should be as

[24] T. Mulgan, *Ethics for a Broken World*, 185-196.

democratic as possible. There are three reasons for this: (1) Each individual has a right to participate in decisions that affect her; (2) Each individual is best-placed to know what will best promote her own interests; and (3) At any point in time, present people are better-placed than their predecessors to decide how competing interpretations of human rights should be balanced against one another. If we seek to entrench *our* interpretation of human rights, we will be imposing our partial opinions on future people. This is neither fair nor efficient.

At the end of section 1 of this reply, I suggested that a credible broken future raises troubling practical questions about how we should live *now*. Can we reasonably justify a refusal to adopt the ethical outlook of the broken world *for ourselves*? The removal of favourable conditions raises an even more disturbing question. Suppose we conclude that, while we can guarantee our own basic needs, our descendants will need to run a survival lottery. Can we still insist on guaranteed survival for ourselves, or should we move in their direction—operating a survival lottery across the generations? If future people must rethink *their* human rights, and replace guarantees with lotteries, then perhaps we should rethink *our* human rights too. (Otherwise, we are simply asserting that our rights are more important than theirs.) Perhaps, faced with a broken future, even rule utilitarians must embrace a more demanding and austere morality.

The survival lottery strikes us as morally unthinkable. But if we leave future people in a place where *they* must think the unthinkable, then perhaps we should think it too. Perhaps the design of a just survival lottery should be *our* central philosophical concern as well.

If the future is too broken—if no fair lottery offers everyone a decent chance of survival—then liberal institutions cannot survive. But in a less-broken future, lessons from the utilitarian

tradition may keep alive the hope of safeguarding fairness, security, and individuality. Human rights and democracy, suitably re-imagined, may have a central role to play in the resulting utilitarian lottery.

University of St Andrews

&

The University of Auckland

If you need to cite this article, please use the following format:

Mulgan, Tim, "Replies to Critics," *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014), 59-92, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

MORAL AND POLITICAL PHILOSOPHY FOR A BROKEN WORLD?

JUSTICE AS FAIRNESS IN A BROKEN WORLD

BY

MARCUS ARVAN

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 95-124
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Justice as Fariness in a Broken World

Marcus Arvan

Abstract. In *Ethics for a Broken World: Imagining Philosophy after Catastrophe*, Tim Mulgan applies a number of influential moral and political theories to a "broken world": a world of environmental catastrophe in which resources are insufficient to meet everyone's basic needs. This paper shows that John Rawls' conception of justice as fairness has very different implications for a broken world than Mulgan suggests it does. §2 briefly summarizes Rawls' conception of justice, including how Rawls uses a hypothetical model—the "original position"—to argue for principles of justice. §3 explains how Mulgan uses a variation of Rawls' original position—a *broken original position*—to argue that justice as fairness requires a "fair survival lottery" in a broken world. §4 shows that the parties to a broken original position have reasons not to agree to such a survival lottery. §5 then shows that Mulgan's argument hangs upon a false assumption: that there are no viable options to adopt in a broken world besides some kind of survival lottery. Finally, §6 shows that the parties to a broken original position would instead rationally agree to a scheme of equal rights and opportunities to earn or forfeit shares of scarce resources on the basis of each person's comparative contribution to human survival.

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 95-124
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

I

In his recent book, *Ethics for a Broken World: Imagining Philosophy after Catastrophe*, Tim Mulgan applies a number of influential moral and political theories to a "broken world: a place where resources are insufficient to meet everyone's basic needs, where a chaotic climate makes life precarious and where each generation is worse off than the last."[1] Mulgan's inquiry is timely and important. Scientific predictions about climate change and its effects strongly suggest that our world may become "broken" in the foreseeable future. It is therefore important to investigate what morality would require of individuals and social-political structures in such a world. One of Mulgan's most striking theses is that a variety of different moral and political theories—Robert Nozick's libertarianism, classical utilitarianism, Thomas Hobbes' contractarianism, John Locke's natural rights theory, and John Rawls' theory of "justice as fairness"—all support the implementation of some kind of "survival lottery" in a broken world, "a bureaucratic procedure to determine who lives and who dies."[2] Although Mulgan argues that different moral and political theories would require somewhat different survival lotteries, in each case the essentials are similar: people would be issued "lottery tickets" that give each person a chance to obtain enough scarce resources to survive. Those whose tickets are selected survive, and those whose tickets are not selected die.

This paper shows, against Mulgan, that John Rawls' famous conception of justice—justice as fairness—does not permit a

[1] Tim Mulgan, *Ethics for a Broken World: Imagining Philosophy After Catastrophe* (Montreal & Kingston-Itacha: McGill-Queen's University Press, 2011).
[2] Ibid: 10-11. For Mulgan's discussion of how libertarianism supports a lottery, see pp. 62-66. For his discussion of utilitarianism and survival lotteries, see pp. 142-6. For his discussion of Hobbes and lotteries, see pp. 157-8; for Locke and lotteries, see p. 159; and for Rawls and lotteries see lecture 15.

survival lottery in a broken world. §2 briefly summarizes Rawls' conception of justice as fairness, including how Rawls uses a hypothetical model—the "original position"—to justify principles of justice. §2 then explains how Mulgan uses a variation of Rawls' original position—a *broken original position*—to argue that justice as fairness requires a "fair survival lottery" in a broken world. §4 shows that the parties to a broken original position have reasons not to agree to such a survival lottery. §5 then shows that Mulgan's argument hangs upon a false assumption: that there are no viable options to adopt in a broken world besides some kind of survival lottery. I show, to the contrary, that the following scheme is a viable alternative: affording each person in a broken world equal rights and opportunities to earn or forfeit shares of scarce resources on the basis of their contribution to the survival of others. Finally, §6 shows precisely why the parties to a broken original position would rationally agree to this alternative over a survival lottery. Because free and equal individuals in a "broken original position" would know that they each have one, and only one, life to live, they should all rationally aim to avoid leaving their fate to mere chance, as a survival lottery requires. They should instead all rationally prefer a competitive scheme in which each person has rights and opportunities to earn or forfeit shares of scarce resources, thereby ensuring, at least as far as is possible in a broken world, that whether they live or die is determined by their choices, talents, and hard work, not mere chance.

II

Justice as Fairness: A Brief Overview

Although Rawls' theory of justice as fairness is complex in its details, its root ideas are simple. Rawls contends that a just society

would conform to principles that free and equal individuals would rationally agree to from an "original position" of fairness: a hypothetical position in which no one is able to arbitrarily privilege themselves or anyone else on any contingent grounds, such as their own identity, race, gender, religion, natural talents, social class, etc.[3] Rawls argued that because no one in the original position knows anything about their own identity, it is rational for everyone in the original position to seek "social primary goods"—goods that will enable them to effectively pursue their goals no matter who they turn out to be: basic rights and liberties, political and economic opportunities, income and wealth, and social props to self-respect.[4] Finally, Rawls argued—on grounds that need not concern us at present—that the parties to the original position should agree to the following principles of justice for distributing these goods *in a fully just society*:

> a. Each person has an equal claim to a fully adequate scheme of equal basic rights and liberties, which scheme is compatible with the same scheme for all; and in this scheme the equal political liberties, and only those liberties, are to be guaranteed their fair value.
>
> b. Social and economic inequalities are to satisfy two conditions: first, they are to be attached to positions and offices open to all under conditions of fair equality of opportunity; and second, they are to be to the greatest benefit of the least advantaged members of society.[5]

[3] See John Rawls, *A Theory of Justice*, rev. ed. (Cambridge, MA: The Belknap Press of Harvard University Press, 1999): §3, "The Main Idea of the Theory of Justice."

[4] J. Rawls, *A Theory of Justice*, p. 79.

[5] These are Rawls' statement of his principles of justice in John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993). For a different, earlier

However, Rawls' argument for these principles is predicated upon, among other things, an assumption of *reasonably favorable conditions*, or conditions of "moderate scarcity."[6] This assumption, obviously, is crucial in the present context. A "broken world"—the kind situation are concerned with in this paper—is not a world of moderate scarcity, but rather one of *extreme scarcity*: it is a world in which there are not enough natural resources for everyone enjoy and exercise traditional liberal-democratic liberties. Thus, if Rawls' conception of justice as fairness is to be properly extended to a broken world, the assumption of reasonably favorable conditions must be *replaced* with an assumption of a broken world. Let us now investigate the implications of doing so.

III

Mulgan's "Broken Original Position" Arguments for a Fair Survival Lottery

Mulgan proposes that the parties to a "broken original position" should assume, behind a Rawlsian veil of ignorance, that they are deliberating to principles of justice to govern a world where:

a. "Breathable air, drinkable water, arable land, and fuel of all kinds are scarce resources that must be conserved and rationed",

formulation see John Rawls, *A Theory of Justice* (Cambridge, MA: The Belknap Press of Harvard University Press, 1971), 302-3.

[6] J. Rawls, *A Theory of Justice*, p. 110. Rawls also predicates his theory on an assumption of "strict-compliance" (see pp. 4-5, 8-9, 216); however, this assumption need not concern us here.

b. "familiar species have disappeared[…] "

c. "many regions that once housed vast civilizations have either sunk beneath the waves or become too arid and hot to sustain life; and[…] human beings live only in higher latitudes, far from the tropics."

d. "Rainfall levels and sunshine hours are largely unpredictable, while extreme weather events such as floods, hurricanes and tidal waves are much more common[…] "

e. "To make the most of good times, without knowing whether they will last for months or decades or days. In bad times, food production falls below what is needed to meet the needs of even a minimal population."[7]

In short, the parties to a "broken original position" are to assume that they are deliberating about a world of (A) scarce and uncertain resources, which (B) make it impossible to predict accurately how many people can be expected to survive from day to day, month to month, and year to year.

Mulgan then simply assumes that some kind of survival lottery is the only viable option for dealing these types of conditions. Mulgan writes:

> Rawls used his original position to design ideal liberal-democratic institutions. Similarly, we want *our* original position to help *design* a survival lottery. We don't ask whether to design a lottery. (Like Rawls' disciple, we say, "Only a fool would ask that question!") We seek a theory of justice for a broken-world society organized around a fair social lottery.[8]

Before we examine whether this is really the case—whether a survival lottery is necessary or just in a broken world—let us examine precisely what kind of survival lottery Mulgan argues free

[7] Tim Mulgan, *Ethics for a Broken World*, pp. 9-10.
[8] Ibid, p. 187.

and equal individuals in a broken original position would rationally agree to.

Mulgan assumes, following Rawls, that the parties to a broken original position would rationally desire *social primary goods*: rights and liberties, opportunities, income and wealth, etc.[9] Mulgan then assumes that because primary goods are scarce resources in a broken world—because "a fair distribution [of primary goods that] gives everyone a bundle that is adequate for a worthwhile life[…] [is] impossible [in a broken world]"[10]—the parties to the broken original position to must agree to some kind of survival lottery in which each person is awarded a lottery ticket that affords them some chance of survival ("In my new broken original position, you know that you must accept *some* survival lottery, and you want one in which everyone has some chance of survival"[11]). Finally, following Rawls, Mulgan suggests that the parties to the broken original position should use the same *maximin* reasoning that Rawls ascribes to the parties in his original position: a strategy of reasoning that maximizes the best outcome for the worst off.[12] This leads Mulgan to defend the following principle of justice for a broken world:

> *Broken general conception (BGC)*: Each person is to receive the most valuable ticket (in a lottery over bundles of social primary goods) that can be guaranteed for all, unless an unequal distribution of tickets is to everyone's advantage.[13]

Although Mulgan never clarifies precisely what a lottery would have to do in order to satisfy this principle, the crucial thing

[9] Ibid.
[10] Ibid.
[11] Ibid.
[12] Ibid., pp. 187-8.
[13] Ibid., p. 188.

about it is this: every person's access to primary goods—whichever bundle of basic rights and liberties, opportunities, income and wealth they receive *enabling them to obtain scarce survival goods* (e.g. food, water, arable land, etc.) would be entirely a matter of chance. Whoever loses the lottery is, quite literally, *out of luck*: no matter what they do, they can never get more primary goods (all-purpose means for obtaining survival goods) than they receive (or do not receive) as a result of the lottery.[14] Let us call this proposal Mulgan's "Maximin Lottery."

Mulgan then introduces a second potential application of Rawls' theory to a broken world. Mulgan suggests that because different people have different attitudes towards risk—"Risk-averse ascetics are content with modest bundles of primary goods, while ambitious gamblers accept a lower possibility of surviving at a higher level of wealth"[15]—it might be rational for the parties to the broken original position to agree to a survival lottery that allocates *flexible tickets*. This is a rather peculiar proposal on Mulgan's part, given that Rawls explicitly argues that the parties to the original position should deliberate as if they are risk-averse.[16] Indeed, if Rawls' arguments for risk aversion in the original position are right, Mulgan's proposal here is simply a non-starter. Still, let us examine Mulgan's proposal on its own terms. Mulgan asks us to, "Consider a very simple society with two groups: risk-averse ascetics and ambitious gamblers. Your lottery has two types of ticket: Safe (high probability of a small bundle) and Risky (lower probability of a larger bundle). People

[14] Of course, luck and individual talents will affect survival odds, as well. However, the point of Mulgan's proposal, I take it, is that the aim of this survival lottery is to distribute goods *that contribute* to human survival to the maximum advantage (i.e. survivability) of the worst off. I thank an anonymous reviewer for encouraging me to clarify this.

[15] Ibid., p. 189.

[16] J. Rawls, *A Theory of Justice*, §§26 and 29.

choose the ticket they want."[17] Mulgan then suggests that as long as everyone is able to pick a lottery ticket that conforms to their most desired level of risk-versus-reward, every person in the broken original position would be comfortable (or "relaxed") with the survival lottery once the veil of ignorance is removed.[18] Next, because Mulgan assumes that because a rational contract is one that everyone in the broken original position is comfortable with (or "relaxed" about), Mulgan concludes that it is rational for the parties to agree to such a lottery.[19] Next, Mulgan imagines what such a society might be like:

> One possibility is a class-based society with two groups: workers and aristocrats. Aristocrats have a better life, but they are disproportionately sacrificed whenever the population must be reduced. Unlike the class-based societies of the distant past, this society would lack resentment and envy. With their different values and attitudes to risk, everyone is equally content with her lot. Workers don't want to trade places with aristocrats, or vice versa. The society is thus both just and stable.[20]

Finally, admitting that real societies are more complicated than this, Mulgan suggests that a just broken-society should involve a *fair procedure* for developing such risk-based survival lotteries, and that a procedure for developing such a lottery would be fair and just "*if* you are relaxed about living in a society governed by that procedure."[21] Let us call this second proposal Mulgan's "Choose-your-own-risk Lottery."

Mulgan never explains which of these two proposals—(1) the Maximin Lottery or (2) "Choose-your-own-risk Lottery"—he

[17] Tim Mulgan, *Ethics for a Broken World*, p. 191.
[18] Ibid., pp. 191-2.
[19] Ibid., p. 192.
[20] Ibid.
[21] Ibid.

believes to be more defensible. Fortunately, this is immaterial for our purposes. We will now see that neither proposal is defensible.

IV

Problems with Mulgan's Arguments

Let us first examine whether it is rational for individuals in a broken original position to agree to either of the survival lotteries Mulgan proposes. First, consider Mulgan's:

> *Broken general conception (BGC):* Each person is to receive the most valuable ticket (in a lottery over bundles of social primary goods) that can be guaranteed for all, unless an unequal distribution of tickets is to everyone's advantage.

Could the parties to a broken original position rationally accept this principle, given Rawls' point that a rational contract is one that individuals behind the veil of ignorance would be disposed to *keep* once the veil is raised?[22] In order for it to be rational for the parties to accept BCG, each individual in the broken original position would have to be *willing to accept its implications* should they turn out to be on the "losing end" of the lottery. Losing the lottery, however, involves receiving fewer primary goods than other people—fewer basic rights, liberties, opportunities, and income and wealth *for obtaining scarce survival goods* (food, drinkable water, arable land, etc.). Losing the lottery may, in other words, essentially consign a person to death (if, for instance, there are not enough scarce goods, and they are not awarded rights to those goods). But of course consignment to

[22] J. Rawls, *A Theory of Justice*, pp. 153-4.

death is hardly something that anyone in the broken original position—behind its veil of ignorance—would be willing to accept and want to uphold should it turn out to their fate (almost everyone, presumably, will want to live once the veil is raised). Thus, it is irrational for the parties to agree to Mulgan's BGC principle on *precisely* the grounds Rawls gives for rejecting utilitarianism: a survival lottery leaves each person's life or death to mere chance—something that many people in the real world desperately do not want, and which individuals in a broken original position should therefore want to avoid.[23]

There is a more technical way to drive Rawls' (and my) point against randomizing home. Rawls argues that anyone behind the veil of ignorance should treat themselves as having three higher-order interests that should guide their deliberations.[24] First, because every person behind the veil of ignorance knows that they are some real person, with real goals, "on the other side" of the veil of ignorance—that is, they know they will turn out to be someone with particular life-goals of their own—the parties should treat themselves as having a higher-order interest in enabling every person they could turn out to be to pursue their actual life goals—whatever they are—effectively. Secondly, however, the parties should know behind the veil of ignorance that every person they can turn out to be is a human being capable of rethinking, revising, and pursuing new life goals. People rethink and revise their life goals all the time, after all. They change career paths, change their minds about whether to have (more) children, decide to end their marriages, etc. Accordingly, Rawls argues that the parties behind the veil of ignorance should want to enable every person they can turn out to be (once the veil is raised) to be able to rethink, revise, and

[23] Ibid., §§3-6 and 44.
[24] J. Rawls, *Political Liberalism*, pp. 74-5, 106.

pursue new life goals. Finally, because the parties are assumed to be seeking an agreement on principles of justice, and have an interest in upholding whatever principles they agree to, Rawls argues that the parties should treat themselves as having a higher-order interest in understanding and upholding fair principles of justice (i.e. whatever principles they agree upon).

These three higher-order interests reveal precisely why it is irrational for the parties to the original position—to any form of it, including a broken original position—to agree to any kind of principle that involves randomness, including any form of survival lottery. Any person who agrees to a randomizing principle might rethink, revise, and want to pursue new life goals that are inconsistent with the randomizing principle's results. We saw this clearly above. It is irrational for the parties to a broken original position to agree to a survival lottery—any survival lottery—for the simple reason that they might not want to accept its results if they turn out to be on the losing end. Given their higher-order interests, it is rational for the parties to seek a better, *non-randomizing* option, an option that enables people to pursue, rethink, and revise, whatever goals they might have, including any anti-survival-lottery goals they might have.

The very same problem afflicts Mulgan's second proposal: the "Choose-your-own-risk Lottery" that distributes different tickets to people depending on their most favored level of risk-aversion. First, as Rawls argues, a rational agreement, again, is one that a person would be willing to uphold once "the results of the agreement are in." But now consider, on the one hand, someone—a gambler—who elects a Risky lottery ticket. Such a lottery ticket requires the person to die if there is a sudden downturn in the availability of scarce resources (as Mulgan writes, "[…] they are disproportionately sacrificed whenever the

population must be reduced"[25]). Such a person would absolutely *not* be willing to uphold this result. They would not "go quietly", submitting willingly to their death (even though it is what their lottery ticket requires). They would rather live in a situation in which they did not have to select a Safe or Risky lottery ticket at all. Similarly, consider a person who selects a Safe ticket, one that only gives them enough scarce resources to survive for a shorter amount of time. Suppose, as it turns out, that even though they live in a broken world, there is a significant period of abundant resources, and the "Aristocrats" in their society (i.e. people who picked the Risky ticket) all get enough resources to live 10 or 20 years longer than those who picked the Safe ticket. Would such a person really be "relaxed" about (or accept and be willing to uphold) the results of such a lottery? Mulgan contends that a society that conformed to such a lottery "would lack resentment and envy. With their different attitudes to risk, everyone is equally content with her lot. Workers don't want to trade places with aristocrats, or vice versa. The society is thus both just and stable."[26] But this intuitively seems false in a broken world. During periods of relative abundance, Workers would be likely to envy and resent the Aristocrats. The Workers would say, "Why should I only live to be 30 when, due to our current period of abundance, the Aristocrats get enough resources to live to age 50?" Conversely, during periods of severe scarcity, the Aristocrats would be likely to say, "Why should I keep up my end of the bargain? I know I selected a Risky ticket, but I do not want to die." The idea that the parties to a broken original position would accept and willingly uphold such implications once the veil of ignorance is raised is simply implausible. But, insofar as this is the case—insofar as they cannot be comfortable with agreeing to the

[25] Tim Mulgan, *Ethics for a Broken World*, p. 192.
[26] Ibid.

Choose-your-own-risk Lottery—they should rationally reject that lottery, at least if some better alternative is available.

Finally, Mulgan's case for the Choose-your-own-risk Lottery makes one additional fatal error. Mulgan assumes that everyone is willing to endure *some* level of risk in a "survival lottery." However, this cannot be assumed. Some people they "could turn out to be" once the veil of ignorance is raised may be unwilling to endure *any* level of risk in a lottery. Such people may be willing to endure significant risks to their life in some domains—for example, hunting or gathering—but not be willing to endure any risk at all in a lottery. It is crucial to understand why this is the case. The parties to the original position, if you recall, are not permitted—thanks to the veil of ignorance—to know anything contingent about themselves. This means not only that they cannot know their own race, gender, religion, talents, etc., but also, whether they are willing to endure any level of risk at all in one domain or another. If I may, allow me to use myself as an example (I am a perfectly relevant case, after all; I am person like any other, and therefore should be considered in an original position, even a hypothetical one for a broken world). The idea of selecting a ticket, the implications of which are that whether I live or die is merely a matter of luck, is absolutely abhorrent to me. I think I would prefer to (A) fight or compete for scarce resources and lose—i.e. not have enough to survive—over (B) leaving my survival to luck alone. But now, if there is anyone like me in the world at all, the parties to the broken original position must take seriously, behind the veil of ignorance, that they could turn out to be me. They must, in other words, take seriously the possibility that they could turn out to be someone who is fundamentally opposed to risking their survival in any survival lottery whatsoever. Mulgan appears not to have countenanced this obvious possibility: that there are some people who, due to contingent facts about themselves—their religion, their

personality, etc.—might be fundamentally against accepting any sort of survival lottery. Mulgan is not entitled to assert that such people do not exist, or even that their preferences for not accepting a lottery are somehow "unfair" to others. For the original position itself—the broken original position, in this case—is supposed to be a model of social and political fairness: its output—the agreement its parties reach—is supposed to specify what is and is not fair in a broken world. Building in a tacit assertion that a survival lottery (and only a survival lottery) is fair is simply question-begging. In order to know whether justice as fairness permits, prohibits, or requires a survival lottery in a broken world, we must ask which principles of justice individuals behind its veil of ignorance would rationally agree to given the assumption that they could turn out to be anyone at all, including people who might be disposed to reject a survival lottery.

V

An Alternative Proposal: A Fair Competition to Earn Scarce Goods by Contributing to Human Survival

A broken world, again, is one in which resources needed for survival are *scarce and unpredictably available*. Such a world contains "times of plenty"—times when there are more than enough resources for larger numbers of people to survive—which alternate unpredictably with times in which there are not enough resources for even a "minimal" population to survive. A broken world, in other words, is a world that is so unpredictable that, although small and larger populations may thrive at times, there

will be other times, in every population, that *some people must die so that others can live*.[27]

Throughout his book, Mulgan assumes that a survival lottery is the only viable way to respond to a broken world. However, there are surely other ways to deal with such a world. Consider, for instance, a competitive environment, in which every person is given an equal right and opportunity to *compete* for scarce resources (e.g. food, water, etc.), where the "winners" of the competition (those who get enough resources to survive) are those who demonstrate themselves *the most capable of contributing to the survival of others*, both long-term (by, say, developing new technologies for growing crops) and short-term (by, say, being particularly capable hunters of scarcely available animal prey, for food). There is no lottery here. On this proposal—call it the *Fair-Competition-to-Contribute-to-Human-Survival* proposal—each person has *equal rights and opportunities* to earn or forfeit scarce resources in a competition to contribute most to the survival of *all*, including the least well-off. Notice, first, that this proposal would actually seem to fit better with Rawls' "general conception" of justice: the conception which holds that justice requires equality, except when inequalities are to the advantage of all, including the worse off. In contrast to Mulgan's survival lotteries—both of which leave the life-prospects of everyone up to chance—this new proposal gives *everyone* the right and opportunity to compete for scarce survival goods. Why doesn't Mulgan consider this alternative?

Some readers might object that, in essence, this scheme is just a different kind of survival lottery. After all, as Rawls pointed out himself, how "capable" a given person is—how hard they are willing to work, how talented they are, etc.—is *itself* a matter of chance: namely, the "natural lottery." As Rawls writes,

[27] Ibid., pp. 9-10.

> [I]t is incorrect that individuals with greater natural endowments and the superior character that has made their development possible have a right to a cooperative scheme that enables them to obtain even further benefits in ways that do not contribute to the advantages of others. We do not deserve our place in the distribution of native endowments, any more than we deserve our initial starting place in society. That we deserve the superior character that enables us to make the effort to cultivate our abilities is also problematic; for such character depends in good part upon fortunate family and social circumstances in early life for which we can claim no credit.[28]

These have, however, always been some of Rawls' more controversial claims.

First, although our talents and character are both plausibly partly determined by chance features out of our control—for instance, by genetics, by how well we are raised, etc.—many have argued that effort (i.e. how hard one works to make the most of one's talents) is more of a matter of free choice than luck.[29] Thus, even if there is some real element of chance involved in the alternative I proposed—the proposal that each person in a broken world should enjoy equal rights and opportunities to earn or forfeit scarce goods—the proposal is not *simply* a "survival lottery" in which people are awarded tickets determining who survives and who dies. Instead, it gives people real personal control over whether they survive, depending on how hard they work to develop their skills to enhance the survivability of others. Now, of course, there may be some people who may be unable to compete effectively for scarce goods under the proposal I defend. A paraplegic, for instance, may not be able to hunt, or otherwise contribute to human survival—in which case, on my proposal,

[28] J. Rawls, *A Theory of Justice*, p. 89.

[29] See Kasper Lippert-Rasmussen, "Justice and Bad Luck", *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/justice-bad-luck/>: §9.

they would not earn scarce goods necessary survival. But, as we will see in more detail shortly, insofar as my proposal distributes scarce resources preferentially to those who contribute most to *human survival*, my proposal indirectly maximizes *every* individual's odds of survival, including the paraplegic.

Second, as Susan Hurley argues in an influential article and subsequent book[30], there are two types of luck: *thin luck* and *thick luck*. Thin luck is a kind of luck that precludes responsibility. If I fall out of an airplane without a parachute, there is simply nothing I can do to avoid hitting the ground. It would be wrong to hold a person responsible for this kind of luck. Thick luck, on the other hand, does not preclude moral responsibility, and it is the kind of luck we have in receiving our natural talents. It is nobody's fault how smart they are, or how nice they are, etc. However, Hurley points out, even though these things are matters of luck, we are still morally responsible for how we respond to them. So, for example, consider a petty criminal who, due to having low natural intelligence and poor upbringing, commits a theft. Although their criminal actions were partly the result of luck (their upbringing, etc.), we do not think they are merely a matter of luck, or that the criminal is not morally responsible for their crimes. Finally, because of this—because people are morally responsible for their choices even when those choices are partly due to the "natural lottery"—Hurley contends that that the natural lottery is irrelevant from the standpoint of justice: it is fair and just to hold people responsible for their choices, even though their choices emanated in part from luck. Hurley's broader point, in other words, is this: insofar as social-political philosophy should treat

[30] See Susan Hurley, "Luck, Responsibility, and the Natural Lottery," *Journal of Political Philosophy* 10 (2002), pp. 79–94, and Susan Hurley, *Justice, Luck and Knowledge* (Oxford: Oxford University Press). Also see Lippert-Rasmussen, "Justice and Bad Luck," §§3-4.

people as morally responsible agents, and thick luck (e.g. "the natural lottery") is compatible with moral responsibility, social and political philosophy should not treat the natural lottery as "mere luck" to be mitigated by social-political institutions. Social and political philosophy should instead be concerned with giving people equal rights and opportunities to exert control over their lives despite whatever luck results from "the natural lottery."

Now, of course, some readers may take issue with Hurley's move here, and indeed, argue that it misses Rawls' more basic point, which is that justice should not arbitrarily advantage or disadvantage people on the basis of contingencies out of their control (which "the natural lottery" is).[31] My point, however, is not that Hurley is correct. The extent to which people are considered responsible for their choices in a (broadly Rawlsian) theory of justice—even if those choices are affected by the natural lottery—is a long-debated issue that cannot be settled here. My point is simply that Rawls' claims about the natural lottery—that how hard a person works is in large part determined by their upbringing, etc.—is one of the *more contentious* aspects of his theory of justice, and for roughly the kinds of reasons Hurley gives. Yes, the natural lottery is out of our control—but, many people want to say, how hard we work, how much we develop our talents, etc., are still the result of *free and responsible choices* we make; choices that a good theory of justice should hold us responsible for, not abstract away from as "simply another contingency" out of our control.

With these points in mind, consider an essential difference between Mulgan's idea of a survival lottery and my alternative: a social scheme in which everyone has an equal right and opportunity to compete for scarce goods, by proving their "value to humanity." Mulgan's survival lotteries *only* involve luck. Once a

[31] I thank an anonymous reviewer for encouraging me to clarify this.

person has a lottery ticket, there is nothing they can do to exert control over their fate: either they will receive primary goods (basic rights, liberties, opportunities, etc.) necessary for obtaining scarce survival goods such as food, water, etc., to survive, or they will not. Thus, whether a person under one of Mulgan's survival lotteries lives or dies is *merely* a matter of whether that person's ticket is selected through a random process. The scheme I am proposing is very different. On my scheme, even if individuals' natural talents are determined in part by a random process (e.g. the "natural lottery"), each person is still to be given—as far as possible[32]—*equal rights and opportunities* to compete with others for scarce survival goods on the basis of their contribution to human survival. As we will now see, although this scheme does not completely eliminate luck—individuals with lesser talents will not be able to compete *as effectively* as people with greater natural talents—it both (A) minimizes the effects of luck on individuals' life prospects, and (B) maximizes the survivability odds of those who cannot compete equally or effectively.

VI

Justice as Fairness in a Broken World

Consider now the following alternative principle of justice for

[32] Obviously, some people – those with physical or mental disabilities, for instance – may be unable, or less able, to exercise the rights and opportunities my proposal involves (viz. competing for scarce survival goods through contributing to human survival). However, as I will explain in more detail later, my proposal, even if not all can exercise the relevant rights and opportunities effectively, still *maximally benefits everyone* in broken conditions, including those who are hindered in these regards (for, as we will see, my proposal maximizes everyone's survival odds). I thank an anonymous reviewer for encouraging me to clarify this.

a broken world—the scheme that I proposed earlier:

> *Broken-World-Justice as Broken-World-Fairness*: each person in a broken world is to be afforded an equal right and opportunity to *earn* access to scarce survival resources (e.g. food, water, medical care) in *direct proportion to their contribution to a social-political scheme that maximizes human survival.*

I call this principle "Broken-World-Justice as Broken-World-Fairness" to emphasize that this principle would be inappropriate and unjust for a world like ours: a world in which there are, in principle, enough resources for all to survive (note: although famine and lack of medical care do exist in our world today, this is not because there are not enough resources, but rather due to a lack of social and political will. As Thomas Pogge and others have argued, there are in principle more than enough resources in our world for all to survive[33]). The principle embodies a kind of "broken justice"—the maximum amount of fairness possible in a brutal, unfair, broken world where not everyone can survive.

Allow me to explain how I want to understand the principle. I assume that there are practical ways to measure how much any given person in a broken society contributes to human survival, both in present and future generations. A person who designs new farming technology that, say, enables society to grow more abundant crops under inhospitable conditions might contribute in some measurable way to an increase in average-lifespan (ALP), an increase of "healthy productive life years" (HPY) in which people in the society are able to work effectively for the common good, and overall survival rate (OSR), or how many people are capable of surviving any given time. In turn, individuals who are capable

[33] Thomas Pogge, *Politics as Usual: What Lies Behind the Pro-Poor Rhetoric* (Malden, MA: Polity Press, 2010).

of using that technology (e.g. "operators") might also contribute some smaller amount to each of those measurables. To make a long story short, people living in a "broken society" might devise some kind of formula for quantifying each person's overall contribution to survival. The Principle of Broken-World Justice as Broken-World Fairness then simply requires giving everyone in society an equal right and opportunity to earn shares of scarce resources. Whoever in society contributes the most to overall human survival earns the greatest shares of scarce resources, enabling them to live longer and therefore contribute more in an ongoing basis, up to the point at which (due, perhaps, to declining abilities in old age) they are no longer able to contribute as effectively. Conversely, those who contribute the least to ongoing human survival are awarded the smallest shares of scarce goods—shares which, depending on prevailing conditions, may or may not be sufficient for such people to survive (i.e. leading to their death).

The question for us is whether it is more rational for the parties to a broken original position to agree to the principle of Broken-World-Justice as Broken-World-Fairness than any other principle, including any kind of survival lottery. I will now argue that the rationality of agreeing to the principle is plain. The parties to a broken original position will know, behind the veil of ignorance, that the principle would maximally satisfy *anyone* they could turn out to be once the veil is raised. Allow me to explain.

Let us begin with the three higher-order interests that Rawls (rightly) ascribes to the parties to the original position. The parties to the broken original position want to enable every person they could turn out to be able to (A) effectively pursue their actual life-goals, (B) rethink, revise, and pursue new life goals, and (C) understand and uphold principles of justice. Let us begin, then, with (A). Does the principle of Broken-World-Justice as Broken-

World-Fairness enable everyone to pursue their actual life-goals as far as possible in a broken world? At first glance, it might not appear to. For what about people who do not want to compete to earn access to scarce goods? What about people who might prefer to run the risk of engaging in a survival lottery instead? The answer, quite simply, is that the principle of Broken-World-Justice as Broken-World-Fairness permits such people to engage in such a lottery on purely voluntary grounds. For the principle only asserts that every person has an equal right and opportunity to compete to earn scarce goods by virtue of their contribution to humanity's survival. Such a right and opportunity is entirely consistent with people deciding, of their own free will, to exercise that right by engaging in a voluntary survival lottery (provided the lottery they freely engage in contributes effectively to human survival). All the principle of Broken-World-Justice as Broken-World-Fairness says is that no one can be forced to engage in a survival lottery. Those who want to compete to earn scarce resources—through innovation, through hard manual labor, etc.—are simply given an equal right and opportunity to compete. Thus, Broken-World-Justice as Broken-World-Fairness maximally enables everyone to pursue their first higher-order interest, at so far as that interest is consistent with their third higher-order (their interest in understanding and upholding fair principles of justice). Whatever a given person's life goals are—whether they want to engage in a voluntary survival lottery, etc.—the principle of Broken-World-Justice as Broken-World-Fairness gives everyone an equal right and opportunity to pursue their goals, at least in proportion to their overall contribution to human survival (more on this momentarily). Finally, Broken-World-Justice as Broken-World-Fairness maximally enables everyone to rethink, revise, and pursue new life goals, at least in proportion to their overall contribution to human survival. Anyone who wants to engage in a new occupation, receive new education, marry, divorce, have

children, engage in a voluntary survival lottery, etc., is given an equal right and opportunity to choose such goals, at least in proportion to their overall contribution to human survival.

Attentive readers might find something puzzling about the arguments just given for Broken-World-Justice as Broken-World-Fairness. Following Rawls, I have assumed that the parties to the broken original position have three higher-order interests. I then argued that the principle of Broken-World-Justice as Broken-World-Fairness maximally satisfies each interest, *at least in proportion to each person's overall contribution to overall survival.* However, I did not justify this italicized caveat—a caveat which places severe constraints on the extent to which any given individual is able to satisfy their higher-order interests. Indeed, Broken-World Justice as Broken-World Fairness only permits people to pursue their three higher-order interests to the extent that each person's right and opportunity to do so contributes to overall human survival. How can this limitation on their three higher-order interests be justified to individuals in the broken original position, behind its veil of ignorance?

The answer is simple. First, the parties to the broken original position can rationally assume that death is the worst possible socially-politically determined result for any given person. Although it is not always true, of course, that death itself is the worst possible outcome for every person (some people voluntarily commit suicide out of a sincere belief that their life is not worth living), death is surely the worst possible socially-politically determined result for any person. For, when we understand death in a social-political context—a context of social and political rules, and laws—the result is being forced to die by society, whether one likes it or not. If anything seems rational for the parties to a broken original position to assume, this does. As a general matter, being told by one's society that "you must die so

that others can live" is the worst outcome any individual can face. Second, because our project in this paper is to extend Rawls' theory of justice as fairness to a broken world, let us assume—as Mulgan's Student A does—that Rawls is right about the rule of social choice that the parties to an original position rationally ought to use: *maximin*, the rule which requires producing the best-possible outcomes for the worst off. Here is the point: when these two points—(1) death being the worst socially-politically determined result for any person, and (2) the rationality of *maximin*—are combined with the parties' three higher-order interests, the result is Broken-World-Justice as Broken-World-Fairness. Broken-World-Justice as Broken–World-Fairness can be justified to *everyone* in the broken original position, including the *worst off*, because it gives everyone an equal right and opportunity to survive in proportion to the extent that they better enable others to survive. The principle, in other words, can be rationally accepted both by its "winners", those who utilize their right and opportunity to help others survive (because they enjoy more scarce goods as a result of their contribution), but also by its "losers": those who do not contribute the most to humanity's survival. Why? *Because Broken-World-Justice as Broken-World-Fairness enables more "losers" to survive than any alternative principle.* "Winners" are given more scarce goods, and so longer, better lives, only insofar as they maximize everyone else's survival odds.

Finally, it is well-worth noting that there are other reasons—reasons that Rawls gives for his principles of ideal justice—for the parties to a broken original position to agree upon Broken-World-Justice as Broken-World-Fairness. Rawls emphasizes that a just society should, for obvious reasons, also be a stable one.[34] The parties to an original position—any original position, including a broken one—should not wish to agree to principles

[34] See J. Rawls, *A Theory of Justice*, pp. 154-9 and §76.

that people in the real world (once the veil is raised) will want to overthrow and replace with new principle. Social strife and instability are in no one's interest. Notice, further, that this seems particularly true in a broken world. First, every minute people in such a world might spend arguing over politics, justice, fairness, etc., is a minute that people are not contributing to human survival. Second, social strife—for instance, violent clashes, riots, etc.—may not only produce harmful social divisions, leading people in society to cooperate poorly for the social good and human survival; such things can also result in the incapacitation or even death of people who contribute effectively to human survival. Social stability and cooperation thus should be of great importance to every individual in the broken original position.

Let us compare, then, Broken-World-Justice as Broken-World-Fairness with Mulgan's two proposed survival lotteries on these dimensions. One of the things about survival lotteries that Mulgan repeatedly tries to downplay throughout his book is that the "losers" of any such lottery—people who society effectively condemns to death—are unlikely to *accept* the results of such a lottery. Mulgan briefly discusses this kind of "instability" worry earlier in his book, within his discussion of Thomas Hobbes. There, Mulgan writes:

> Hobbes insisted that anyone could resist if the sovereign threatened his life. Won't this sanction all lottery losers to rebel? This result seems inevitable, *if* we follow Hobbes and regard the universal fear of violent death as the overriding human motivation. But[…] this disposition is not universal, and must be cultivated by the sovereign. In a broken world, a Hobbesian sovereign might encourage other motivations: perhaps a sense of honour or a concern for future generations. If lottery losers feel honour bound to submit to their fate, our sovereign will sleep more soundly![35]

[35] Tim Mulgan, *Ethics for a Broken World*, p. 158.

Mulgan's argument here, however, is simultaneously over-stated and overly optimistic. First, it is over-stated in the sense that we do not have to ascribe a universal overriding fear of death to generate social instability. All that has to be the case for social instability to occur is that a significant amount of lottery losers to fear death enough to rebel against the lottery's results. Second, there is every reason to think that there would be a significant amount of such people, and that however much a sovereign, or society more generally, might attempt to cultivate a sense of honor and sacrifice in people, significant numbers of people will still be likely, at least over time, to rebel against the results of any survival lottery. One main reason to think this is that similar forms of discontent and rebellion despite social-indoctrination are a common theme in dystopian fiction. Consider, for example, the famous novel (and feature film) *The Hunger Games*. In *The Hunger Games*, in the aftermath of a great war, society has instituted a lottery in which, every year, a dozen children selected at random are forced to fight one another to the death in an arena. The expressed purposes of the lottery—which the government convinces large numbers of people to accept—is simple: it is intended to both remind people of the great costs of violence (of how many men, women, and children died in the great war), and as an expression of penance for the civilian insurrection that led to the great war (citizens are told that a dozen children must be sacrificed each year in order to atone for their predecessors' sins). Now, of course, this is a very different type of lottery than any of Mulgan's survival lotteries, but still, they intuitively share a common problem. In *The Hunger Games*, citizens inevitably revolt. As a result of one child's inspiring behavior in the arena, the common citizens become so incensed with the annual lottery that they violently overthrow the government. And of course the reason they do so is obvious enough: human beings tend not to like seeing their children

selected and killed at random. But how is a survival lottery of the sort that Mulgan defends any different? People in a broken world would surely *not* enjoy seeing themselves or their family members randomly selected for death in a survival lottery, any more than people in a "Hunger Games world" could be expected to just sit by, for generations, and watch their children die. For these reasons, no matter how effective of a propaganda machine a broken society might devise for convincing people to go along with a survival lottery, we can expect—and the parties to the broken original position can expect—for such a system to be unstable.

My principle of Broken-World-Justice as Broken-World-Fairness is very different. First, Broken-World Justice as Broken-World Fairness gives "losers"—people who do not compete effectively for scarce goods—compelling reasons not to protest or rebel. Anyone living in a society that conforms to the principle would *know* that their society's system of rewards—giving more scarce goods to those who contribute most to human survival—is the *optimal* scheme to ensure that the most people survive. Those who "lose" under such a system, in other words, could not seriously think that some other social system could have made them better off, or more likely to survive. Broken-World-Justice as Broken-World-Fairness embodies a rewards-system that enables the *most people to survive*. Thus, even "losers" under the system should be able to recognize that they have no good reason to rebel or overthrow it. No workable alternative social-political system could make it more likely that *they* survive than Broken-World-Justice as Broken-World-Fairness.

Second, although it does effectively consign some people to death—it is, again, impossible to ensure in a broken world that everyone survives—Broken-World-Justice as Broken-World-Fairness gives *everyone an equal right and opportunity* to compete to

earn scarce goods by proving their ability to contribute to the survival of others. The psychological ramifications of everyone having such a right and opportunity should not be underestimated. It is, intuitively, far easier to accept having to die if one can say to oneself, "At least I was given a right and opportunity to compete for scarce goods", than it is to accept having to die as a result of a random lottery. And while, of course, this may come as little consolation to the person who must die, any person who dies in a world governed by Broken-World-Justice as Broken-World-Fairness will know, again, that that principle them the very *best chance* of survival (since, again, it distributes scarce goods preferentially to those who maximize human survivability). Since it gives each person a better probability of living than any Mulgan-esque survival lottery, Broken-World-Justice as Broken-World-Fairness will give every person, including losers, the *correct impression* that they both live under a scheme that (A) maximizes their survival odds, but also (B) puts their fate *as much as possible into their own hands*, rather than the hands of fate—both of which can be expected to maximally satisfy every person under broken conditions *as far as is possible* under such harsh conditions.

VII

Conclusion

Tim Mulgan's book *Ethics for a Broken World* raises a timely and important issue: how social and political structures should be organized in a broken world in which there are not enough resources for everyone to survive. Mulgan suggests that existing moral and political theories generally entail that a broken world should governed by "survival lotteries": randomized bureaucratic

procedures that determine who lives and who dies. I have argued in this paper that there is a better option, and that John Rawls' conception of justice as fairness requires it. Instituting a survival lottery is not a fair and just way to respond to a broken world. Justice as fairness requires affording each individual in a broken world equal rights and opportunities to earn and forfeit shares of scarce resources in proportion to how well they contribute to helping others survive. Justice as fairness, in other words, requires a *fair competition* for scarce goods in a broken world, where the aim of the competition is to maximize human survivability. This answer is not only, I believe, justified by Rawls' theoretical framework. It is also intuitively compelling.

University of Tampa

If you need to cite this article, please use the following format:

Arvan, Marcus, "Justice as Fairness in a Broken World," *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014), 95-124, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

MORAL AND POLITICAL PHILOSOPHY FOR A BROKEN WORLD?

GOAL DIRECTNESS AND FUTURE-FOCUSED ACCOUNTS OF WELL-BEING

BY

AARON WILSON

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 125-156
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Goal Directedness and Future-Focused Accounts of Well-Being

Aaron Wilson

Abstract. Well-being may not be the sole moral good, but a comprehensive ethics which did not take such considerations into account must surely strike us as odd, if not outright naive. The claim of this paper is first of all that the value of a theoretical approach to well-being lies not just in unifying familiar examples of our experience, but as an intellectual tool for application to new, unfamiliar cases. Such cases, where simple folk theories of well-being let us down, are a by-product of changes in technology and society. A theory of well-being, if at all possible, should be constructed in such a way that it could offer us insight into these puzzling cases as they occur. I then propose a sketch of how an explicitly future-focused account of well-being might look.

 Philosophy and Public Issues (New Series), Vol. 4, No. 2 (2014): 125-156
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

I

Introduction

The goal of a moral theory should ultimately be to assist us in the making of judgements concerning how to act. Many such judgements are made on the basis of the potential impact that our actions might have for well-being of ourselves and those around us. An adequate moral theory, therefore, requires an adequate theory of well-being. The notion of 'well-being' that I have in mind here is that which we talk about when we describe a certain happening as being 'good for' an individual, and can be likened to James Griffin's account of 'prudential value'[1]. In this paper I wish to make two claims about theories of well-being. The first is a methodological point, according to which a theory of well-being is of greatest value when it has potential applications to novel situations rather than just familiar ones. The second claim I wish to make is that if one accepts my proposal concerning the goal of a theory of well-being, then we should prefer to construct such a theory in a way that does not depend on features which are biased towards the 'average' human. My own preferred approach shares some similarities with Simon Keller's theory of well-being based on goals[2], however with some marked differences. Specifically, I reject his claim that goals are best thought of as only one source of well-being amongst several, and that it is the achievement of goals by individuals *whose goals they are* which is relevant to well-being.

[1] James Griffin, *Well-being: Its meaning, measurement and moral importance* (Clarendon Press 1986), pp. 3-4.
[2] Simon Keller, "Welfare and the achievement of goals" *Philosophical Studies* 121 (2004), pp. 27-41.

II

Focusing on the Future

The majority of moral judgements that we are called upon to make are simple and intuitive, and usually don't require a systematised theory of well-being in order for us to make a pronouncement on them. When someone loses their job or catches a cold, a folk theory of well-being is sufficient to tell us that such individuals have experienced harm. Such judgements come to us readily and often direct us well, however one area where folk theories of all kinds struggle is in their ability to provide consistency, or to deal with the unfamiliar. A philosophical account of well-being is of value not just in *explaining* the familiar cases for which the folk-theory is already an adequate guide, but because the philosophical account, if it is well-fashioned, should be able to guide our judgements when the folk theory lets us down. Following Feinberg I call these unfamiliar examples 'puzzling cases'[3], those in which our intuitions are either absent or else pull us in opposite directions. Such unfamiliar cases, while obviously rarer than the everyday cases, do nonetheless crop up more often than not, especially when one takes the point of view of human history as a whole.

As societies move from past to future, intuitions about things like well-being are liable to be disrupted by the new perspectives which greater access to information brings. The Athenians were convinced that some individuals were born slaves, and that the peoples to the north were mere barbarians to whom the goods accorded to citizens, such as freedom, self-determination and the right to property simply did not apply. Nearer our own time we have questioned the entrenched views about differences between

[3] Joel Feinberg, *The Moral Limits of the Criminal Law*, volume one (New York: Oxford University Press 1987), chapter 2.

genders, races, the insane, and for the first time seriously considered a moral duty of care towards animals. Questions are likewise beginning to be asked about whether environments, species, ecosystems and cultures can be harmed. The rise of secularism in parts of the world throws the previously widespread belief in an immortal soul into doubt, changing our ideas about how and when the subject of well-being persists. In almost all cases such judgements pull away from the folk theory of the time, often because they involve expanding existing concerns to beings previously deemed ineligible for one reason or another. Such changes in the scope of well-being are an inescapable consequence of human development. If we assume that our future as a society will be anything like our past, then our existing folk theories will steadily be shown to be inadequate, and new ethical challenges will present themselves. I propose that it is in dealing with an unknown future that a philosophical theory of well-being, one that goes beyond folk intuitions is required.

Of course, we cannot look directly into the future to see what challenges await us. The best we can do in the present is to try and account for those cases in own experience which are puzzling, in that they seem to conflict with some feature of our folk theory. It seems sensible to assume that in order for a theory of well-being to be 'future-proof', it must at least be able to be account for these present day puzzling cases, ideally whilst preserving as much of the folk theory as is salvageable. In the remainder of this article, I will attempt to show how my preferred theory of well-being based on goals is better able to deal with these puzzling cases than a number of the most popular alternative positions, particularly hedonist, desire-satisfaction and objective list accounts.

My use of these terms in the context of this discussion must necessarily be quite rough, as each of these labels in fact

encompasses a wide diversity of opinion, however in the interests of clarity it would be best to be explicit as to what exactly I take each of these approaches to well-being to amount to. By a 'hedonist' account, I mean any theory which takes facts about well-being to be ultimately reducible to facts about the balance of positive psychological experiences which an individual undergoes compared with negative ones. The paradigm cases of positive psychological experiences are pleasure and happiness, compared with pain and unhappiness on the other, although more sophisticated accounts may take into consideration more complex experiences such as love, satisfaction, frustration or tranquility. A desire-satisfaction account is one which takes an individual's well-being to be grounded in what that individual *wants*. A 'desire', on this account, is usually taken to be a specific kind of psychological state which disposes an individual to seek certain outcomes. Roughly speaking, when a state of affairs is an instance of benefit to an individual this is to be accounted for in terms of that individual's having a desiring attitude to that state of affairs. Most such theories resist the claim that a state of affair's being desired is both a necessary *and sufficient* condition of that state of affairs being a source of well-being for a person, however. Self-destructive or otherwise irrational desires are usually taken to be actively detrimental to a person's well-being rather than conducive to it. Finally, an objective list account is exactly what it sounds like: a list of things which are objectively good for the person who has them. Plausible candidates for such goods include virtue, liberty and happiness. Note that hybridised versions of these approaches are also common.

III

Puzzling Cases

In this section I wish to briefly outline two cases which cause problems for folk accounts of well-being. These are the issues of death and posthumous harm and the well-being of non-humans. Note that in choosing to focus on these two cases in particular I do not intend to suggest that these are the *only* puzzling cases deserving of scrutiny, only that focus on these cases is particularly instructive.

Death, it is widely supposed, is the cessation not just of life but also of the person itself. The folk theoretical approach to death is clear: life, except in some tragic cases, is a source of well-being for a person. Things which prolong one's life contribute to their well-being, and death, viewed as the cessation of life, is a great harm. Similarly the folk are quite clear that some sources of the things we take to be good for a person, like leaving a positive legacy or the health of their loved ones, relate to events which will occur after their own death. These attitudes are so firmly entrenched in the folk theory of well-being that many would consider it trivially true that death is harmful to the person that undergoes it, and that there is some value for them (even if we are initially reluctant to call it well-being) in events that occur after death. The suggestion that my own death, shortly followed by the death of all my acquaintances and the destruction of the Earth itself would not do me any harm just seems manifestly false. I will not seek to defend the claim that we should regard death and posthumous events as sources of harm, only that this is what the folk theory claims. If we regard a fit with the folk theory as desirable, then we should naturally prefer that a theory allow for the harmfulness of death and posthumous events, in the absence of compelling reason against. The puzzle in this case, however, is that well-being is value *to a subject*, the very subject

whose existence is assumed to end with death[4]. This too is a part of our folk theory, for the idea of a well-being in the abstract, devoid of a subject, simply seems like a contradiction in terms. Siding with the folk in cases revolving around death require that our theory of well-being depend on something that is not destroyed in the transition from living to dead.

Hedonistic theories, even those which are construed broadly enough to include complex or multiply realisable mental states such as 'happiness' or 'contentment' are ill-equipped to face the problem of death, for even if there were some doubt about personal survival post-mortem, anything like an experience or a mental state as we understand it must be rendered impossible once the mortal vessel has ceased to function. Capacity for pleasure is lost, as is the capacity for any feeling of satisfaction or dissatisfaction. Desire satisfaction accounts fare better, although they too experience problems of their own. Even though we might possess a desire for things which occur outside the span of our own lives, the problem remains of whom exactly is benefitted when that desire is fulfilled. It is possible that such a theorist might allow that the satisfactions of posthumous desires benefit deceased individuals retroactively, but this view too leads to some implausible consequences. If posthumous desire-satisfactions reach back in time, why not all desires? Perhaps I am unknowingly better off today thanks to the ice-cream I will eat twenty years from now. Also, suppose that I desire ice-cream today, but change my mind tomorrow. Would the satisfaction of this desire in the future still benefit me, even though I desired it for only a day? It seems odd to assume it would, but this exactly the assumption we make when we allow desires we had when we were alive to benefit us in death. We lose our desires when we die, just as surely as if we had changed our mind about them. This

[4] Thomas Nagel, *Mortal Questions* (Cambridge University Press 1979), chapter 1.

treatment of posthumous desire is of necessity somewhat rough, and it is possible a sufficiently sophisticated account of such might fare better faced with such cases. If this is to be achieved, however, then it seems that it must be achieved by bringing the desire-satisfaction account closer to the theory I intend to defend in this paper. Indeed, my own preferred approach may be viewed as a refinement of the desire-satisfaction approach rather than a member of a truly different species.

As regards those theories which take well-being to consist in a list of some objectively valuable things, the issue is simply what goods we may envisage that could both survive an individual and still realistically be described as *their* 'well-being'? Most of the objective goods countenanced by such theories are conditions like liberty, or the development of skills, but these seem to be just as rooted in the life of the subject as previous candidates.

I take it for granted that at least some non-humans experience harm, and so possess well-being. That said there a great many factors involved in human well-being which are indeed species-specific. Commonly cited components of well-being include development of virtue or the cultivation of abilities of appreciation of certain things (which is Raz's characterisation of what is 'good for' a being[5]), both of which we generally cannot attribute to non-humans.

Most attempts at species-independent theories of well-being have focused on commonalities between human and animal well-being, arguing that animal well-being amounts to human well-being less those goods requiring distinctly human faculties to realise. We all pursue pain avoidance behaviours, for instance, and exhibit distress when our basic needs are not met. Even

[5] Joseph Raz, "The Role of Well-being," in *Philosophical Perspectives* 18 (2004), pp. 269-94.

theorists normally at odds with hedonistic approaches to human well-being have taken an approach to the well-being of animals that differs very little from what a hedonist would endorse. Walker in particular proposes an Aristotelian account of 'flourishing' which acknowledges factors such as health and comfort as necessities and therefore constitutive of the good life.[6] Although animals are incapable of the ultimate aims of a human *eudaimon*, these environmental factors form a kind of 'base' for flourishing which both man and animal can achieve, thereby providing a basis for the ethical treatment of animals.

The key assumption at work in a theory like Walker's is that all beings capable of experiencing harm or benefit are appropriately similar to typical humans for comparison to be made. Instead of trying to find a truly adaptable, neurology-independent basis for well-being that could be applied to hypothetical aliens or artificial intelligences, the attempt is once again to hope that all judgements about well-being can be guided by experience of the most familiar, cosy cases of human experience. With regards to the well-being of animals this tendency is especially troubling. Although we seemingly cannot conceive of a subject of well-being that is not a human agent in disguise, our attempts to explain animal behaviour have little in common with the way we interpret human action.

Sheep, for instance, spend most of their time grazing, and this is variously taken as evidence that either the sheep experiences pleasure from eating grass, or that it must have a deeply held desire to graze. Either way we interpret this as evidence that 'sheep well-being' must consist in adequate grazing, in this case a

[6] Rebecca Walker, "The good life for non-human animals: What virtue requires of humans," in Walker and Ivanhoe (eds.) *Working Virtue* (Oxford: Oxford University Press, 2007).

simple connection is made between behaviour and well-being. Explaining human action is rarely amenable to such simple assessment however. When a human eats, we do not always assume that this means eating is good for them in every case. A human might eat because they're bored or unhappy, or as the result of a psychological compulsion that they would do better to excise. Human well-being takes the form it does because of the way it is rooted in the human condition as a whole, in which explicitly human faculties of self-reflection shape the landscape of what is valuable, including those kinds of value which depend primarily on faculties we share with animals. I do not wish to deny that human and animal motivations are never importantly similar, only that we are unjustified in assuming that the difference between human and animal motivation is simply that humans have some additional sources of well-being which animals lack. There seems a contradiction here, in that we assume outright that animals are not so complex that their motivations can be more convoluted than they appear at face value, but then attempt to jury-rig a logic to their well-being designed for a species with a far more intricate set of goals. Our usual candidates for theories of well-being, as we have already seen, require assumptions of precisely this kind in order to apply. The difference between these cases and the ones involving death is that whereas in the former it seemed as though the necessary states of being were notably absent, in these cases the issue is that the necessary states might be altogether too alien to fit within the paradigm designed for a single species.

IV

Distinguishing the 'Subject' and 'Vessel' of Well-Being

Perhaps the most seemingly intractable obstacle thrown up by the puzzling cases is that there are times when we want to allow that a person's well-being can be affected even when the subject of that well-being cannot be easily established. In the earlier discussion of death for instance, the concern was raised that dead persons cannot have well-being because they had ceased to exist as persons, and as such we were unable to reconcile our intuitive notion that death is usually harmful to us with the equally intuitive notion that harm requires a subject. One option we have already explored is to claim that certain events that occur after death retroactively affect the value of a person's life. For example, Bertrand Russell's life might have been more valuable had his efforts campaigning for nuclear disarmament actually hastened the abolition of atomic weaponry. Although this is quite a plausible assessment of how posthumous events can affect the value of a life, it is unclear why we should regard this value as 'well-being' rather than some other kind, such as moral value[7]. Well-being must amount to something which is good *for* the person to whom it applies, rather than something which is merely good *about* them, which is what the Russell example appears to suggest. Compare this with the tragic case of the victims of the Dunblane massacre in 1996. It seems plausible to assert that the lives of the victims gained some value retroactively over the following weeks when the shocking events of the shooting precipitated the banning of handguns in the UK, potentially saving many lives. Few would argue that this enhanced the *well-being* of the victims, though. In order to explain why well-being, as opposed to moral value or 'significance' of a life might be able to increase or decrease posthumously, we need to give a

[7] James Griffin, *Well-being*, pp. 2-4.

characterisation to the relationship between a person and their well-being which supports that assessment.

The seeming difficulty here, I suggest, is a consequence of mistaking questions of the *subject* or owner of well-being for questions of the *vessel*, or what one might call the bearer of that value. We have established that for some value to be a case of well-being there must be an individual whose well-being it is. We have yet to establish that such value *inheres* in that person, that this person is both the subject *and* the vessel for well-being, and in fact we have reason to believe it does not. The most promising strategy for reconciling death as the termination of the subject with death as a source of harm for the subject is to take the vessel of well-being to be something which *does indeed* persist beyond death, or at least remained relevant to the establishment of harm and benefit. Whatever this relevant feature is, however, *cannot* be a property of a person, because objects which do not exist cannot instantiate properties, and the death cases we examined assume the termination of the person's existence. The person's *body* does of course persist, but this is now simply inanimate matter, no more a likely vessel for well-being than sand from the beach. If it is indeed sensible to talk about well-being with reference to death and the dead, then despite the fact that well-being can only be understood with reference to a person by definition, it cannot be a condition which inheres in persons, even the person to whom well-being applies. Contrast this with other inhering states or conditions of objects, for example being a certain temperature. If an object ceases to be then it can no longer possess temperature, because that condition of being a certain temperature has no subject in which to inhere. Our best option for reconciling death with a theory of well-being, therefore, is to reject the view that well-being is a persisting state of the person altogether. It is perfectly simple to make use of concepts like 'harm' and 'well-being' without assuming that such prudential value must

correspond with such a state. Prudential value may require a subject to which it applies, but we are under no obligation to take the subject of application to likewise be the vessel of that value also. In terms of David Velleman's contrasting of 'momentary' well-being with well-being over an extended period of time[8], I can be understood as claiming that the best prospects for a theory of well-being which preserve intuitions about the harmfulness of death involve rejecting the latter notion in favour of the former.

In taking this step we would be giving up on a deeply entrenched way of thinking about well-being, although I suggest abandoning our intuitions about death would be the larger sacrifice. I suggest that this entrenchment arises from the fact that in our everyday talk it is natural to take stock of our well-being by reflecting on our psychological attitudes, which are indeed persisting conditions. Additionally we can talk about our mental states in terms of being happy, happier than we were, even the happiest we've ever been, and that we are mostly, completely or not at all content. As such the natural assumption that we are liable to make is that well-being is a *state we are in* at a given time that represents a quantity of value, much like a bank account, from which external events can add or subtract amounts. I previously claimed that we should respect the folk theory of well-being insofar as that is possible, but it appears that abandoning the bank account view will ultimately be a price worth paying.

The motivation for the bank account approach to well-being does not, it seems, stem from any inherent attractiveness to a folk theory of well-being. In fact, I suggest that there is nothing about the way we talk of well-being to suggest that *value for* a person must be *value in* a person at all. Rather, if there is any advantage to speaking of well-being in this way, then this advantage is in its

[8] David Velleman "Well-Being and Time," in *Pacific Philosophical Quarterly* 72 (1991), 48-77.

amenability to the making of social policy decisions rather than being in accordance with our most natural ways of thinking. If well-being is quantitative then it is (theoretically) comparable between people, and allows for easier judgements about the distribution of society's goods. Rawls' minmax principle requires the identification, at least in principle, of the person in a society with the lowest welfare, and insists that inequality in distribution be tolerated only if this person's welfare would be higher under such a distribution than under a more equal one. Utilitarian theories have a similar need for this quantitative account of welfare. Utilitarianism is committed to the rectitude of the greatest aggregate welfare, and as such requires that there is something which can, at least in principle, be aggregated. Even amongst these theorists however, there is a certain amount of wariness against taking the concept of comparable, quantitative welfare too literally. Rawls is keen to make clear that relative welfare cannot be easily calculated from the original position, and so brings in a notion of 'primary goods' to stand in for welfare in the judgements regarding fair distribution instead, whereas the various difficulties involved in measuring utility hardly need repeating here.

In fact, we observe that writers less concerned with the problems of aggregation already tend to bypass the bank-account approach to well-being altogether. Feinberg in his study of harm and benefit is concerned not with utility maximisation but with the practical application of the harm principle, and as such concerns himself only with occurrences of harm and not with the relation between incidents of harm and a person's overall well-being, yet his project does not suffer as a result of this difference of focus.

The precise articulation of harm being a 'setback to an interest' is also attributable to Feinberg[9], although he is more concerned to differentiate types of interests that can be harmed than explaining in detail what an interest actually is and how it is constituted. In claiming harm to be the setting back of someone's interests, Feinberg is attempting to provide the same theoretical resource that we are striving for: the facts about a person that explain why something has an effect on their well-being. The fact that he additionally considers moral wrong to be something which is to be viewed from an individual perspective is an early point of divergence between what he and I propose, according to which Feinberg's account is similar to that implied by the private ownership theory of moral value proposed by Moore[10]. It is worth noting that Feinberg's implied endorsement of this approach to moral value may simply be a consequence of his focus on establishing the boundaries of the harm principle rather than developing a comprehensive theory of welfare, a fact which is also likely responsible for his lack of detail on the nature of interests. He instead seems to take a person's interests for granted and focus on which harms to them are allowable. Nonetheless, Feinberg's notion of well-being as connected with interests suggests a promising avenue by which to develop a 'future-focused' account of well-being. Interests, I propose, should be our preferred candidate for the *vessel* of well-being.

Let us say then that the vessel of well-being is indeed an interest. We might then say that harm to a person consists of harm done to, or perhaps we might more accurately say damage or a setback to, something which is an interest of that person. Benefit, conversely, is to be thought of as promotion of an

[9] Joel Feinberg , *The Moral Limits of the Criminal Law*, volume one, chapter 1.
[10] L.W. Sumner, *Welfare, Happiness and Ethics*, (Oxford: Oxford University Press 1996), section 3.1.

interest. What I have in mind when we speak of 'promoting' an interest is simply making that state of affairs more probable, or working to bring it about in some way. This, I suggest, is an acceptably natural way of thinking about well-being that accords suitably well with our folk theory. If a consequence of adopting such a view is that we must take talk of well-being in the aggregate, or over time, to be nothing more than a useful heuristic for assessing the best, most equitable disbursement of resources then so be it.

I suggest the following terminology for interests. An interest of an individual is a *state of affairs* which has some value, either positive or negative, for that individual. For something to be 'in a person's interests', is for that state of affairs to come about or to persist. Suitable candidates for interests are not restricted to psychological states, or to states of affairs which involve persons themselves in some way. In this way, the account of well-being I offer distinguishes itself from desire-satisfaction and hedonist accounts. Positive psychological states or the satisfaction of desires may well contribute to well-being some, or indeed most of the time on the account I propose, however neither one of these is a requirement of a state of affair's having value for a person. Leaving aside the question of what makes a state of affairs an interest for the moment, taking states of affairs to be the vessels of well-being rather than some feature of persons or persons themselves has some immediate advantages for a theory of well-being.

For one thing, in disentangling the subject of well-being from the vessel of well-being, we gain the ability to talk about value for a person after that person has ceased to exist. Suppose we wish to say that it is in my interests that I have great-great-grandchildren. It is no more a problem to say that my great-great-grandchildren stand in a relation of 'valuable for' with me than it is to say that

they stand in a relation of 'being the descendants of' with me. The 'interest' relation makes no implicit reference to my being alive at the time, or to my having certain properties any more than the 'descendant of' relation does. This is not the only advantage to such an approach however.

If the vessels of well-being are thought to be unrestricted states of affairs rather than residing in a quality of persons, then we may expect an easier time speaking of the well-being of a non-human entity. Recall that many of the existing accounts struggled to accommodate the possibility that novel kinds of entity might be possessed of a capacity for well-being just as potentially rich as that of a human. The account of interests as states of affairs does not have that problem, for there are no entities which are excluded *in principle* from taking certain states of affairs as sources of value. Consider a simple, single-celled organism such as an amoeba. It is generally speaking difficult to speak of such a creature as having well-being precisely because it is assumed that the vessels of well-being are things which the amoeba cannot possess, such as psychological states. As previously noted, however, it is a virtue of a theory of well-being that it is adaptable to novel situations. While it is entirely possible that the amoeba does in fact lack the capacity to have well-being, we should consider it an advantage of a theory that it does not exclude any living thing from even the *possibility* of having well-being. We must be able to ask questions about amoeboid well-being, even if we ultimately conclude that it has none, in the same way that we may one day be called upon to ask questions concerning the well-being of machines or ecosystems.

V

The Interest-Making Relation

If the vessel of well-being is the interest conceived of as a state of affairs, we then require an account of what grounds the relationship between an individual and their interests.

The proposal I have in mind for this is in many respects similar to the theory of achievement of goals put forward by Keller, most notably in taking those things which are actively sought by an individual to be a source of value for them[11], and in advocating for an 'Unrestricted' view of such sources of well-being. While he terms these sought-out states of affairs 'goals' I prefer to retain the term 'interests', as I will ultimately seek to claim that some things can be sought after implicitly, or more or less passively in the absence of explicit adoption of that thing as a goal. Furthermore, my proposal concerning interests is in fact intended to be broader than Keller's theory of goals, as he considers the achievement of goals to only be one among many possible sources of well-being, rather than a theory of well-being itself. Additionally I reject Keller's view that it is the personal achievement of goals by the person who has them which is always a necessary component of their value to well-being.

The Unrestricted View of well-being holds that the achievement of goals, or the coming to pass of interested states of affairs in my terminology, is intrinsically good for an individual irrespective of what those goals actually are. If a state of affairs is an interest of an individual then it will be a source of some well-being to them irrespective of whether or not such goals are those that would be adopted by a rational being or whether the promotion of such a goal might in fact be a setback to some more important interest. Although it is intuitively plausible to

[11] Simon Keller, "Welfare and the achievement of goals," p. 32

hold that at least some goals make this contribution to well-being, it is less apparent that this is true of all the goals that people set themselves. To use Keller's own example, eating a handful of gravel does not seem to advance one's well-being, even if one has fervently wanted to do so their entire life[12].

The objection to the Unrestricted View on this basis is that interests themselves can seemingly be assessed as valuable independently of their status as interests in the first place. Unfortunately it is beyond the scope of this paper to present a complete defence of this position here, so I must ask the reader to simply assume it in the discussion that follows. I do however hold that the criteria I present for a state of affairs' being an interest is sufficiently sophisticated to rule out bizarre goals such as eating gravel from being true interests in most cases, although it is likely that there will be at least some instances where bizarre interests are indeed permitted by my account. In actual fact, I consider that this additional flexibility might be a valuable feature of a future-focused account of well-being. If our theory were such that all unusual or novel interests were to be excluded in principle, then we would fail in our goal of providing a theoretical account suitable to be applied to unfamiliar cases. Provided that there are adequate restrictions in place against making our account of well-being overly permissive, we should not be concerned that it allows for at least some unexpected consequences concerning the interests of entities.

In simplistic terms, I propose that an individual's interests are the states of affairs which that individual is disposed to bring about. To be the object of some kind of directed effort on the part of an agent is all that there is to being an interest. In basing my account of interests on dispositional states it could be argued that this approach is more of a refinement to the desire-

[12] Ibid., p. 28.

satisfaction based account than an outright replacement for it. I would not resist this comparison, however I take the notion of a 'desire' to be too narrow to capture what I intend when I speak of 'seeking to bring about' a state of affairs.

We do not come to understand what an individual's interests are simply by observing their behaviour and 'reading off' their dispositional states, as we would come to understand their desires. For one thing, it can be assumed that at least some behaviour is a product of coercion or of false beliefs, which do not represent the underlying attitudes of the individual in question. This is not the criticism I have in mind when I say that simply taking behaviour to indicate dispositional attitudes is too simplistic however. The desire-satisfaction theorist can account for such cases easily enough simply by insisting that only action undertaken in the absence of coercion and in the absence of false beliefs should be taken to indicate the presence of a desire. At the very least we should adhere to such a restriction in our own project, but it is not the case that in inventorying all the desires of a person we arrive at a list of their interests also. An individual's interests, rather, are calculated based on the best possible interpretation we can give to their behaviour conceived of as goal-directed, and viewed as a whole. I should make it clear at this juncture that I should not be read as claiming that interpretation is simply the means by which we *discover* or attribute interests to individuals, but the stronger claim that a person's interests just are those states of affairs which the best possible interpretation of their behaviour would indicate their interests to be. What I propose might be described as a kind of *interest functionalism*, analogous to functional accounts of mental content. Interests, as I have already argued, do not reside in the mind in anything like the manner that it is assumed that mental content is bound to the mind, but we may nonetheless take a functionalist stance on how states of affairs come to be the interests of persons. One

immediate objection that one might raise to such a view is that it appears to entail that in cases where two equally good interpretations of an agent's behaviour are possible, their interests will be indeterminate. To this objection I would first of all caution that such cases are likely to occur relatively infrequently in practice. The interpretive project I have in mind is intended to draw upon all available data in attributing interests to individuals, including (where such things are possible) their own self-reports. While self-reports are of course not foolproof guides to the best interests of a person, we should only expect indeterminacy of interests to the extent that there is ever a serious discrepancy between the behaviour of a person and *their own* sincere interpretation of the goals underlying their behaviour. If a person is indeed in such a sorry state, then why would we expect their goals to be amenable to a single coherent interpretation in the first place? Such cases may well exist, but if so, then all we need conclude is that in these rare cases, a person's interests are under no obligation to be any more settled than their owner is. This certainly need not disadvantage us much from the point of view of deciding how we should act towards such people. Even if the totality of the interests attributable to such individuals are indeterminate, chances are that many interests within that totality can indeed be attributed to them unproblematically. As to the remainder, we may simply wish to withhold judgement for the time being in the hope that new information will reveal dispositional states that were previously hidden.

The fact that understanding the interests of an individual is an inherently interpretative project is crucial to the account I propose. In our everyday dealings with people, we seldom arrive at decisions about their personalities or values on the basis of isolated incidents. Instead we observe them over time, noting their tendencies towards certain ways of behaving and approaching life in general. This was amongst MacIntyre's

insights when he claimed that human action is intelligible only through an understanding of the narrative histories surrounding those actions[13]. Whether MacIntyre was correct that narratives are the *only* way in which human lives and action can be sensibly considered remains a contentious issue (see Strawson[14] and Woods[15]), but what is of greatest importance for our project is the relationship he proposed between interpretation of any kind and the concept of a virtue. Virtues, he claimed, were not comparable to mere dispositions to act a certain way at a certain time, or even a disjunction of such dispositions. Rather they were present in different ways throughout a person's life. For example we come to recognise courage not by recognising the set of dispositions which together suffice for courage, but by locating a person's actions within a narrative that renders them intelligible. This is the proper way of understanding interests, I propose. Just as the virtue of courage cannot be straightforwardly identified with the tendency to stand one's ground in battle, an interest cannot be identified with a single set of dispositions, self-reports or desires, but must be interpreted from them when viewed within the proper context. This form of interpretation need not be strictly a narrative one, as I believe it is possible to accept my proposal concerning interests as constructed in this way without subscribing to such a view of the self or action in general. Also, in practice narrative accounts of the person commonly rely on stereotypes which impose value based on how a person is to be categorised rather than their idiosyncratic approach to life, for example narrative histories of what constitutes a good life for a woman, or a member of a slave class. On this basis, we may be

13 Alasdair MacIntyre, *After Virtue* (Gerald Duckworth & Co. Ltd. 1981), chapter 15.

14 Galen Strawson, "Against Narrativity," *Ratio* 17 (2004), pp. 428-452, at 428.

15 Angela Woods, "The Limits of Narrative: provocation for the medical humanities," *Med Humanities* 37 (2011), pp. 73-78, at 73-74.

inclined to reject a specifically *narrative* kind of interpretive project, but the importance of interpretation itself cannot be denied.

We engage in a similar interpretive activity in the process of reflection on our own interests as we do with those of others. Even if we assume that we have a special access to many of our own desires and preferences as they occur, often we will need to consider critically whether these desires are merely fleeting or evidence of a more deeply held aspiration. Our interests are quite literally a *product of* this critical reflection on behaviour and mental states. As such, we have no special access to them the way we have access to some of the features of our mental lives. This, I suggest, is quite intuitive. No child is born knowing what they want to be when they grow up, but rather must work hard to reveal their own attitudes over the course of their lives. A fleeting desire or passing fancy is not a sound basis for determining the course of one's life. Instead one must come to an understanding of oneself in order to make such a decision. Interests, as viewed by the person who possesses them, are those states of affairs which they *judge* to be good for them, rather than those which they simply feel a desire for or approval towards. Let us suppose that on a particularly trying day Van Gogh formed a strong desire to give up his artistic career and become a businessman. He probably realised at the time that this was an unusual desire for him, that yesterday he wanted to be a painter and tomorrow he would once again. This desire would only be evidence of an *interest* in becoming a businessman, rather than a fleeting desire, if it was plausibly corroborated by everything else Van Gogh knew about himself and his goals.

The distinction between narrow 'desires' and broad 'interests' in place, we can see how the interpretive project can allow us to calculate a person's sources of well-being in a satisfying way when

their attitudes and behaviour are more stable. If an athlete spends enough time training, one possible explanation is that there is value for him in one day competing in a prestigious sporting event. In that case, we would be entitled to view a knee injury that ended his athletic career as a setback to that interest and therefore harmful to him. Another, equally possible explanation however is that his overbearing parents pressured him into a life that he finds tedious and unrewarding, however after years of brainwashing he has come to internalise his parents' wishes and comes to believe in the desirability of his training, even though he gets no more satisfaction from it than he used to. In the latter case it appears that there would be no value from his point of view in his success as an athlete, so a knee injury would not be harmful to him, or at least not harmful for the same reasons. It may actually be beneficial to him, in that it enables him to leave the career that he found unrewarding with impunity. In either instance we could imagine that his beliefs and desires about what held value for him were the same, and yet we arrived at differing conclusions about what was harmful to him. Something makes the difference in this example, and that something is the broader context in which the person's attitudes and behaviour is understood, and that could only be arrived at through the process of interpretation I have described. This is the strength of an account based on 'interests' as I have formulated them, over narrower dispositional attitudes.

VI

Issues for Well-Being Based on Interests

Based on this analysis of what interests actually are, we can point to a number of consequences such an account has for the

kinds of things that can be sources of well-being. I shall now deal with a few of what I consider the most important of these.

For one, the only entities which have the capacity for well-being are those capable of behaviour sophisticated enough for the interpretive project to get off the ground. This rules out simple deterministic devices such as thermometers from having interests, since the behaviour of a thermometer cannot be viewed as directed towards any specific end. More complex non-human entities like ecosystems, however, can be interpreted in this way. Free of interference, ecosystems generally move towards the greatest possible degree of biodiversity, with a web of predatory behaviours engineered to support this goal. It is then sensible to speak of an ecosystem being harmed by any action which would decrease biodiversity within it. Secondly, as interests require goal-directedness, only contingent states of affairs can be sources of harm or benefit. States of affairs which are necessary or impossible cannot be related to well-being. It might be the case that everything I undertake to bring about in my life is dependent on the truth of some necessary state of affairs, for instance [Pv¬P], where P might stand for literally any contingent state of affairs. It seems incorrect to say however that because [Pv¬P]'s holding is a necessary condition of any interest I have being fulfilled, I derive some benefit from it, except in the most trivial sense.

It might potentially be questioned whether or not all interests that a person has actually require the process of interpretation I have described in order to be revealed. To what extent, if any, are some of the interests of persons universally possessed, at least amongst persons of a certain species? It is strongly intuitive that ensuring a balance of positive experiences in one's life over negative is an interest which all sane humans have (at least when it is not overruled by some more valued interest of theirs) simply

by virtue of their being humans and possessing the appropriate capacities for happiness and sadness. Also, all living things are programmed to promote the survival of their genes, so could this be an interest universal to all living things? I would be inclined to resist this line of argument as far as possible. In any population of sufficient size there will be outliers and exceptions to any norm. Buddhist monks or medieval flagellants had behaviour patterns that did not suggest either of these supposedly universal interests were held by them, and interpretation of a person's behaviour, in the broadest sense, is how we a person's interests are established in the first place. I am not claiming that these admittedly unusual individuals lacked any inclination towards seeking pleasure over pain, or were never even tempted to act in accordance with the biological urge to procreate. Nor is it necessary for me to deny that giving in to such temptation would even have been a source of happiness for them. What I deny is simply that mental states which are pleasurable are conducive to well-being in virtue of any intrinsic value of the state itself, in the absence of a revealed interest with which it is in accordance. In order for such states to be interests of a person, their behaviour and attitudes must direct them towards pursuit of such states. No experiences, not even those which are usually considered desirable by default receive automatic elevation to the status of 'interests' without going through the same process of interpretation as any other state of affairs.

If interests are constituted by interpreting how people act and self-report over time, then we would expect many of our interests to be quite broadly defined. Even so, there is scope in this approach for more passively held interests that are not *explicitly* sought out by an individual. No one would deny that I have an interest in a meteorite not striking the Earth and wiping out all life on the planet, despite my not being able to act in such a way as to affect the probability of this occurrence. I may never

outwardly act as though I believe such an event might be bad for me, or ever articulate it, but if one observes my behaviour and vocalisations then one will realise that many of the interests which I can be attributed as having depend for their success on all life on Earth not experiencing fiery death in the near future, which would be sufficient grounds for ruling that such a thing was against my interests.

Once we allow that some interests can be held passively, however, it could be objected that this implies that we have infinite ranges of interests at any given moment, since many such interests are merely implied by other interests of ours. At this moment in time I am writing a sentence, but precisely which interests of mine am I evidencing in doing so? It is maybe correct to say that I have an interest in finishing typing this paragraph, and an interest in finishing this paper, and an interest in a successful career as a philosopher which is best served by work on this and other papers in the future. I might even have an interest in one day being able to reflect on my life's work with satisfaction, any one of which interpretations are supported by my writing that sentence a little while ago. I do not however consider this kind of infinity a particularly troubling possibility however. Many states of affairs which have value for me will be valuable for me under more than one description. The writing of a sentence may have value for me because it is a part of a paper, the completion of which I value independently, at the same time as I value it as a part of a larger body of work, or as a job well done in its own right. Interests, it seems, may 'nest' one within the other. Some interest of mine may be promoted by the coming to pass of some other state of affairs, just as my writing of a sentence promotes my satisfaction at a body of work that I can one day feel proud of. We should not resist this conclusion, but rather embrace it. If we were unwilling to allow certain states of affairs to be good for us because their value to us were subsumed

into some other, more general goal, then our theory of well-being would be immediately crippled. Our interest in maximising our own personal well-being would be the only possible interest that could not be so subsumed. This would be an undesirable consequence of our theory, and so I propose that we allow that sources of well-being sometimes have value as constituent parts of more general goals, even if this does lead to our having infinitely many ways that we might be benefited at any time.

The existence of passive interests draws attention to an important distinction between *value for* a person and *value to* a person. When we speak of something as being *valuable to* a person, it seems that what we have in mind is that such a thing is looked upon in an approving way by that person. Quite literally, it is valued *by* them in virtue of some attitude they have to that thing. When we speak of something as being valuable for a person though, we should not be interpreted as claiming that this person necessarily has some kind of valuing attitude. Looking upon a state of affairs as good or valuable is neither necessary nor sufficient for something's being *valuable for* that person. Four-year-olds rarely have valuing attitudes towards fresh vegetables, yet obtaining sufficient vitamins from such foods are undoubtedly valuable for anyone for whom health is a source of well-being. Similarly, not everything towards which someone has a valuing attitude is necessarily relevant to their well-being. One may consider talent at playing the banjo to be valuable without in fact enjoying the music from this particular instrument. A valuing attitude may often be indicative of an interest, but is no more a guarantee that a state of affairs will be a source of well-being, taken in isolation, than a desire is.

VII

Revisiting Death and Non-Humans

I suggest that thinking of interests in this way, as states of affairs which agents undertake to bring about, can assist us in a satisfactory approach to the puzzling cases I drew attention to earlier in this paper. The issue with such cases, we established, was that sometimes the *vessel* of well-being appeared to be destroyed but the capacity for well-being was not, while at other times certain entities which appeared able to possess well-being did not meet the standards necessary for being such vessels, at least not without some pretty large assumptions being made.

The theory of interests as states of affairs aims to solve that problem by making interests themselves the vessels of well-being. According to such an account, when we speak of well-being as valuable we are in actual fact speaking of the value of certain states of affairs *for* some individual. If I claim that someone lives a charmed life of great well-being, I am claiming that their life is full of things which are valuable for them. Although I may speak in a loose sense as though their well-being is a thing which somehow resides within that person, we must avoid taking such talk too literally. The way we should understand talk of having well-being, I have argued, is analogous to the way we should talk of someone as having many admirers. The admiration directed towards someone is evidently a fact concerning them, but admiration does not *inhere* in its object. It if it is sensible to speak of such a thing as inhering at all it must surely inhere in this person's admirers, who are the vessels of that admiration. Admiration is a thing which may plausibly survive the destruction of the thing which is admired, but not the destruction of the things which do the admiring. Well-being, I have claimed, is something like this.

When I die, the facts about which states of affairs I undertook to bring about in life will remain unchanged. While I was capable of behaviour, I adopted goals, sometimes explicitly and sometimes implicitly. A corpse is incapable of such goal-directed behaviour. In any case, if we assume that death involves the termination of a person's existence altogether, it seems inappropriate by definition to take any activity we do observe from the decomposing mortal coil to be indication of the *person's* interests. Our interpretive project is therefore limited, as it should be, to the events of that person's life. Based on this, we can come to a conclusion about what states of affairs had value for that person, and acknowledge that even post-mortem, such occurrences still have this feature which qualifies them for consideration as harms and benefits.

As to the question of whether or not death itself is a harm, given what we have said so far about interests, the answer is 'sometimes'. Death itself is harmful only to the extent that it is a setback to one's interests. Quite possibly this would be true of most deaths, although by no means all. The harmfulness of death, now that it can be established to be potentially harmful at all, depends entirely on whether or not it would adversely affect any of the things which I cared about. Suppose I wished my children to continue to live long and happy lives, or that a charitable foundation I set up continues to do its work into the future, then these interests could not be harmed by my dying. Other interests of mine, for instance those interpreted from my fondness for good coffee and good books, will indeed be harmed in as much as I will never experience any of those things again.

The understanding we can offer to non-human well-being is similarly advantaged. The account of well-being I offer is based on interpretations of behaviour rather than any specific

psychological or somatic capacity. The benefit to the issue of non-human well-being is twofold.

Firstly, no entity capable of interpretable behaviour is excluded from the domain of creatures with well-being. The presence or absence of neurological features indicative of pleasure, pain, rationality, or a theory of mind or self are not relevant to determining whether or not a given entity might be deserving of consideration as having well-being. Of course, whatever knowledge we have about the capacities of such creatures must be brought to bear in carrying out the necessary interpretive project, but by taking interests as the vessels of well-being rather than psychological capacities we do not exclude any entities capable of behaviour from consideration until after their behaviour has been investigated. It might be that we ultimately deem the evidence for their having interests to be insufficient, but at least we have the ability to even ask such questions of these beings.

Secondly, a view of well-being as based on interests does not make the assumption that the human experience is the paradigm of how things can be 'good for' an individual. Much discussion of the well-being of animals, as we have previously noted, makes the assumption that the well-being of animals is like that of humans, but lacking access to those sources of harm and benefit which their more limited faculties deny them. Well-being based on goal-directed behaviour allows for the possibility that some animals may have sources of well-being that humans lack, or that both humans and animals lack some sources of well-being applicable to different, novel entities.

VIII

A Future-Focused Account of Well-Being

As useful as it is to be able to apply our theory of well-being to the puzzling cases, such theories, we have decided, are to earn their keep by their applicability to novel situations, involving circumstances and entities not currently included within the existing sphere of concern. Here, too, I propose we may be optimistic concerning the prospects of the interest based approach I describe. The theory of interests I describe is far less limited in terms of what kinds of things it can take to be sources of well-being, and what kinds of beings have the capacity for such value. Historical change has always brought with it changes of this kind, and we have no reason to doubt that it will continue to do so. My proposal is simply this: that instead of dismantling our systematised theory of well-being every time our folk theory receives a shock, we should aim to construct a theoretical account that is capable of accommodating these changes in our intuitions. As new entities or new situations begin to demand our attention, these intuitions about what can experience harm and under what circumstances are of course bound to change. This approach suggests that such change can be viewed as a change in our evidence, demanding revision of how we *apply* our theory, rather than the theory itself. If such flexibility in a theory is possible without sacrificing fit with our existing folk conceptions, we should probably pursue such a theory above others.

University of Manchester

If you need to cite this article, please use the following format:

Wilson, Aaron, "Goal Directedness and Future-Focused Accounts of Well-Being," *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014), 125-156, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

MORAL AND POLITICAL PHILOSOPHY FOR A BROKEN WORLD?

CLIMATE CHANGE, MORAL INTUITIONS, AND MORAL DEMANDINGNESS

BY

BRIAN BERKEY

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 157-189
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Climate Change, Moral Intuitions, and Moral Demandingness

Brian Berkey

Abstract. In this paper I argue that reflection on the threat of climate change brings out a distinct challenge for appeals to what I call the *Anti-Demandingness Intuition* (ADI), according to which a view about our obligations can be rejected if it would, as a general matter, require very large sacrifices of us. The ADI is often appealed to in order to reject the view that well off people are obligated to make substantial sacrifices in order to aid the global poor, but the appeal to the same intuition is much less intuitively plausible against the view that we are obligated to make great sacrifices if that is the only way to avoid severe climate change. I claim that there are no plausible grounds on which to accept the ADI with respect to addressing global poverty while rejecting it with respect to avoiding severe climate change. I conclude that we should accept that morality is far more demanding than we typically accept, and suggest two lessons of my discussion regarding the practice of appealing to intuitions in moral argument.

 Philosophy and Public Issues (New Series), Vol. 4, No. 2 (2014): 157-189
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

I

Introduction

In the Preface to his *Ethics for a Broken World*, Tim Mulgan suggests that the inhabitants of a world "broken" by severe climate change will be angry with us, the people of what he calls the "affluent" age.[1] They will "see us as the self-obsessed breakers of their world," and will, perhaps, "think of us as we think of those past generations who practiced slavery or burnt heretics."[2]

Those who reject the view that we ought to take steps to reduce greenhouse gas (GHG) emissions on the ground that climate change is not really occurring, or is not caused by human behavior, of course, share much in common with past defenders of slavery and heretic burning. They tend, for example, to have self-interested reasons for preferring that the status-quo remain in place, and attempt to justify this preference in moral terms by appealing to an all-too-convenient ideology that is wholly lacking in rational or empirical support. And those who accept the overwhelming scientific consensus that human activities are the primary cause of global warming,[3] but refuse to make any effort to reduce their own GHG emissions, and/or to advocate for policy changes that would help mitigate warming, can plausibly be described as "self-obsessed [potential] breakers" of the world that

[1] Tim Mulgan, *Ethics for a Broken World: Imagining Philosophy After Catastrophe* (Montreal & Kingston-Itacha: McGill-Queen's University Press, 2011), xi-xii.
[2] Ibid.
[3] It has recently been reported that the forthcoming Intergovernmental Panel on Climate Change report will assert that it is at least 95% certain that human activities, and in particular the burning of fossil fuels, are the principle cause of climate change. See, for example, Justin Gillis, "Climate Panel Cites Near Certainty on Warming," *New York Times*, August 19th, 2013. Accessed August 23rd, 2013. http://www.nytimes.com/2013/08/20/science/earth/extremely-likely-that-human-activity-is-driving-climate-change-panel-finds.html.

future generations will be forced to inhabit.[4] It seems, then, that members of future generations might justifiably look back upon typical affluent people of our time in the way that Mulgan imagines.

Mulgan's central concern in the book, however, is not what those living in the broken world that our activities may bring about would or ought to think about the behavior or characters of contemporary affluent people generally, but rather what they would or ought to think about what he calls "affluent philosophy,"[5] by which he means, roughly, the mainstream of contemporary Anglo-American moral and political philosophy. He suggests that there are a number of general features of affluent philosophy that inhabitants of the broken world would likely find troubling, or at least deeply puzzling, including the relative neglect of intergenerational issues,[6] the apparently pervasive assumption that future people would, at least on the whole, be better off than present people (which may help explain the neglect of intergenerational ethics/justice),[7] and the widespread reliance on individual intuitions in philosophical arguments.[8]

But while it is certainly true that moral and political philosophers have often treated intergenerational issues as less

[4] As an anonymous reviewer suggests, this description does not apply to those who accept that current affluent people have very demanding obligations, but believe that our resources ought to be exclusively, or at least nearly exclusively, devoted to addressing issues other than climate change, such as global poverty and disease. For a view of roughly this sort, see Bjorn Lomborg, *Cool It: The Skeptical Environmentalist's Guide to Global Warming* (New York: Random House, 2007).

[5] T. Mulgan, *Ethics for a Broken World*, xi.

[6] Ibid., xi.

[7] Ibid., 8.

[8] Ibid., 7-8.

than central,[9] there is clearly a growing awareness that the threat of climate change makes any approach to basic questions of ethics/justice that implies that these issues are of only peripheral or secondary concern indefensible. And this is surely at least in part because it is now widely recognized that we can no longer safely operate on the assumption that future generations will, on the whole, be better off than present people. Indeed, in nearly all of the recent philosophical work addressed to the range of ethical challenges raised by the threat of severe climate change, it is at least implicitly assumed that if severe climate change occurs, and future generations are left to occupy a broken world, we, that is, present affluent people, will have committed serious wrongs. The thought that we have an obligation to do what is necessary to avoid severe climate change, then, seems to function as what Mulgan calls a "decisive intuition," that is, as "a judgement any acceptable moral theory must accommodate,"[10] at least within a significant subset of current debates in moral and political philosophy.

Mulgan suggests that the inhabitants of the broken world would have different intuitions than many of us have in cases that are typically thought to ground objections to utilitarianism, including cases in which we can give money to charity rather than spending it on non-necessities for ourselves, and cases in which we can save many people from serious harm by seriously harming a smaller number of people ourselves.[11] If this is right, it suggests that it is a mistake to treat these intuitions as decisive, as many have wanted to. More generally, it suggests that it may be methodologically problematic to rely heavily on individual case-

[9] Mulgan makes this point explicitly with reference to the work of John Rawls; Ibid., 174.
[10] Tim Mulgan, *Future People: A Moderate Consequentialist Account of Our Obligations to Future Generations* (New York: Oxford University Press, 2006), 2.
[11] T. Mulgan, *Ethics for a Broken World*, 82-8.

based intuitions in moral argument.[12] But if it is appropriate for inquiry into questions regarding our obligations to future generations to be guided, as a general matter, by the intuition that the occurrence of severe climate change would constitute sufficient reason to conclude that current affluent people have committed serious wrongs, then we must at least allow that certain kinds of intuitions (though perhaps not individual case-based ones) can carry substantial weight in moral argument.

Roughly speaking, we might distinguish individual case-based intuitions from intuitions to the effect that certain generally describable implications of a moral theory or account of our obligations render that theory or account unacceptable.[13] By

[12] This view is powerfully defended in Chapter 4 of Peter Unger, *Living High and Letting Die: Our Illusion of Innocence* (New York: Oxford University Press, 1996); see also Tim Mulgan, "The Future of Philosophy," *Metaphilosophy* 44 (2013): 241-53, at 248-9. For a defense of the method of appealing primarily to case-based intuitions, see F. M. Kamm, *Intricate Ethics: Rights, Responsibilities, and Permissible Harm* (New York: Oxford University Press, 2007), 5.

[13] What I call "individual case-based intuitions" are referred to by Joakim Sandberg and Niklus Juth as "practical intuitions"; they contrast practical intuitions with what they call "theoretical intuitions," which they define as "intuitions about abstract moral principles or ideas, or about what makes actions moral or immoral generally and what morality is about" ("Ethics and Intuitions: A Reply to Singer," *Journal of Ethics* 15 (2011): 209-26, at 213). Their category of theoretical intuitions is similar, though not identical to my category of intuitions concerning generally describable implications. This is because their category includes intuitions about the plausibility of moral principles themselves, as well as intuitions about the moral (ir)relevance of distinctions such as that between doing harm and allowing harm, and of factors such as physical distance (Ibid., 214); mine, however, includes only intuitions to the effect that a theory or principle either must or must not have a certain kind of implication, and is therefore narrower. Nonetheless, many intuitions concerning generally describable implications will be at least closely related to intuitions about the (ir)relevance of distinctions or factors; for example, the intuition that a theory cannot imply that it is permissible to seriously harm some in order to prevent similar serious harms from being suffered by a

'generally describable implications', I mean implications that can be described without reference to particular cases, for example the implication that it is permissible to avoidably bring about a broken world, or the implication that it is permissible to seriously harm some in order to prevent similar serious harms for a greater number of others. Claims of the form "theory A is unacceptable because it has generally describable counterintuitive implication X" can, we might think, have much greater force against a view than any claim of the form "theory A is unacceptable because, in individual case Y, it has counterintuitive implication Z."[14]

If we accept that intuitions to the effect that certain generally describable implications either must follow from any acceptable moral theory, or cannot be allowed to follow from any acceptable

greater number of others is clearly very closely related to the intuition that the distinction between doing harm and allowing harm is morally relevant.

[14] If this is right, then whether and to what extent the fact that a theory has certain generally describable implications that might strike some as counterintuitive can be taken to count against the theory will depend on the extent to which the fact that those who find the generally describable implications counterintuitive is itself explained by the fact that they have intuitions about certain particular cases with relevant features. For example, I suspect that a large part of the explanation of many people's having the intuition (to the extent that they have it) that a theory is unacceptable in virtue of having the generally describable implication that it is permissible to harm some in order to prevent similar harms for a larger number of others is that these people tend to have intuitions about cases of the sort that are typically thought to constitute counterexamples to (act)-consequentialism (e.g. Mulgan's case of the Sheriff hanging one innocent person in order to prevent the deaths of several innocent people in a riot; *Ethics for a Broken World*, 83), and these intuitions are brought to mind and influence their intuitions about the acceptability of the generally describable implication. In many cases it will likely be difficult to separate out the independent counterintuitive force of a generally describable implication from that which is attributable to the influence of related intuitions about particular cases. But if we think that there are good reasons to be skeptical of appeals to case-based intuitions, there does not seem to be any clear alternative to attempting to do so.

theory, can have substantial force in moral arguments, then among the most difficult challenges for moral theorizing will be to determine how to adjudicate between such intuitions when they conflict. In the remainder of this paper, I want to explore one conflict, made salient in large part by the threat of severe climate change, that seems to take this form. I'll argue that there is reason to think that any acceptable resolution of the conflict will, given relevant features of the world, commit us to a much more demanding account of the obligations of the affluent than many philosophers have been willing to accept. More generally, I'll suggest that reflecting on the conflict can help to reveal an important type of challenge to the widely accepted practice of appealing to intuitions in moral argument. This challenge should lead us to think more carefully about the conditions in which we are justified in assigning substantial weight in moral argument to intuitions, and, relatedly, about the conditions in which we ought to be more suspicious of our intuitions (including those that we might initially be inclined to treat as decisive).

II

Climate Change, Global Poverty, and the Anti-Demandingness Intuition

The conflict that I will focus on arises in large part because of the recently recognized fact that much of our ordinary, everyday behavior, and, as we might say, our collective "way of life," threatens to bring about a broken world in which our descendants will have to live. The longer we put off taking the steps that are necessary to ensure that global temperature increases are limited to an extent that is sufficiently likely to avoid at least many of the more serious potential effects of global

warming, the more difficult and costly it will become to do what is necessary to avoid "breaking" the world.[15] Indeed, our current circumstances may already be such that doing what is necessary to avoid leaving our descendants a broken world would require massive sacrifices of resources, and more generally of the lifestyles that we are accustomed to living. To the extent that this is the case, the intuition that we are obligated to avoid leaving our descendants a broken world, which I will call the *Mitigation-Obligation Intuition (MOI)*, will conflict with the intuition, shared by many, though not all, that morality cannot require that we make massive sacrifices of resources, time, and our most valued projects in order to improve the lives of others, or to make the world impersonally better.[16] Any such requirement, according to proponents of this latter intuition, can be rejected on the ground that it is objectionably demanding.

[15] It is generally agreed that avoiding "dangerous anthropogenic interference with the climate system" (*United Nations Framework Convention on Climate Change*, 1992. Accessed August 24th, 2013. http://unfccc.int/resource/docs/convkp/conveng.pdf) requires limiting warming to less than 2° C above pre-industrial levels, although some, including the Alliance of Small Island States, and over 100 countries in total (Alliance of Small Island States. "Opening Statement, Ad Hoc Working Group on the Durban Platform for Enhanced Action," April 29th, 2013. Accessed August 24th, 2013. http://aosis.org/for-immediate-release-aosis-opening-statement-adp-2/) argue that warming that exceeds 1.5° is unacceptable, in particular because it is likely that allowing temperatures to rise by more than 1.5° will cause sea levels to rise to an extent that will threaten the existence of several small island nations.

[16] The latter of these intuitions is expressed and defended in a way that has been particularly influential in contemporary debates by Bernard Williams. See especially his "A Critique of Utilitarianism," in J.J.C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973), 110-18; and "Persons, Character, and Morality," in *Moral Luck* (Cambridge: Cambridge University Press, 1981), 10-19.

The intuition that there are significant limits on how demanding morality can be, which I will call the *Anti-Demandingness Intuition (ADI)*, has often been appealed to in response to arguments that purport to show that we have, or at least could have,[17] extensive obligations to make sacrifices in order to save the lives of some, and improve those of others among the global poor. One of the most powerful and best-known arguments for this view is that of Peter Singer, who famously compares refusing to donate a modest amount of money to an effective aid agency to allowing a child to drown in a shallow pond.[18] When iterated, this argument yields a view on which we are obligated to make very large sacrifices. There are various ways in which philosophers have attempted to explain precisely what is supposed to be objectionable about views that require such large sacrifices.[19] I will, for the most part, set aside

[17] Some reject the view that we have very demanding obligations to make sacrifices in order to aid the global poor either because they think that, at least collectively, modest sacrifices would be sufficient to provide all of the aid required to meet our obligations, or because they think that, as an empirical matter, aid is not an effective means of improving the lives of the victims of global poverty, and therefore not morally obligatory. Even if true, however, neither of these claims provides reason to reject the view that *if* very large sacrifices were both necessary and sufficient to alleviate the plight of the global poor, such sacrifices would be required.

[18] Peter Singer, "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1 (1972): 229-43.

[19] In addition to Williams's work, referred to in note 16, see Samuel Scheffler, *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*, *Revised Edition* (New York: Oxford University Press, 1994), 9-10, 55-62; Susan Wolf, "Moral Saints," *Journal of Philosophy* 79 (1982): 419-39. Also relevant is Liam Murphy's claim that there is no reason to think that any such explanation, any "underlying rationale" for the belief that there must be a limit to morality's demands, is necessary, since the belief is itself widely held and plausible, and "none of [the rationales that have been offered]…seems to have any greater plausibility than the simple claim that

these differences, since the central question that I want to consider is whether the appeal to the ADI as a means of rejecting the view that we are (or could be) obligated to make extremely demanding sacrifices in order to save or improve the lives of those suffering from global poverty and poverty-related afflictions can be accepted, assuming that the intuition that, given the threat of severe climate change, we are obligated to do what is necessary to mitigate warming (that is, the MOI) is correct.[20] The answer to this question will not depend on how we understand the ground of the ADI, since if it is illegitimate to appeal to the ADI, however grounded, in order to reject the view that we are obligated to do what is necessary to avoid severe climate change, then any defense of appealing to the ADI in order to reject the view that we have extremely demanding obligations to aid the global poor cannot simply involve referring to or developing a particular way of grounding the ADI. Instead, it will have to be claimed that there is a (or multiple) morally relevant difference(s) between obligations to avoid severe climate change, on the one hand, and purported obligations to aid the global poor, on the other, that explains why appeals to the ADI can have the force against the latter that they lack against the former.

The ADI suggests that any moral theory or account of our obligations that implies that we are required to make very large

there is such a limit" ("The Demands of Beneficence," *Philosophy and Public Affairs* 22 (1993): 267-92, at 274).

[20] I do not consider the possibility of resolving the conflict between the MOI and the ADI by rejecting the MOI here. I suspect that few will find this to be an attractive option, and it seems to me clearly unacceptable. Still, a complete defense of the view that the appeal to the ADI must be rejected with respect to obligations to the global poor would require an argument against giving up the MOI. My aim here is merely to suggest that there are substantial and generally unacknowledged costs to endorsing the appeal to the ADI in the case of global poverty relief, and that insofar as its proponents are unwilling to accept those costs, they must give up their commitment to the ADI.

sacrifices, or to radically change the way that we live our lives, can be rejected on intuitive grounds. This is, at least roughly, the basis on which many have sought to reject views like Singer's about the extent of our obligations to the global poor. But it would appear that if the ADI can justify rejection of the view that we are, or could be, obligated to radically change the way that we live our lives in order to improve, and in many cases save, the lives of the global poor, then it must also justify rejecting the view that we are, or could be, obligated to radically change the way that we live our lives in order to avoid leaving our descendants with a broken world.[21] After all, the changes to our lifestyles and standards of living that might be required in order to avoid severe climate change could be, and perhaps actually are, just as or even more radical than what would be required to ensure that all of the global poor are, for example, provided with sufficient resources and opportunities to live a decent life.

[21] I have not distinguished between several distinct ways of understanding the notion that "we" are obligated to radically change the way that we live our lives. The view might be understood to mean only that there is a collective obligation to alter our way of life that applies to, for example, all affluent people together, such that there is not necessarily also an obligation that applies to each affluent individual to radically alter his or her own lifestyle (for the view that a group can be obligated to do something without any of the individual members of the group being obligated to do their relevant part, see Frank Jackson, "Group Morality," in *Metaphysics and Morality: Essays in Honour of J.J.C. Smart*, eds. Philip Pettit, Richard Sylvan, and Jean Norman (Oxford: Blackwell, 1987)). It might also be understood to mean that there is both a collective and an individual obligation to radically alter the prevailing affluent way of life. Lastly, it might be understood to mean only that each affluent individual is obligated to radically alter his or her way of life insofar as doing so will contribute to alleviating the suffering of the global poor or the threat of severe climate change. Since the satisfaction of any of these obligations would entail large sacrifices for current affluent individuals, none of these views avoids conflict with the ADI, and so for my purposes I do not need to distinguish between them.

III

The Conflict Between the MOI and the ADI

It is, however, deeply implausible to think that we might lack an obligation to do what is necessary to avoid severe climate change because such an obligation would be objectionably demanding. Indeed, as I noted above, the intuition that we *are* obligated to do what is necessary to avoid allowing global warming to exceed certain limits (2°C, or, more demandingly, 1.5°C) effectively functions as a decisive intuition in all of the recent philosophical work on the ethical implications of the threat of climate change.[22] It is noteworthy, however, that many of the prominent contributors to the growing philosophical literature on climate change are also among the philosophers who endorse the view that we have more demanding obligations to the global poor than many others are willing to accept.[23] Dale Jamieson, for example, says that "While people can reasonably disagree about exactly how demanding morality is with respect to duties to the desperate, there is little question in my mind that it is much more demanding than common sense morality or our everyday behavior suggests."[24] Henry Shue argues that all individuals have a basic right to subsistence, where this means more than that they have a right to what is necessary for survival. In particular, he

[22] Nearly all of the views defended in the broader literature on our obligations to future generations also imply that we are obligated to avoid allowing severe climate change to occur. An exception is Thomas Schwartz, "Obligations to Posterity," in *Obligations to Future Generations*, eds. R.I. Sikora and Brian Barry (Philadelphia: Temple University Press, 1978).

[23] Among this group is Peter Singer, whose work on climate change includes "One Atmosphere," in *One World: The Ethics of Globalization* (New Haven: Yale University Press, 2002) and Chapter 9 of *Practical Ethics, Third Edition* (New York: Cambridge University Press, 2011).

[24] Dale Jamieson, "Duties to the Distant: Aid, Assistance, and Intervention in the Developing World," *Journal of Ethics* 9 (2005): 151-170, at 153.

argues that this right entails that all are entitled to "what is needed for a decent chance at a reasonably healthy and active life of more or less normal length, barring tragic interventions."[25] There are, on Shue's view, obligations correlative to this right that apply to the affluent (both individually and institutionally) to provide the resources necessary to ensure that it is satisfied, including by making significant economic sacrifices. Finally, Simon Caney suggests that a plausible set of principles of global distributive justice will include not only a universal right to subsistence, but also global principles of equality of opportunity and equal pay for equal work, as well as a general prioritarian principle stating that benefits to a person matter more the worse off she is.[26] Although Caney denies that these principles are as demanding as, for example, the global utilitarianism endorsed by Singer, and even suggests that their being less demanding should be taken to count in their favor,[27] it is clear that compliance with these principles would require that typical affluent Americans, for example, accept significantly lower incomes and greater competition for desirable positions from those who are currently effectively excluded due to lack of educational opportunities, among other causes. And these are certainly among the sacrifices that, at least in an indirect sense, proponents of the ADI believe that we are not obligated to make.[28]

[25] Henry Shue, *Basic Rights: Subsistence, Affluence, and U.S. Foreign Policy* (Princeton: Princeton University Press, 1980), 23.

[26] Simon Caney, *Justice Beyond Borders: A Global Political Theory* (Oxford: Oxford University Press, 2005), 122-4.

[27] Ibid., 123-4.

[28] Those who reject accounts of our obligations on demandingness grounds often claim that we cannot be required to, for example, give up the pursuit of our most valued projects or refrain from heavily prioritizing the interests of our nearest and dearest, even when doing so could save lives or provide desperately needed benefits for the very badly off. Since many valuable projects require large resource investments, and since the affluent would

I suspect that it is more than a mere coincidence that climate change has disproportionately been taken up as a subject of philosophical interest by those who are already inclined to accept at least fairly demanding views about other morally pressing matters. But I also expect that even those who are generally attracted to less demanding views will tend to be strongly committed to the MOI, and will not be willing to abandon it simply because it conflicts with the ADI. If we are justified in treating the MOI as a decisive intuition, then we must accept that the ADI cannot be treated as a decisive intuition (since it conflicts with the MOI). Relatedly, we must accept that any defense of the view that we do not or cannot have extremely demanding obligations to aid the global poor must be capable of explaining why we lack these obligations given that we have, or at least could have, extremely demanding obligations to avoid severe climate change. If no such explanation can be given, then we will have good reason to think that our obligations to the global poor are significantly more extensive than many, including those who have appealed to the ADI in order to reject very demanding views, have been willing to accept. We will also be forced to accept that, in a world like ours, in which the affluent must make moral choices in the face of widespread and crushing poverty, *as well as* the serious threat of severe climate change, the demands on us are very likely to be extreme.

Before moving on to consider how the view that we have potentially very demanding obligations to avoid severe climate change, but lack very demanding obligations to aid the global poor, might be defended, it will be helpful to say a bit more about

significantly compromise the interests of, for example, their children by promoting and accepting global equality of opportunity and equal pay for equal work, it is clear that adherence to Caney's principles would at least tend to impact the affluent in ways that proponents of the ADI deny that they are obligated to accept.

the conflict between the MOI and the ADI. It might be suggested, against the view that there really is a conflict between these two intuitions, that severe climate change can in fact be avoided without affluent people making very large sacrifices. What is required in order to meet our mitigation obligations, we might think, is primarily aggressive investment in research and development of alternative energy technologies to replace fossil fuels, along with further efforts to increase the efficiency with which we can use GHG-emitting fuels as we await the development and implementation of alternatives. These measures are, to be sure, far from costless, but adopting them would not require that we accept substantial reductions in our quality of life, in large part because they would not require that we make substantial short-term reductions in our energy consumption. Shue emphasizes the importance of aggressively pursuing the development of alternative forms of energy in part because he is convinced that in the absence of such alternatives, the affluent will in fact be unwilling to reduce their GHG emissions to levels that are low enough to sufficiently limit the threat of severe climate change.[29] Of course, this prediction about the behavior of the affluent is virtually certain to be correct, and so there is obvious value in focusing, as Shue does, on what we are obligated to do about the threat of severe climate change given that we can only realistically expect large reductions in GHG emissions once alternative sources of energy are widely available, reliable, and inexpensive. But we also have both philosophical and practical reasons to ask whether the refusal by affluent people to significantly reduce their GHG-emitting energy use in the short term is necessarily justified. First, as some experts seem to think,

[29] Henry Shue, "Responsibility to Future Generations and the Technological Transition," in *Perspectives on Climate Change: Science, Economics, Politics, Ethics*, eds. Walter Sinnott-Armstrong and Richard Howarth (Amsterdam: Elsevier, 2005), 266.

it may in fact already be too late to do what is necessary to avoid severe climate change without making significant near-term reductions in GHG emissions, so that, as a practical matter, it is essential to seriously consider the possibility of accepting substantial collective near-term reductions in energy use.[30] But even if we are not yet in a situation in which avoiding severe climate change requires large sacrifices in quality of life for the affluent, thinking about what we *would be* obligated to do if we were in such a situation might help to shed light on other difficult moral issues, such as the extent of our obligations to the global poor.

Consider the following case, which (apart from the clarity of the evidence assumed) may be at least reasonably close to representing the actual situation of present affluent people:

> *Dire Climate Change Threat*: While reliable non-GHG-emitting energy sources are not yet widely and cheaply available, it has become clear that in order to keep global temperature increases below the 2°C threshold, global emissions must be reduced by 40% almost immediately, and further reductions

[30] In a posting to his e-mail list from April 15th, 2013, climate scientist James Hansen notes that back in 2005 he warned that we would need to get on a path with decreasing emissions by 2015 in order to avoid "build[ing] into the climate system future changes that will be out of our control." Because we have failed to do so, he adds, "the climate dice are now loaded" ("Making Things Clearer: Exaggeration, Jumping the Gun, and the Venus Syndrome." Accessed August 26th, 2013. http://www.columbia.edu/~jeh1/mailings/2013/20130415_Exaggerations.pdf). If Hansen is right, then there is good reason to believe that we either are already, or will be within just a few years, in a situation in which avoiding severe climate change will be possible only if we make drastic short-term reductions in GHG emissions, and in which alternative energy sources are not yet widely available enough to fully make up for the necessary reductions in fossil fuel use.

> must be achieved in fairly short order. If these reductions are not made, billions of people will, a few generations in the future, unavoidably endure severe weather events that will kill many and threaten the provision of basic needs for many more, suffer from lack of reliable access to clean water, and face significant difficulty obtaining basic health care. The affluent could reduce their emissions by the amount that is necessary to avoid these results, and still have lives that are well worth living, although the reductions would entail a notably lower quality of life for nearly all affluent people.

In these circumstances, the choice that we face is, in effect, between accepting significant quality of life sacrifices, or else ensuring that our descendants will be left with a broken world. Would we be justified in doing the latter, since the former would require us to, for example, visit friends and family who live far away less often, walk, bike, or take public transit whenever possible (even when it would be much more convenient to drive), live in smaller homes, keep those homes notably cooler in the winter and warmer in the summer, purchase fewer products the manufacture and/or transport of which is energy intensive, and give up valued projects the pursuit of which requires substantial energy use?

I expect that very few people would be inclined to reject the view that we are obligated to make these sacrifices on demandingness grounds, and that instead most will continue to share the MOI in this case. If this is right, then thinking about the moral implications of the threat of severe climate change reveals that, at the very least, the ADI cannot be taken to support the view that very demanding sacrifices can never be systematically required of all affluent people in virtue of general facts about the

state of the world in morally relevant respects.[31] And since this seems to be the view that many proponents of the ADI have in fact taken it to support, the conflict with the view that I have suggested we must accept about *Dire Climate Change Threat* shows that if the view that we lack very demanding obligations to aid the global poor can be defended at all, it cannot be defended by appeal to the ADI alone. More specifically, it will have to be defended by arguing that there is a morally relevant difference, or multiple morally relevant differences, between *Dire Climate Change Threat* and the following case, which there is reason to think is at least close to representing the actual situation of affluent people with respect to current global poverty:

> *Dire Global Poverty*: Approximately 18 million people die each year (approximately 50,000 per day) from preventable, poverty-related causes.[32] Approximately 7 million of these people are children under the age of five.[33] Over 3 billion

[31] Those who appeal to the ADI in order to reject the view that affluent people generally have, for example, very demanding obligations to aid the global poor sometimes allow that particular affluent people can, at least in principle, find themselves with very demanding obligations to provide aid to particular people. If an affluent person finds himself in a situation in which the only way that he can, for example, rescue a nearby drowning child involves taking out and thereby ensuring the destruction of a boat that represents a large portion of his net worth, and that he needs in order to continue pursuit of his highly valued project of participating in sailing competitions, many who oppose demanding obligations to aid the global poor will nonetheless accept that he must make the large sacrifice in order to save the child. The ADI, then, is not typically taken, even by its proponents, to support the view that we can *never* be obligated to make very large sacrifices in order to aid others.

[32] Roger C. Riddell, "Aiding the World's Poor: New Challenges for Donor States," in *Giving Well: The Ethics of Philanthropy*, eds. Patricia Illingworth, Thomas Pogge, and Leif Wenar (New York: Oxford University Press, 2011), 86-7.

[33] UNICEF. "Progress Toward Millennium Development Goal 4: Key Facts and Figures." Accessed August 27th, 2013.

people live on less than $2.50 per day,[34] and approximately 600 million children live on less than $1 per day.[35] Approximately 400 million children lack access to safe drinking water, and around 270 million "have no access to health care services."[36] The affluent could sacrifice the resources necessary to vastly reduce, if not eliminate these deaths and deprivations relatively quickly and still have lives that are well worth living, although the sacrifices would entail a notably lower quality of life for nearly all affluent people.

What morally relevant differences might there be between *Dire Climate Change Threat* and *Dire Global Poverty* that could support the view that we have very demanding obligations in the former but not in the latter?

IV

Morally Relevant Differences?

It might be suggested that *Dire Climate Change Threat* involves a certain kind of morally important change in our circumstances, and that this explains why we have extensive obligations in this case that we lack in *Dire Global Poverty*. It might be added that the ADI is an intuition that we tend to have primarily in response to

http://www.childinfo.org/mortality.html.

[34] Patricia Illingworth, Thomas Pogge, and Leif Wenar, "Introduction: The Ethics of Philanthropy," in *Giving Well*, eds. P. Illingworth, T. Pogge, and L. Wenar, 3. The figure is in 2005 Purchasing Power Parity adjusted dollars.

[35] UNICEF. "Millennium Development Goals: 1. Eradicate Extreme Poverty and Hunger." Accessed August 27th, 2013.
http://www.unicef.org/mdg/poverty.html.

[36] Ibid.

cases in which we could prevent *widespread and familiar* types of suffering and deprivation at substantial cost to ourselves, rather than an intuition that we can never, under any circumstances, have extremely demanding obligations to respond to morally disastrous features of the world.

This view is suggested by Samuel Scheffler's explanation of his claim that "moral norms [...] must be capable of being integrated in a coherent and attractive way into an individual human life."[37] Scheffler says that this claim should be understood to mean that "within generous limits, morality makes room for personal projects and relationships. *In ordinary circumstances*, it is permissible for agents to develop and pursue a wide range of personal projects and to cultivate personal relationships of many different kinds."[38] Surely "ordinary circumstances" must include circumstances in which many millions of people die each year from poverty-related causes, and hundreds of millions more live on less than $1 per day and lack access to clean water and basic health care. After all, these are the circumstances in which the world's poor have lived for generations – there is nothing at all out of the ordinary about widespread poverty-related death and deprivation. We might think, however, that circumstances in which our behavior threatens to leave our descendants with a broken world are not ordinary at all, and that it is this fact that, in some sense, explains why we have extremely demanding obligations in *Dire Climate Change Threat*, but lack them in *Dire Global Poverty*.

Some might point to both our intuitions about cases and our actual behavior in order to suggest that we do at least implicitly take it that there is less moral reason to respond to "ordinary"

[37] Samuel Scheffler, "Potential Congruence," in *Morality and Self-Interest*, ed. Paul Bloomfield (New York: Oxford University Press, 2008), 118.
[38] Ibid., 119, emphasis added.

threats to human life and well-being by making significant sacrifices than there is to respond to unusual threats by making similar sacrifices. For example, we tend to believe that we have strong moral reasons to provide "emergency aid" in response to unusual, high-profile devastating events such as the September 11th attacks, Hurricane Katrina, and the 2004 Boxing Day tsunami, but not to believe that there are equally strong reasons to provide aid to those suffering from chronic poverty. And our behavior, at least as individuals, reflects this.[39] American households are reported to have given $1.93 billion to tsunami relief efforts, with 30% of households giving,[40] $2.8 billion to 9/11 relief efforts,[41] with 66% of households giving,[42] and $5.3 billion to Hurricane Katrina relief efforts, again with 66% of households giving.[43] By contrast, we can estimate that American households gave a total of approximately $4 billion to international aid in 2005,[44] with a portion of this total surely going

[39] Aid provided by governments is less disproportionately directed toward emergency relief, although given the scope of the suffering and death caused by chronic poverty in relation to the scope of the suffering and death caused by emergencies, government aid is still somewhat disproportionately directed toward emergency relief (Riddell, "Aiding the World's Poor," 86-7).

[40] Philanthropy News Digest. "9/11 Motivated Americans to Give When Disaster Strikes." Accessed August 27th, 2013. http://foundationcenter.org/pnd/news/story.jhtml?id=352800015.

[41] Ibid.

[42] Aaron Smith, "How Sept. 11th Changed Charity in America," *CNN Money*, September 6th, 2011. Accessed August 27th, 2013. http://money.cnn.com/2011/09/06/news/economy/katrina_donations_911/.

[43] Ibid.

[44] Individual charitable contributions for 2005 are reported to have been $199.07 billion (Rob Reich, "Toward a Political Theory of Philanthropy," in *Giving Well*, eds. P. Illingworth, T. Pogge, and L. Wenar, 190, citing Melissa S. Brown, *Giving USA 2006: The Annual Report on Philanthropy for the Year 2005* (Glenview: Giving USA Foundation, 2006)), and Americans typically give 2%

to emergency aid rather to aid the victims of chronic poverty, despite the fact that the deaths from chronic poverty-related causes outnumber the deaths from emergency-related causes by approximately 20-1.[45] In addition, the total amount provided for international aid is likely only as high as it is because of a relatively small number of larger donors; most Americans surely give nothing at all.[46]

It is difficult, however, to see what moral basis there might be for thinking that our obligations to aid are stronger in emergencies, and more generally in response to "non-ordinary" threats to well-being, than they are in the case of chronic poverty. After all, the lives of the victims of chronic poverty involve a constant struggle for survival, continual deprivation, and typically much suffering, while the lives of emergency victims may or may not have been particularly bad prior to the threat posed by the relevant emergency. Chronic poverty is, then, as Peter Unger puts it, "*far worse* than almost any emergency [...]we may say that [...][it is] a *chronic horror*."[47] Because the victims of chronic poverty have, on the whole, worse lives than the victims of emergencies (who are only sometimes among the world's poorest people), it would seem that there is, if anything, *greater reason* to provide aid to the global poor than to the victims of emergencies, all else being

of their donations to international aid (P. Illingworth, T. Pogge, and L. Wenar, "Introduction," 3).

45 R. Riddell, "Aiding the World's Poor," 86.

46 Peter Unger reports that in 1993, over 4 million Americans, targeted because of past charitable behavior, were sent appeals for donations from UNICEF, and that less than 1% of these people donated anything at all (*Living High and Letting Die*, 7).

47 Ibid., 42, emphasis in original. For similar points, see R. Riddell, "Aiding the World's Poor," 86-7; and Elizabeth Ashford, "Obligations of Justice and Beneficence to Aid the Severely Poor," in *Giving Well*, eds. P. Illingworth, T. Pogge, and L. Wenar, 42-3.

equal.[48] The fact that their suffering and deprivation are ordinary, far from making it the case that it is less morally urgent to aid them, seems to *add to* the case for prioritizing their claims over the claims of others in need. It seems clear, then, that we cannot appeal to the fact that global poverty, unlike climate change, is an ordinary type of threat to human well-being in order to defend the view that our obligations to the global poor are less demanding than our obligations to avoid severe climate change.

We might, however, accept that we have at least as much reason to aid the victims of global poverty as to aid the victims of familiar types of emergencies, but claim that it does not follow from this that we can have extremely demanding obligations to aid the global poor. After all, we seem not to be committed, either in our intuitive responses to cases or in our typical behavior, to the view that we can be obligated to make very large sacrifices in response to emergencies. Typical donations to emergency relief from well off people are quite modest, given the resources available to such people, and it is far from widely accepted that they are obligated to give more. And we might think that we can explain why we should accept that we are obligated to make much greater sacrifices in *Dire Climate Change Threat* by pointing out that severe climate change would not merely bring about a large emergency, or even a series of large emergencies, but instead would create *a much more extensive chronic horror* than current global poverty. On this view, *Dire Climate*

[48] One way to explain this thought is by appeal to a prioritarian principle stating that, all else equal, benefits to a person matter more the worse off she is; for discussion of such a principle, see Derek Parfit, "Equality or Priority," in *The Ideal of Equality*, eds. Matthew Clayton and Andrew Williams (Basingstoke: Palgrave Macmillan, 2000). As I noted above, a principle of this sort is endorsed by Caney (*Justice Beyond Borders*, 123); it is also endorsed by Thomas Pogge ("How International Nongovernmental Organizations Should Act," in *Giving Well*, eds. P. Illingworth, T. Pogge, and L. Wenar, 50).

Change Threat generates much more demanding obligations than *Dire Global Poverty* simply because severe climate change threatens to bring about even greater and more widespread death, deprivation, and suffering than is caused by current global poverty. We might say that *Dire Climate Change Threat* involves circumstances that are non-ordinary, in the relevant sense, because of the threat that the world will come to contain extraordinary levels (that is, much greater levels than we currently observe) of premature death and suffering that could have been avoided.

There are two reasons that this view is unacceptable. First, although the claim that, all else equal, we are obligated to sacrifice more in order to prevent more deaths, deprivation, and suffering is surely correct, there does not seem to be any principled reason to think that we can be obligated to make very large sacrifices in order to prevent, say, *many billions* of people from facing these ills (as, we can imagine, would occur if we did not act in *Dire Climate Change Threat*), but not in order to prevent *merely several billion* from facing similar ills (as will occur in the next few years if we do not act in response to *Dire Global Poverty*). In both cases, making the relevant sacrifices would make the world vastly better, in impersonal terms, than it would otherwise be. The fact that there may be even more potential suffering and death at stake in *Dire Climate Change Threat* does not seem sufficient to justify the view that there is a vast difference in how much we can be obligated to sacrifice in the two cases, since in both cases whatever sacrifices we do make would bring about massive improvements in the state of the world in morally relevant respects.

The second reason that an appeal to differences in the scale of death, deprivation, and suffering at stake cannot justify the view that we have much more demanding obligations in *Dire Climate Change Threat* than in *Dire Global Poverty* is that the MOI does not

lose its apparently decisive force if we assume that the effects of climate change will, on the whole, be no worse than the effects of current global poverty. Imagine an extended version of *Dire Climate Change Threat* in which it is stipulated that, whether because of technological advances, political arrangements that ensure that as many people as possible have their basic needs met,[49] or for other reasons, the number of people who die or suffer seriously from climate change-related causes attributable to our emitting behavior will be *somewhat lower* than the number of people who will die or suffer in the next several years as a result of poverty-related causes, but still well into the *billions*. Surely we do not think that in this version of *Dire Climate Change Threat*, it would be permissible to refuse to make the sacrifices necessary in order to avoid severe climate change.

Perhaps the problem with not acting in *Dire Climate Change Threat*, however, is not that more people will die and suffer than if we do not act in *Dire Global Poverty*, but rather that failing to make large sacrifices will necessarily make the world *worse than it currently is*, whereas failing to make similar sacrifices in *Dire Global Poverty* is compatible with continued improvement in the moral state of the world. This view is suggested by the "Progressive Consequentialism" that Dale Jamieson and Robert Elliot claim is worth taking seriously as a candidate moral theory.[50] This theory, roughly, says that "a [permissible] action is one whose

[49] Mulgan portrays the inhabitants of the broken world as living within political systems that are, at least generally, committed to maximizing the number of people whose basic needs are met through survival lotteries (*Ethics for a Broken World*, 10-11). It seems at least conceivable that policies of this sort could keep the typical number of people who die each year from lack of basic resources below the levels that we see today due to global poverty, even in a broken world with far fewer resources than we currently possess.

[50] Dale Jamieson and Robert Elliot, "Progressive Consequentialism," *Philosophical Perspectives* 23 (2009): 241-51.

consequences improve the world[...]what [it] requires of agents is that they act in such a way as to increase value in the world."[51] They add that according to Progressive Consequentialism, "Our mission as moral agents is to leave the world better than we found it. This struggle for improvement should be constant. The more we accomplish, the more that is demanded. Ourselves and others are held to even higher standards as the world improves."[52] Since *Dire Global Poverty* involves bad conditions that are already a part of the world, making only small sacrifices to improve these conditions will, all else equal, improve the world relative to the baseline of current conditions, and so making only small sacrifices appears to be permissible according to Progressive Consequentialism.[53] On the other hand, continuing to improve the world in the face of the threat of severe climate change may require very large sacrifices, so that Progressive Consequentialism demands such sacrifices in *Dire Climate Change Threat*.[54]

There are, however, two important problems with Progressive Consequentialism that render the attempt to resolve our conflict by appealing to it problematic. First, it is not clear that the

[51] Ibid., 244.

[52] Ibid., 245.

[53] Jamieson and Elliot add to the requirement to improve the world what they refer to as an "efficiency condition," which says, effectively, that agents must maximize the amount of improvement that they achieve given the amount of effort that they expend (Ibid., 244-5, 248). Since it will, at least typically, require more effort of agents to make, for example, a $1,000 donation to OXFAM, than to make a $50 donation, it appears that Progressive Consequentialism will not require larger donations when smaller donations will bring about at least some improvement in the world.

[54] Jamieson and Elliot are motivated to develop Progressive Consequentialism as a consequentialist response to the demandingness objection (roughly, appeals to the ADI typically made against Act-Consequentialism). But they allow that if it were to become difficult to improve the world, morality would, as a result, become quite demanding (Ibid., 242).

implications of Progressive Consequentialism regarding our obligations to the global poor can be moderate, as Jamieson and Elliot seem to suggest they are. Consider, for example, what the view might imply about how much a typical affluent person is obligated to donate to organizations such as OXFAM. On the one hand, we might think that it implies that there is no obligation to donate at all, and that it is permissible for an affluent person to spend all of her wealth on herself. After all, each purchase will improve the world, even if only by a very small amount, since, we can assume, both buyer and seller will benefit from such transactions.[55] Because of this, so long as it would require more effort of an affluent person to give any of her money to aid the poor than to spend it on herself,[56] it looks as though Progressive Consequentialism will imply that it is permissible for her to spend all of her money on herself. But this result is surely *objectionably undemanding*; even proponents of the ADI acknowledge that any acceptable view will include some (less than very demanding) obligation to contribute to aiding the global poor.

If, on the other hand, we understand Progressive Consequentialism so that it requires affluent agents to make at least one donation of any amount to aid the global poor, then it appears that whatever explains why this first act is required will also ensure that a second such act will be required, and a third, and so on. For example, if our account of what constitutes "improving the world" for the purposes of our theory implies

[55] If some such transactions would harm third parties to a greater extent than the parties to them would benefit, then they would be ruled out by Progressive Consequentialism. But some plausible theories of harm make this unlikely, if not impossible, and in any event, this will only limit the range of self-interested purchases that are permissible, rather than yielding any obligation to provide aid to the global poor.

[56] See note 53.

that in order to meet that requirement, an affluent person must make a certain sacrifice in order to aid the global poor, then it is clear that the account will also imply that further sacrifices of the same sort are required. Although such a view may be able to justify a limit to these required sacrifices that makes the view less demanding than, say, Singer's view, it seems quite unlikely that it will be able to justify a limit that proponents of the ADI will find acceptable.

The second problem for Progressive Consequentialism is even more serious. This is that it seems to commit us to a form of what I will call *reverse discounting*.[57] Because Progressive Consequentialism assigns to us the aim of continually improving the world, one thing that we will have to ensure when deciding what to do is that we do not now act in ways that will make it impossible, or even much more difficult, to continue to improve the world in the future. We will have to prefer courses of action that are likely to allow for steady, incremental improvements in the state of the world to be made over courses of action that would provide large benefits now but are also likely to lead to future actions that will bring about even small reductions in the

[57] There is debate, in the literature on climate change and more generally, about whether it is permissible for us to discount the interests of future people relative to those of present people, that is, to count the interests of future people for proportionately less than the interests of present people, simply because the former will live in the future (this is sometimes referred to as "pure time discounting," to contrast it with forms of discounting that might be justified by, for example, greater uncertainty about effects in the future as compared with effects in the present). Reverse discounting, then, involves counting the interests of future people for proportionately *more* than the interests of present people. Most philosophers reject the view that pure time discounting of future interests is permissible. See, for example, Tyler Cowen and Derek Parfit, "Against the Social Discount Rate," in *Justice Between Age Groups and Generations*, eds. Peter Laslett and James S. Fishkin (New Haven: Yale University Press, 1992).

moral state of the world, since these latter actions would be wrong, and it is surely better to choose a course of action that will involve no wrongdoing than one that will involve wrongdoing.

For example, imagine that we have the following two options: 1) We can increase the well-being of 100 current people from 50 to 51, and ensure that we will also, 50 years from now, act so as to provide an improvement for 100 people who are not yet born from 80 to 85; 2) We can increase the well-being of 100 current people from 50 to 70, but in doing so will ensure that 50 years from now we will act so as to reduce the well being of 100 people who are not yet born from 80 to 79. It appears that Progressive Consequentialism rules out choosing option 2, since it entails wrongdoing on our part, while option 1 does not. But this entails that *we must consider a small loss to better off people who will live in the future to be of greater moral importance than a much larger benefit to worse off present people.* And this means that, at least in certain kinds of cases, Progressive Consequentialism requires us to discount present interests relative to future interests. And this is clearly unacceptable. We cannot, then, accept the view that our fundamental moral obligation is to ensure that the world continuously improves, and so we cannot appeal to this claim in order to defend the view that we have more demanding obligations in *Dire Climate Change Threat* than we have in *Dire Global Poverty*.

Perhaps the most promising approach to defending the view that we have very demanding obligations in *Dire Climate Change Threat*, but not in *Dire Global Poverty*, would involve appealing to the distinction between doing harm and allowing harm. It might be argued that if we continue to emit greenhouse gases at dangerous levels, we will be doing harm to future generations, whereas if we fail to act in response to global poverty we will merely be allowing harm to come to those among the global poor

whom we might have helped. It might be further argued that although we can have very demanding obligations to avoid doing harm to people, we cannot have similarly demanding obligations to avoid allowing harm to befall people.[58]

I cannot provide a complete response to this line of argument here, but I will try to point out some reasons to be skeptical that it can succeed in defense of the view that we have very demanding obligations in *Dire Climate Change Threat*, but not in *Dire Global Poverty*. First, the much discussed Non-Identity Problem makes it questionable whether, by failing to act so as to avoid severe climate change, we would in fact *harm* future people.[59] Our deciding to reduce emissions drastically would, at least over a long enough period of time, yield a future with an

[58] An anonymous reviewer suggests that a slightly different defense of the view that we have demanding obligations in *Dire Climate Change Threat*, but not in *Dire Global Poverty*, would claim that even if bringing about a broken world would not harm anyone, it would involve actively bringing about a bad state of affairs, while refraining from aiding the current global poor would merely allow a bad state of affairs to occur. It could then be claimed that we can have demanding obligations to avoid actively bringing about bad states of affairs, but not to avoid allowing bad states of affairs to occur. It seems to me, however, that if our actively bringing about a state of affairs can be morally problematic despite not harming anyone (or, perhaps, violating their rights), this can only be because there was an alternative available to us that would have brought about a better state of affairs. In other words, the objection to what we in fact did could only be that we acted in a way that brought about a state of affairs that is worse than the state of affairs that would have come about as a result of our acting in some other way. But this is just as true of whatever we might do in preference to aiding the global poor as it is of our acting in a way that brings about a broken world. Because of this, the proposed line of defense cannot succeed.

[59] The seminal discussion of the Non-Identity Problem is in Chapter 16 of Derek Parfit's *Reasons and Persons* (New York: Oxford University Press, 1984); see also Parfit's "Energy Policy and the Further Future: The Identity Problem," in *Energy and the Future*, eds. Douglas MacLean and Peter Brown (Totowa: Rowman & Allanheld, 1983).

entirely different set of people than would have existed had we decided to continue emitting at dangerous levels. If we make the latter decision, then as long as our emitting behavior did not cause anyone to exist and have a life that is, on the whole, not worth living, no one will have been made worse off as a result of that behavior than they otherwise would have been. So, if harming a person requires making her worse off than she otherwise would have been, our behavior may not harm any future people.

Even if our continuing to emit at high levels would harm future people,[60] however, it is far from clear, given the Non-Identity Problem, that the moral reasons against harming them (assuming that they will have lives that are on the whole worth living) are as weighty as the reasons against harming present people. I may have much greater reason not to harm a present person than I have not to allow a similar harm to befall a present person, without also having much greater reason not to harm a future person, who would not exist at all if I did not harm him, than I have not to allow a similar harm to befall a present person. It may, for example, be much worse to take the food that a person needs in order to avoid going hungry for several days than to refrain from providing someone who would otherwise go hungry with similar food. But even if this is the case, it is not clear that it must also be the case that causing someone to exist who will experience food deprivation for several days (but will have a

[60] A number of philosophers have defended accounts of harm according to which it is possible for us to harm future people even if we do not make them worse off than they otherwise would have been. See, for example, Seana Shiffrin, "Wrongful Life, Procreative Responsibility, and the Significance of Harm," *Legal Theory* 5 (1999): 117-48; Lukas H. Meyer, "Past and Future: The Case for a Threshold Notion of Harm," in *Rights, Culture, and the Law: Themes from the Legal and Political Philosophy of Joseph Raz,* eds. Lukas H. Meyer, Stanley L. Paulson, and Thomas W. Pogge (New York: Oxford University Press, 2003).

life that is on the whole worth living) is much worse than failing to provide needed food to a present hungry person. Indeed, it seems at least plausible that the latter is in fact worse. If this is right, then it provides a significant reason to think that our obligations in *Dire Global Poverty* must, all else equal, be at least as demanding as our obligations in *Dire Climate Change Threat*. Since we cannot plausibly reject the view that we have very demanding obligations in *Dire Climate Change Threat*, my discussion suggests that we must accept that we have similarly demanding obligations in *Dire Global Poverty*. And since we live in a world that contains the morally significant features of *Dire Global Poverty*, and at least many of the morally significant features of *Dire Climate Change Threat*, we have good reason to believe that the obligations of the affluent are *extremely demanding* indeed.

V

Appeals to Intuition?

This result is, of course, deeply counterintuitive. But my discussion has suggested that thinking about the moral significance of the threat of severe climate change helps to reveal that appeals to widely accepted intuitions, such as the ADI, are deeply problematic. Furthermore, the problem is not that these intuitions provide acceptable guidance in familiar circumstances, but become misleading when we begin to consider their implications in unfamiliar types of cases, such as that involving the threat of severe climate change. Rather, it is that they can be generally misleading, but that we are sometimes only able to recognize that this is so as a result of thinking about what they suggest about our obligations in particular kinds of cases. If this is right, it suggests that we must be much more cautious about

appealing to intuitions in moral argument than many philosophers have been. With regard to an intuition like the ADI, we should be careful to consider whether we are willing to accept it as decisive across the full range of relevant cases. If thinking about novel cases, such as that involving the threat of severe climate change, reveals that we are not, then we must take on the difficult task of determining how best to resolve the conflict of intuitions that we face. This will often require, as I have attempted to do here, considering whether there are morally relevant differences between the novel cases and the cases in response to which we were inclined to apply the initial intuition. It will also tend to require taking seriously the possibility that our initial intuitions are seriously mistaken, and that we may have to accept surprising, and perhaps unsettling, moral conclusions.[61]

Stanford University

[61] I am grateful to the participants in the Postdoctoral Workshop at Stanford University's Center for Ethics in Society for helpful discussion, and to Mark Budolfson, Liam Shields, and two anonymous reviewers for helpful written comments.

If you need to cite this article, please use the following format:

Berkey, Brian, "Climate Change, Moral Intuitions, and Moral Demandingness," *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014), 157-189, edited by S. Maffettone, G. Pellegrino and M. Bocchiola