Does Climate Change Policy Depend Importantly on Population Ethics?

Deflationary Responses to the Challenges of Population Ethics for Public Policy

Gustaf Arrhenius, Mark Budolfson, and Dean Spears

14 August 2019

forthcoming in *Climate Change and Philosophy*, Oxford UP

> "To plan an appropriate response to climate change, it is important to evaluate each of the alternative responses that are available. How can we take into account changes in the world's population? Should society aim to promote the total of people's wellbeing in the world, or their average wellbeing, or something else? The answer to this question will make a great difference to the conclusions we reach." (Pachauri, Mayer, & Intergovernmental Panel on Climate Change (2015)).

1. Introduction

The International Panel on Climate Change (IPCC) and some leading philosophers and economists have expressed unease about the implications of population change for evaluating responses to climate change and other intergenerational policy challenges. Their unease derives from a common view among those who investigate the questions of population ethics, that is, theories about the value of outcomes where the number of people, the quality of their lives, and their identities may vary. The view is that we do not know what to do about intergenerational policy until we know what to do about population ethics. John Broome, in particular, has prominently voiced the concern that

climate policy could turn critically on unresolved questions in population ethics.[1] The worry expressed by Broome and reflected in the quote from the IPCC above might be stated as follows:

> **Worry:** Because climate change, climate policy, the size of the population, and population policy all may have effects on one another, and because population ethics is so theoretically unresolved as to permit a wide range of reasonable disagreement about social evaluation, our ignorance of the correct population ethic implies serious practical ignorance about what climate policies to pursue.[2]

In this chapter, we argue that the Worry is not obviously well-founded: we may already know enough to make good choices about climate policy even without further progress in population ethics, and further progress might not make much difference to the conclusions that are ultimately correct. More generally, we highlight some reasons – some philosophical, some empirical – why intergenerational policymaking might not be very sensitive to classic arguments from population ethics in the way that have often been assumed.

To understand why the IPCC and many others share the Worry, we must begin by noting that intergenerational policymaking seems to require a concept of goodness that aggregates consequences for many different people (perhaps even non-humans), with different properties, living at different times. Most of these people are not yet alive. Most of them will only ever be born depending on which particular climate policy is chosen. But any response to climate change requires integrating over the consequences for all of them.

For example, consider the Integrated Assessment Models (IAMs) of climate policy constructed by economists and other researchers. In 2018, William Nordhaus was awarded the Economics prize to the memory of Alfred Nobel, partly for his family of climate policy IAMs. IAMs like Nordhaus' choose an optimal carbon tax policy, balancing the disadvantages of more expensive energy with the advantages of reduced global warming. More broadly, reducing fossil fuel

---

[1] See, e.g., Broome (1992), (2004), ch. 1, and (2012b).

[2] "We do not know what value to set on changes in the world's population. If the population shrinks as a result of climate change, we do not know how to evaluate that change. Yet we have reason to think that changes in population may be one of the most morally significant effects of climate change. The small chance of catastrophe may be a major component in the expected value of harm caused by climate change, and the loss of population may be a major component of the badness of catastrophe. … So we face a particularly intractable problem of uncertainty, which prevents us from working out what we should do. Yet we have to act; climate change will not wait while we sort ourselves out" (Broome (2012a), pg. 183-185).

consumption could increase present-day economic costs for both poor people and rich people; could slow economic growth and poverty alleviation in the developing world; and could prevent future harm from temperature increases — increases which will help some people, but hurt many more people, and have consequences for inequality. The socially optimal carbon tax or fossil fuel policy depends on taking all of these and other relevant factors into proper account – which seems to require weighing the aggregate of these consequences conditional on different policy options.

So, choosing a policy response to climate change seems to demand an aggregative concept of goodness — an axiology. Those who study axiology have devoted considerable theoretical attention to population ethics: to the questions of how rankings of aggregate social goodness extend to ranking outcomes in which different people and different numbers of people exist. Parfit (1984) identified many of the core questions of population ethics, which are widely regarded to remain open. A number of candidate resolutions have been offered in the literature, but a formal literature involving impossibility theorems — led by Arrhenius (2000a), (2000b) and subsequent work — has demonstrated that each approach (and all possible approaches) has one or more seemingly counterintuitive implication. These theorems may appear to show that our considered moral beliefs are mutually inconsistent, that is, that necessarily at least one of our considered moral beliefs is false. Since consistency is, arguably, a necessary condition for moral justification, it may appear that we are forced to conclude that there is no moral theory which can be justified. Moreover, we would then lack the theoretical tools needed to evaluate climate options in which the number of people, the quality of their lives, and their identities will differ.

In Section 2 we introduce in more detail these paradoxes and the related population axiology literature, with special focus on Parfit's well-known Repugnant Conclusion. With this introduction in hand, Section 3 offers the first and simplest of two deflationary responses to the Worry: it may be, given the actual facts of climate change, that all axiologies agree on a particular policy response. In this case, there would be a clear dominance conclusion, and the puzzles of population ethics would be practically irrelevant (albeit still theoretically challenging). Section 4 offers the second more complex deflationary response: despite the impossibility results from Arrhenius, it is nonetheless possible to prove the possibility of axiologies that satisfy *bounded* versions of all of the desiderata from the population ethics literature that Arrhenius's proofs marshal. In this way, an incomplete population axiology that is defined over the practically relevant bounded space can avoid the Repugnant Conclusion and satisfy other relevant bounded versions of the adequacy conditions in population ethics. Assuming that we only need to consider the bounded versions of the adequacy

conditions when we consider policy issues, and that analogous impossibility theorems cannot be proved in the bounded domain, we can for practical purposes put the impossibility theorems that have haunted population ethics to the side.

These deflationary responses do not show that theoretical progress towards population axiology should not continue. Indeed, as we shall show below, an important consequence of the second deflationary response is that it shows the need of more scrutiny of what the core intuitions behind the adequacy conditions in population ethics really are, and further investigation of axiologies on bounded domains. The upshot is that responding to climate change, and policy analysis more generally, may not need to wait for greater consensus in population ethics on unbounded domains, and that the possibility of deflationary responses to the impossibility theorems deserves further attention.

2. Population axiology and the Repugnant Conclusion

Population axiology concerns how to evaluate populations of different sizes in regard to their goodness: how to assign a value to increases and decreases in population size. The first few papers in this field were not published until the late 1960s and it did not become a significant field until Derek Parfit's famous book *Reasons and Persons*, published in 1984. It is now a very lively field of inquiry.

As John Broome has noted, policymakers seem to almost universally ignore the effects of policy on population size. Why do they ignore it? One possible explanation is that many people have what Broome calls the **Intuition of Neutrality**, which holds that adding a person to the world's population makes the world neither better nor worse.[3] Hence, effects on population size is something that we do not need to think about, or if we do need to think about it, it is because it makes people's lives better or worse; other than that, having a bigger or smaller population does not make any difference to the value of outcomes.

There are likely to be limits to Neutrality. For example, most people would probably agree that if population growth leads to having many people with very bad lives, then that would make the world worse. In light of this, *pace* Broome, we think it is likely that people endorse not the Intuition of Neutrality, but rather the more limited **Asymmetry Intuition** (which also appeared earlier in the literature): We have no moral reasons for or against creating people with positive welfare stemming from the welfare these people would enjoy, but, on the other hand, we have moral reasons against

[3] For a more detailed discussion of the neutrality intuition, see Broome (2004), (2010).

creating people with negative welfare stemming from the negative welfare these people would suffer. Hence, we are neutral about adding people with positive welfare.[4] However, assuming that future people have positive or neutral welfare, the idea is that population size is neutral in terms of value and that we can ignore this aspect when considering different policies.

However, Neutrality and Asymmetry each on their own lead to inconsistency given some other beliefs that most of us share. Consider the following two possible additions to the present population *A*, each of which would be the result of an alternative climate policy:

- Population *B* consists of a number of people with very low positive welfare, and
- Population *C* is a population of the same size as *B* but made up of people with very high welfare.

According to Neutrality and Asymmetry, either adding *B* or adding *C* to *A* each would make the resulting populations equally good, given full comparability.[5] But surely, when other things are equal, it must be better to create people with very high welfare rather than people with very low welfare. Hence, population *A+C* is better than population *A+B*, which contradicts Neutrality and Asymmetry. So they are false. And because they are false, climate policy-making must value population size by aggregating welfare to measure how good and bad outcomes are.

The opening quotation from the IPCC listed two alternative approaches to aggregating welfare. One approach is Total Utilitarianism: when we evaluate future populations in respect of population change, we look at the total welfare in the different possible outcomes and rank them by how much total welfare they contain. According to this view, we should maximize the total amount of welfare in the world. So if there are more people with lives worth living, then that is better.

Now a problem with this view is that it has a number of very counterintuitive implications. Much theoretical attention in population ethics has focused on a particular implication of Total Utilitarianism. Total welfare can be increased in two ways when the size of the population is no longer fixed: by keeping the population at a constant size and making people's lives better, or by increasing the size of the population by adding new people with lives worth living. So, according to

---

[4] This formulation is from Arrhenius (forthcoming), (2000b). For earlier formulations, see McMahan (1981); Parfit (1982).

[5] Giving up full comparability isn't sufficient to save the neutrality and asymmetry intuition, see Arrhenius (forthcoming) and Broome (2004).

Total Utilitarianism, a future with an enormous population with lives barely worth living could be better than a future with a smaller population with very high individual quality of life. But the idea that it would be better to radically increase the world's population at the expense of future people's individual welfare seems repugnant to many, and rather a reason to reject Total Utilitarianism. It is an instance of Parfit's infamous Repugnant Conclusion:

> **Repugnant Conclusion:** For any population consisting of people with very high positive welfare, there is a better population in which everyone has a very low positive welfare, other things being equal.[6]

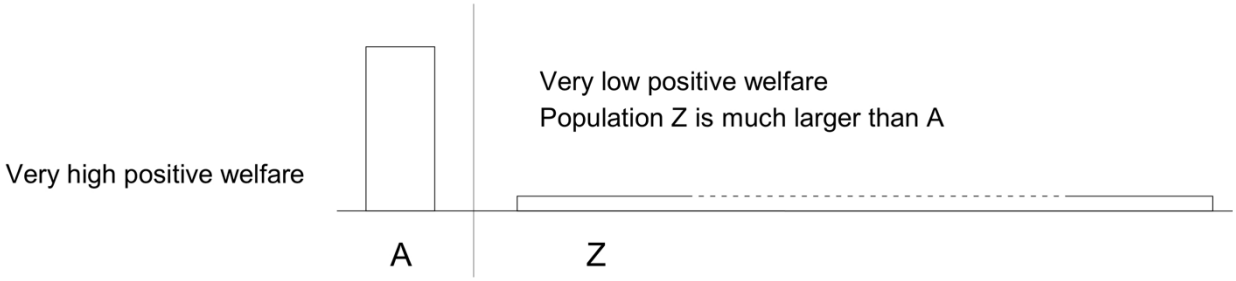

Figure 1: Repugnant Conclusion

In Figure 1, the width of each block represents the number of people; the height represents their lifetime welfare. Dashes indicate that the block in question should be much wider than shown, that is, the population size is much larger than shown. These populations could consist of all the past, present and future lives, or all the present and future lives, or all the lives during some shorter time span in the future such as the next generation, or all the lives that are causally affected by, or consequences of a certain action or series of actions, and so forth.

---

[6] Here's how Parfit (1984), p. 388 formulates the conclusion: "For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living." Hence, our formulation from Arrhenius (2000b) is more general than his. The *ceteris paribus* clause in the formulation is meant to imply that the compared populations are roughly equal in all other putatively axiologically relevant aspect apart from individual welfare levels. Although it is through Parfit's writings that this implication of Total Utilitarianism has become widely discussed, it was already noted by Henry Sidgwick (1907), p. 415, before the turn of the century. For other early sources of the Repugnant Conclusion, see Broad (1979), pp. 249–250, McTaggart (1927), pp. 452–453, and Narveson (1967).

All the lives in the diagram have positive welfare, or, as we also could put it, all the people have lives worth living. The A-people have very high welfare whereas the Z-people have very low positive welfare. The reason for this could be that in the Z-lives there are, to paraphrase Parfit, only enough ecstasies to just outweigh the agonies, or that the good things in those lives are of uniformly poor quality, *e.g.*, eating potatoes and listening to Muzak.[7] Or it could be that the Z-people have quite short lives as compared to the A-people. We could imagine that in A, the people live for, say, 80 years whereas in Z the average life expectancy is, say, 35 years, like in some developing countries in the 1970s. However, because there are many more people in Z, the total sum of welfare in Z is greater than in A. Hence, a theory like Total Utilitarianism, according to which we should maximize the welfare in the world, ranks Z as better than A --- an instance of the Repugnant Conclusion.

As the name indicates, many people find the Repugnant Conclusion a reason to reject Total Utilitarianism; to these, the idea that we can make the world better by expanding the population at the expense of future people's individual quality of life seems very counterintuitive. The Repugnant Conclusion has sometimes been taken in the literature as the major objection to Total Utilitarianism that allegedly disqualifies it as a plausible axiology.[8]

The other approach mentioned by the IPCC is to maximize *average* welfare in the world. This is what Average Utilitarianism tells us to do. Returning to Figure 1, in the case of the A and Z populations the average principle recommends A, because average welfare is much higher in A than in Z. Hence, Average Utilitarianism avoids Parfit's Repugnant Conclusion, which may seem to count in its favour.[9] Unfortunately, it has even worse problems. One problem with maximizing average welfare is that it implies that it can be better to add one group of people to the population rather than some other group, even if each person in the former group has a life that is not worth living and each person in the latter group has a life that is worth living. This is illustrated in Figure 2:

---

[7] See Parfit (1984), p. 388 and Parfit (1986), p. 148.

[8] There are other implications of Total Utilitarianism in population ethics that arguably are even more counterintuitive than the Repugnant Conclusion, see e.g., Arrhenius (forthcoming), (2000b), (2011). More on this below.

[9] As explained below, Budolfson & Spears (2018c) have argued that Parfit's initial illustration is only a subset of the classical Repugnant Conclusion, and that we should understand it to include a version (based on addition to a base population, explained in their paper) that is implied by Average Utilitarianism and other axiologies that are commonly taken to avoid the repugnant conclusion. Throughout this section, for clarity we maintain the standard terminology in the population literature, except where it is clear we are discussing the argument of Budolfson and Spears. Anglin (1977) and Arrhenius (2000b), ch. 3, 10 note that Average Utilitarianism implies a version of the Repugnant Conclusion to the effect that that for any population with very high welfare, it can be worse to add this population rather than a population with very low welfare. As Anglin summarized simply: "in some cases the average principle also leads to the Repugnant Conclusion" (p. 746).

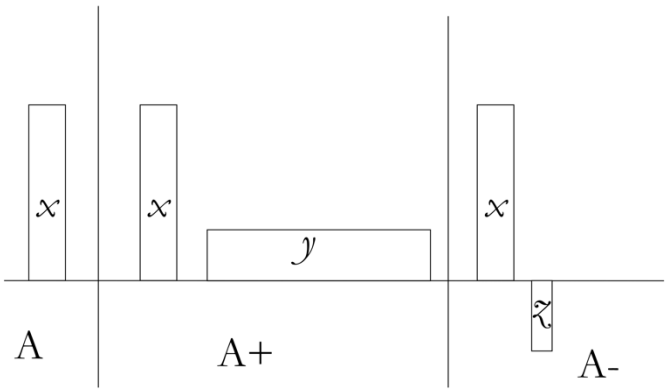

Figure 2: The Sadistic Conclusion

Here, we have the A population where the x-people's quality of life is very high. Assume that we can either increase population either by adding the y-people that have quite low but positive welfare—their lives are worth living—or by adding the z people, all of whom are suffering horribly—their lives are not worth living.

Because adding a lot of people with very low but positive welfare can decrease the average welfare of the population more than adding fewer people suffering horribly, it might be better, according to Average Utilitarianism, to add the suffering lives (the z-people) rather than the lives worth living (the y-people). Again, we have a very counterintuitive conclusion on our hands. This is what Arrhenius called the Sadistic Conclusion:

> **Sadistic Conclusion:** It can be better to expand the population by adding people with negative welfare rather than adding people with positive welfare, other things being equal.[10]

The path away from the Repugnant Conclusion towards the Sadistic Conclusion illustrates the puzzles that motivate the Worry. There may be no principle for evaluating populations that is not in some way very counterintuitive. This possibility was originally raised informally by Parfit, who presented a number of paradoxes in population ethics. Much of the important theoretical progress since then has been in formalization of these conclusions and axiologies, as well as many others, and their integration into rigorous proofs.

---

[10] See e.g., Arrhenius (2000b), (2000a).

This literature has progressed, at first, through a dialogue in which researchers proposed and formalized alternative population axiologies (Greaves (2017)). Each was specially formulated to avoid versions of the Repugnant Conclusion, and then further explored by researchers. So, Ng (1989) introduced a variable-value axiology, in which the average utility of a population is inflated by a positively increasing, concave function of population size, such that social evaluation asymptotes from nearly-Total Utilitarianism to nearly-Average Utilitarianism as population size increases. Like Average Utilitarianism, Ng's theory does not escape the Sadistic Conclusion. Blackorby & Donaldson (1984) and later Blackorby, Bossert, & Donaldson (1995) propose Critical-Level Generalized Utilitarianism; this approach also avoids the Repugnant Conclusion at the cost of implying the Sadistic Conclusion. Other approaches, such as Sider (1991)'s theoretical example of Geometrism, or Asheim & Zuber (2014)'s Rank-Dependent Generalized Utilitarianism, attend to people's *rank* within a population, like maximin does. These avoid the Repugnant Conclusion, but have other implausible properties, including in cases where population size does not change, such as recommending redistribution from the worst off to the best off in some cases.[11]

None of these proposals has resolved the paradoxes. Led by Arrhenius (2000b), the literature has now established a number of impossibility theorems that demonstrate that no axiology can simultaneously satisfy various sets of very compelling adequacy conditions or principles. Trying to satisfy all of them at the same time leads to contradiction. These conditions are of the type that we have been considering—for example, what Arrhenius calls the Egalitarian Dominance Condition, which states that one population A is better than another same-sized population B if A is perfectly equal and every person in A is better off than every person in B. This condition is incompatible with several other compelling conditions, including conditions that are formulated to rule out the Repugnant and the Sadistic Conclusions. The first and perhaps most well-known of these impossibility theorems is the following:

> **Impossibility Theorem** (Arrhenius (2000a)): There is no welfarist axiology that satisfies the Dominance, the Addition, and the Minimal Non-Extreme Priority Principle and avoids the Repugnant, the Sadistic and the Anti-Egalitarian Conclusion.[12]

---

[11] See Arrhenius (forthcoming), (2000a); Arrhenius, Ryberg, & Tännsjö (2014).

[12] For theorems with logically weaker and intuitively even more compelling conditions, see Arrhenius (forthcoming), (2000a), (2001), (2011).

Although we refer the reader to the formal statement by Arrhenius (2000a), we emphasize here that each of the conditions listed in the theorem is intuitively compelling. For example, the Dominance Condition is simply that if everyone in population A is better off than everyone in population B, then A is better than B. Moreover, as Arrhenius has shown, there are theorems with logically weaker and intuitively even more compelling conditions.[13]

Impossibilities such as these are the challenges that motivate the Worry. One type of response to this challenge that we will set aside here is to offer a purported philosophical *resolution* to the challenge of the Repugnant Conclusion. Most of these purported resolutions argue that the Repugnant Conclusion should simply be accepted as true. For example, Hare (1988); Huemer (2008); Mackie (1985); Tännsjö (2002), and Gustafsson (forthcoming) have all offered arguments in favour of endorsing the Repugnant Conclusion, because of various arguments that the apparent repugnance of the conclusion is illusory or based on misunderstanding. One drawback with this resolution is that the theorems with logically weaker conditions are not based on avoidance of the Repugnant Conclusion but on the intuitively more compelling **Very Repugnant Conclusion**: For any perfectly equal population with very high positive welfare, and for any number of lives with very negative welfare, there is a population consisting of the lives with negative welfare and lives with very low positive welfare which is better than the high welfare population, other things being equal.[14]

More recently, Budolfson & Spears (2018c) have offered an alternative type of resolution of the Repugnant Conclusion. They argue that Parfit's original example of the Repugnant Conclusion should be understood as describing only a proper subset of instances of the Repugnant Conclusion, and that the full set of instances of the Repugnant Conclusion should be understood to include a broader set, including cases in which there is a base population that is unaffected by the choice between a larger or a smaller population.[15] Given their more general characterization of the Repugnant Conclusion, they prove that all of the most commonly discussed aggregative welfarist

---

[13] See, e.g., Arrhenius (forthcoming), (2000a), (2001), (2011).

[14] See, e.g., Arrhenius (forthcoming), (2000b), (2011). For a detailed discussion of other problems with debunking arguments with regard to the Repugnant Conclusion, including Hare et al.'s arguments, see Arrhenius (forthcoming), ch. 3, (2000b).

[15] Budolfson and Spears' concept of avoiding all instances of the repugnant conclusion including those with non-zero base populations is comparable to Arrhenius' Strong Quality Addition Principle (Arrhenius (forthcoming), (2000b)), which is violated by both Total and Average Utilitarianism (and some other population axiologies). Arrhenius draws, however, a different conclusion from this result, namely that the Strong Quality Addition Principle should be rejected as an adequacy condition since it rules out too many axiologies in one fell swoop and thus is in that sense too strong.

population axiologies imply at least one instance of it. They then argue that since the Repugnant Conclusion so understood is a problem for all of the most commonly discussed welfarist axiologies, it can no longer be reasonable to assume that a plausible axiology must avoid it.

We set aside these purported solutions in this paper. The problem we focus on is what the upshot of the population ethics literature is for policy on the assumption that there is no resolution to the challenges of population axiology at hand.

3. First Deflationary Response: Axiologies May Agree about Climate Change

The open theoretical questions of population axiology only turn out to be a practical problem for a policy challenge if population axiologies sufficiently disagree about the best policy response to that challenge. To see how this could turn out not to be the case in connection with climate change, consider the toy illustrative example in Figure 3.

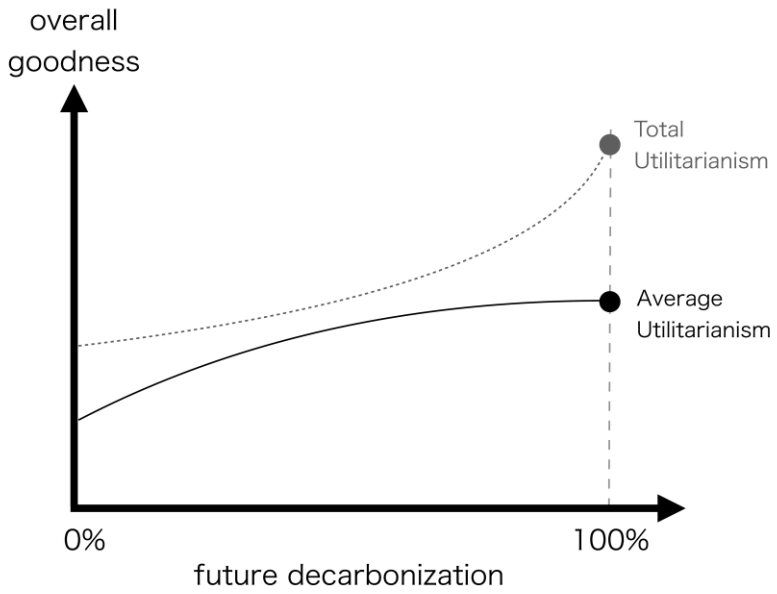

Figure 3: Two population axiologies recommend the same "corner solution" to optimal decarbonization

If figure 3 correctly described the full climate policy problem, then the Worry could be false, even though the candidate population axiologies differ. In the figure, the ethical question under consideration is what future decarbonization rate should be achieved: 100%, 0%, or some other optimum in between? The recommendations of two population axiologies are considered. These

give different evaluations of different options. Total Utilitarianism rises convexly as the decarbonization rate increases; Average Utilitarianism rises only concavely. Thus, Average Utilitarianism thinks that a decarbonization rate of 90% would be only slightly worse than 100%, but Total Utilitarianism thinks 90% would be much worse than 100%.

Note that Average and Total Utilitarianism even have different *scales* for goodness: neither their lowest level of goodness nor their highest levels of goodness are the same number, and their evaluations cover ranges of different length. This is important because some responses to normative uncertainty — such as Expected Moral Value — recommend an average or expectation over alternative theories (Budolfson & Spears (2018a); Bykvist (2017); Bykvist, MacAskill, & Ord (2019); Greaves & Ord (2017); Hedden (2016)). This moral-expectation approach has found difficulty in the need to compare evaluation quantities across theories, but that problem is not relevant in the case of Figure 3, because the two axiologies agree on the optimum.

The point of Figure 3 is that both Average and Total Utilitarianism recommend the same *corner solution*. In optimization, a "corner solution" is when the optimal policy is equal to a boundary constraint. Because Average and Total Utilitarianism both recommend full decarbonization, in this example, there is no *practical* disagreement between them, only *theoretical* disagreement. Whether or not actual climate policy is well-described by figure 1 is substantially an empirical question (concerning economics, demography, climate science, etc.), although also a normative one (because different losses, such as of life and wealth, must be aggregated). However, it is not implausible that actual climate policy questions could be resolved by dominance — that is to say, by agreement across candidate axiologies. For example, if we are confident that a particular set of future lives would be full only of terrible suffering and thus not worth living, and if by preventing those lives from occurring we prevent some harmful carbon emissions, and if furthermore we know these are the only relevant considerations, then all plausible population axiologies recommend not creating those lives.

Although that example was fanciful, another might be quite realistic (see Scovronick et al. (2017) for detailed evaluation of the following). Consider investments in human development in developing countries, with a special focus on women's social status and the education and well-being of girls. This would have a range of likely consequences, which we can assume for hypothesis that we know with certainty (which would be confidence beyond the actual reach of social science):

- The women who receive the program and the lives lived by other people in their places and times would be better: an increase in the near-term average.
- Long-term average well-being would be improved by reduced climate change and by accelerated economic development.
- Some 21st century lives that would have been worth living would not be lived, because of empowered young women choosing to reduce their fertility. (Under Total Utilitarian-like theories, this would be a social cost.)
- Because of the reduced threat of climate change, the expected number of future good lives lived increases by more than the number of 21st century lives reduced.

In this case, the total expected number of lives lived would increase, average well-being would increase within every time period, and average across-time well-being would increase because the average human would live later in historical time. Moreover, it is not implausible that the welfare of the worst-off lives would be higher (a property that matters to some egalitarian views), although this was not specified above. So, according to every plausible axiology in the literature and more — including Average utilitarianism and related views, Total Utilitarianism and related views, maximin, and others — implementing the human development policy is recommended, in expectation. The upshot is that we can know whether to implement the policy without knowing the correct population axiology, and also without a general solution to moral uncertainty. In this case, the Worry would be deflated.

More generally, other practical policy questions that are commonly taken to hinge on the choice of population axiology may be resolved by similar dominance arguments or corner solutions.[16] This would depend on social, economic, and scientific facts. For example, some have argued that an implication of Total Utilitarianism is that substantially more resources should be invested in preventing human extinction (Beckstead (2013); Bostrom (2013)). However, it may be that commonly-discussed policy options (such as asteroid deflection) offer a small marginal benefit

[16] One exception to this possibility is the welfare of non-human animals. The number and well-being of nonhuman animals is generally governed by ecological forces such as natural selection, to a greater extent than the number and well-being of humans, which is regulated, in part, through complex technology and culture. In many cases, the implication of this fact may be that the average well-being of non-human animal species is kept within a narrow species-specific range, while adjustment to changing conditions occurs in population size (on the extensive rather than the intensive margin, in economists' language). If so, Average and Total Utilitarianism, as extended to non-human animals, may give very different recommendations. See Hsiung & Sunstein (2006), and Budolfson & Spears (2018b) for more on climate and non-human animals.

of further investment as compared to merely pursuing standard economic growth, technological progress, and human development. The reason being that such standard policies would have large *co-benefits* against existential risk, perhaps because war of mass destruction or resistant, pandemic infectious disease would be less likely, or because survival-promoting technologies would be invented. If so, both Average and Total Utilitarianism would recommend serious investment in thoughtful, long-term human development, economic growth, and technical progress: Average Utilitarianism because it increases average well-being, and Total Utilitarianism because it does this while also offering the co-benefit of promoting survival. To be sure, this would not be the set of policies that humanity is currently pursuing, but it would not be a major reallocation into activities that only have the benefit of reducing existential risk, and nor would it turn on the choice of population axiology.

Of course, it may be that the climate policy menu under consideration does not yield one dominating option. Also, there could be additional considerations, such as bounded political capital. If political capital is scarce, a politician who needs to compromise across politically linked issues (such as climate policy and domestic health care or tax policy) may care about *how much worse* 95% would be than 100%, which cannot be settled by this sort of dominance-identification procedure. Still, this is a promising avenue for further research that should be pursued in light of the impossibility theorems in population axiology.

4. Second Deflationary Response: Bounded Population Principles

The Repugnant Conclusion --- and especially the search for a sensible population-sensitive social welfare function that does not imply the Repugnant Conclusion --- has been a central focus of the population ethics literature since Parfit (1984) introduced it. For example, Arrhenius, Ryberg, & Tannsjö (2014) has called it "one of the cardinal challenges of modern ethics" and Greaves (2017) introduces the Repugnant Conclusion as "the key objection" to Total Utilitarianism and related views. Because most of the literature on population axiology takes it as an adequacy condition that an acceptable social welfare function should not imply the Repugnant Conclusion, researchers have proven that many social welfare functions, in addition to total utilitarianism, imply the Repugnant Conclusion if the populations being evaluated can be unboundedly large. As noted above, Arrhenius (2000a), (2000b) presents an impossibility theorem that proves that no social welfare function can escape implying the Repugnant Conclusion, if the function is defined for unboundedly large

population and has desirable --- and plausibly ethically necessary --- properties. Such properties are formalized as axioms for Arrhenius' theorems.

These are impressive and rigorous philosophical results. But what are the implications for policy analysis? Do these results show that the assumptions of many leading policy analyses are illegitimate, as suggested by the quotes above from IPSP and John Broome? More generally, how should policy analysis respond to these results? Arrhenius notes that one response could be a thoroughgoing scepticism or paralysis. However, he is much more enthusiastic about the possibility of a deflationary response: namely, to "try to find a way to explain away the relevance of the [Repugnant Conclusion and associated impossibility] theorem for moral justification."[17]

Our goal in this section is to articulate another deflationary response to the impossibility theorems to the effect that policy analysis can in some cases legitimately ignore them and the Repugnant Conclusion when that analysis applies to bounded problems, as Arrhenius's impossibility theorems assume unboundedness. We show that unlike unbounded cases, in bounded cases that are relevant to policy analysis, it is indeed possible to identify an axiology that captures all of the intuitions that support Total Utilitarianism while also avoiding the Repugnant Conclusion. This shows that it may be possible to endorse both the intuitions that motivate Total Utilitarianism and the intuition that tells against accepting the Repugnant Conclusion. The idea is that there might be a mere *appearance* of conflict between these intuitions that arises from taking our intuitions about the realistic range of cases relevant to policy as also extending to cases in the unbounded penumbra.

In other words, this second deflationary response to the Worry exploits the possibility of interpreting the intuitively compelling axioms of population ethics as restricted to a bounded domain.[18] An adequacy condition to avoid the Repugnant Conclusion on unbounded space has no implications for such a family of bounded axiologies. As we detail below, in our formal argument, our approach is not to reject that populations can be unboundedly large; instead, we propose bounded *axioms* that, in some cases, apply to only some of the space of possible populations.

---

[17] Arrhenius (forthcoming), ch. 13, (2000b), ch. 12.

[18] Shiell (2008) offers a formal proof of an intuition (related to a point made by Parfit (1984), pg. 387), namely that within a truncated domain, Total Utilitarianism need not imply the Repugnant Conclusion within that domain. In this way, Shiell's proof depends essentially on truncating the choice set. In contrast, our proof below does not truncate the choice set. Our axiological principles cover the entire choice set, fully specify how to rank all outcomes within a policy-relevant range, but do not fully specify how rank all outcomes beyond that range. Moreover, the principles also satisfy certain bounded analogues of the central population ethics desiderata involved in the impossibility theorems in the area.

4.1 Axiology with population size bounds

The practically relevant set of policy options that humanity will ever face is a bounded set, along many dimensions. This is partly because the set of practically relevant population sizes is bounded. This is true even if the possible values of social welfare are unbounded, in part because policy choices could only have boundedly large effects on individual welfare. In making the empirical observation that the set of practically relevant population sizes is bounded, we have in mind a very large upper bound. The upper bound could be much larger than the largest set that an expert predicts could ever be relevant. It is sufficient for our purposes, for example, that the bound be 10^{80}, which is an estimate of the number of atoms in the universe, or 10^{58}, which is the estimate of Bostrom (2013) of the number of simulated human lives that a superintelligence could create with the available energy in the universe. The lower bound on the policy relevant set of population sizes is the number of humans who already have ever been born.

In this vein, the previous section noted that, even outside of population ethics, practical policy analyses are untroubled by imaginable, *unbounded* marginal utilities or counts of small harms; in this section, we formalize that observation by weakening some axioms of population ethics to a bounded domain. We can consider axioms that only apply to a very large but bounded subset of the potentially unbounded complete, imaginable social choice set, and choose a family of axiologies that (a) satisfies attractive axioms defined over the bounded set and (b) has no implications about the Repugnant Conclusion. A requirement to avoid the Repugnant Conclusion has no implications for this bounded family of axiologies.

The purpose of axiomatic representation theorems is to rule in and rule out sets of functional forms. In general, a representation theorem permits a *family* of function shapes that leaves certain features unspecified. For an example in the context of axiologies, critical level generalized utilitarianism is consistent with concave or affine transformations of utility and with positive or zero critical levels; each of these combinations would have different normative implications. Similarly, a family of population-sensitive axiologies could leave unspecified how populations are evaluated outside of the bounded set. Such a family of axiologies would ignore the Repugnant Conclusion --- while fully specifying the social evaluation on the bounded set.

The literature has identified the following very general characterization of the space of a number of important aggregative (here meaning non-person-affecting), welfarist axiologies:

$$W = g(n)[h(n^{-1} \sum_i f(x_i)) - h(f(a))],^{23}$$

where:

- n is population size,
- x_i is the utility of person i,
- a is 0 or positive and is a critical level for adding a life to be a social improvement.
- The functions f, g, and h are all non-decreasing. If f and h are both the identity function, then we have utilitarianism. If f is concave and h is the identity function, then we have additively separable prioritarianism. If f is concave and $h = f^1$, we have a type of non-separable egalitarianism.

This general functional form is intended to clarify that the shape of g could be chosen independently of any combination of otherwise permissible features for the other elements of the function. It includes as special cases many axiologies in the literature, although not rank-dependent axiologies such as maximin or Zuber and Asheim's (2014) rank-dependent generalized utilitarianism. In Total Utilitarianism g is linear; in Average Utilitarianism g is constant; and in Ng's Theory X' g is concave.

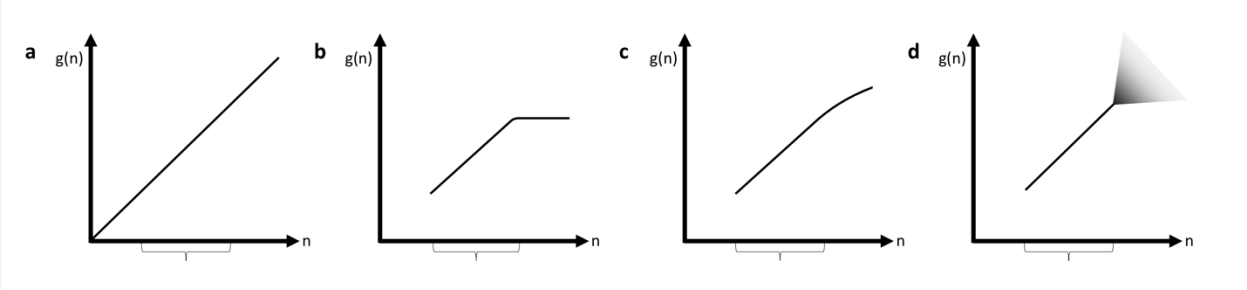

Figure 4. Families of social evaluations that cohere with totalist axioms on the bounded set

Note: Curly braces on the horizontal axis note the finite bounded set

23 Compare Budolfson & Spears (2018c) and Greaves & Ord (2017).

Figure 4 illustrates a possibility for *g* that is the focus of this section of the paper: a family of functional forms for g could be chosen that *fully specifies g* on the bounded policy-relevant set, while avoiding the Repugnant Conclusion and *taking no stand* on the shape of *g* outside the bounded set. Functional forms **a**, **b**, **c**, and **d** would rank policy options over the practically relevant set identically, for any given specification of *f*, *h*, and *a*. Form **a** matches Total Utilitarianism, if *f* and *h* are the identity function. Forms **b**, **c**, and **d** are blank at populations smaller that the bounded choice set, to illustrate that they do not make assumptions about how to rank populations this small. It is not essential to our argument that the bounded set have either a zero or a positive lower bound: the possibility of a lower bound greater than zero represents the minimum on policy-relevant population sizes due to the fact that billions of humans have already been born.

Forms **a**, **b**, and **c** have different implications for the Repugnant Conclusion, and may or may not invoke other undesirable properties outside of the practically relevant set. Form **d** is not a fully specified function form, but is merely a representation of the possibility of a decision-maker remaining uncertain about options outside of the bounded set. The existence of functional forms **a**, **b**, and **c** and of the options in **d** tells us that a climate policy-maker could say:

> Because over the practically relevant set of policy options I am both attracted to totalist intuitions (or axioms), and I am fully comfortable with a generalized total social welfare function; and because this practically relevant set is bounded, I should make policy according to any of **a**, **b**, **c**, or **d**. I remain troubled by the Repugnant Conclusion and only by the Repugnant Conclusion when I concentrate on it, but that can be a problem for future research, because it does not threaten my conviction about how policy options should be ordered in the practically relevant set of policy options.

Of course, someone with less totalist intuitions, for example someone who leans more toward Average Utilitarianism, wouldn't be able to say this. Likewise for theories that do not fall under the general characterisation above, such as rank-order theories and person affecting theories. Still, it shows that restricting the applicability of the axioms to bounded sets opens up for convergence on policy recommendations for a number of different theories.

4.2 Possibility Proof for Escaping the Repugnant Conclusion while Satisfying Bounded versions of Population Ethics Desiderata

With the preceding in hand, we now prove that there is a principled motivation to avoid the Repugnant Conclusion by adopting a family of g functions that do so – i.e. by adopting a set of bounded axiologies. Moreover, the motivation is in the axiomatic spirit of Arrhenius' theorems: rather than beginning from a functional form, the motivation for adopting a family of g functions is given by slightly modified versions of some of the traditional axioms of population ethics such that they make no claim beyond the large bounded set of relevant options.

For example, in one of his pioneering informal results Parfit (1984), make use of the controversial (since it makes it easy to derive the Repugnant Conclusion) Mere Addition Principle:

> **Mere Addition:** An addition of people with positive welfare does not make a population worse, other things being equal.[24]

This axiom could be weakened to:

> **Bounded Mere Addition:** An addition of people with positive welfare does not make a population worse, other things being equal, if each population (with and without the addition) is within the bounded domain.

One could similarly modify Arrhenius' Non-Sadism Condition to a Bounded Non-Sadism Condition, and the Egalitarian Dominance Condition to a Bounded Egalitarian Dominance Condition. In each case, the modified axiom would reflect an analogous axiological intuition as the original axiom, but with the restriction that it only applies to comparisons of populations within the bounded set. Such bounded axioms would simply make no claims about ranking populations outside of the bounded set. Relatedly, but outside of an axiomatic framework, one could assess the constructive argument that Broome (2004) presents for generalized, Critical-Level Total

[24] See also Blackorby, Bossert, & Donaldson (2005), Arrhenius (forthcoming), (2000b). Like many contributors to the debate, Arrhenius and Blackorby et al. rejects the Mere Addition Principle as an adequacy condition for a satisfactory population axiology.

Utilitarianism, but --- unlike Broome --- only assess and apply the argument while considering populations within the bounded set.[25]

Would such bounded axioms be intuitively compelling? Because they are logically weaker than their unbounded counterparts, they must be at least as compelling. The impossibilities of population ethics are only interesting because the original axioms are compelling. Anyone who agrees with the original axioms will also agree with these, which are weaker: they make the same claims about fewer cases. And they may attract the new support of cautious evaluators who are hesitant to make axiomatic claims about unbounded populations.

In particular, consider a social evaluator who accepts the axiom of a complete and transitive social order for all populations, and accepts anonymity and same-number Pareto for all populations, but then accepts only the Bounded Mere Addition and similarly modified and bounded versions of Separability and the other axioms that Blackorby & Donaldson (1984) demonstrate entail generalized Critical Level Total Utilitarianism. Such a set of axioms would entail a family of social welfare functions – each same-number utilitarian – where g is increasing and linear over the bounded set, and could have any shape outside of the bounded set (perhaps disciplined by further continuity axioms). In particular, the resulting axiologies need not be separable outside of the bounded set. Such bounded axioms would also rule out a positive critical level within the bounded set, due to Bounded Mere Addition. The modified axioms would provide a principled motivation for the social evaluator to use this family of social welfare functions. Such an axiology would be sufficient for a climate IAM and to answer any question posed by climate ethics, and the Repugnant Conclusion is not entailed.

More broadly, we now prove:

> **Possibility Theorem for Bounded Axiologies:** There exist complete welfarist axiologies that satisfy the Bounded Dominance, the Bounded Addition, and the Bounded Minimal Non-Extreme Priority Principles and avoid the Repugnant, the Bounded Sadistic, and the Bounded Anti-Egalitarian Conclusion.

---

[25] Of course, a more substantive axiology such as Critical-Level Total Utilitarianism could still have unintuitive violations of other bounded conditions; for example, Critical-Level Total Utilitarianism violates a Bounded Non-Sadism that modifies the Non-Sadism axiom to only apply to the bounded set.

The proof is by example. Forms **b** and **c** from Figure 4 satisfy the theorem, as does any form of W in which h and f are the identity functions, g is the identity function on the bounded set (as in Total Utilitarianism), and g is everywhere non-decreasing and is bounded above outside the bounded set. At very large population sizes outside of the bounded set, this family of axiologies would imply the (unbounded) Sadistic Conclusion, just as Ng's Theory X' does – but that is no contradiction, because the Possibility Theorem only requires avoiding the Sadistic Conclusion in the bounded set. Note that bounded Average Utilitarianism (g is constant in the bounded set) is not an example consistent with the Possibility Theorem because it does not satisfy avoiding even the Bounded Sadistic Conclusion; nor does Theory X', if g is concave within the bounded set.

A worry, however, is that the impossibility theorems might reappear over a bounded domain by further reformulating the adequacy conditions to take into account that we are now dealing with a bounded domain. Such reformulations can be done in multiple ways, one straightforward example is as follows:

> **Bounded Repugnant Conclusion I:** In the bounded domain, for any population consisting of people with very high positive welfare, there is a better population in which everyone has a very low positive welfare, other things being equal.

Rather trivially, this cannot be an implication of axiologies that verify the Possibility Theorem above. Consider, for example, the largest population size within the bounded domain, and assume each member of that population has a very high welfare. Because this involves the largest population size within the domain, there cannot be a population with much lower welfare that is better.

However, there are other reformulations of the Repugnant Conclusion that are not as easily avoided in the bounded domain. Here is one example:

> **Bounded Repugnant Conclusion II:** In the bounded domain, there are very large populations consisting of people each with very high positive welfare for which there are better populations in which everyone has a very low positive welfare, other things being equal.

The idea behind the Bounded Repugnant Conclusion II is the intuition that if a population is sufficiently big and everyone enjoys very high welfare, then such a population is better than each of

the populations with only very low positive welfare in the domain. In light of this, it could be argued that what is fundamental to repugnance is merely the existence of a **Large Quantity-Quality Tradeoff**. This is surely one candidate for being the main intuition behind the counterintuitiveness of the Repugnant Conclusion. According to this take on the Repugnant Conclusion, unboundedness is not essential to repugnance. This raises the question of what is essential to the Repugnant Conclusion, and how many versions or instances there may be. As it is sometimes expressed, there can be various instances of the Repugnant Conclusion (Parfit (2016)). If so, perhaps any satisfactory population axiology should not imply any instances of it.

Depending on the size of the domain, the size of the very large populations, and on what the difference is between lives with very high and very low welfare, Bounded Total Utilitarianism might imply Bounded Repugnant Conclusion II. For example, let's assume that a life with very high welfare is at least 100 times better than a life with very low positive welfare and let's use Bostrom's estimate, mentioned above, of 10^{58} simulated human lives as an upper bound on the size of possible populations. It follows from Bounded Total Utilitarianism that there is a very high welfare level such that for any population up to size 10^{56} enjoying this level, there is a better very low welfare population in the domain. So, according to Bounded Total Utilitarianism, a population with lives barely worth living would be better than an enormous population with very high individual quality of life. And given that an intuitively sufficiently large population with very high welfare is smaller than 10^{56}, which seems intuitively compelling (compare Parfit's specification of "at least 10 billion people"), Bounded Total Utilitarianism implies the Bounded Repugnant Conclusion II in this domain.

One can, of course, argue for other smaller upper bounds on the size of possible populations and for other differences between very high and very low positive welfare lives. However, what this shows is that the unbounded scope of the classical Repugnant Conclusion is not needed to produce extreme quantity-quality trade-offs More importantly, it shows that there may be impossibility theorems looming even in the bounded domain with the adequacy conditions from the unrestricted domain appropriately adjusted. Of course, this has to be appropriately shown by proving such theorems.

The mere fact that some set of axioms is impossible to combine is not sufficient, of course, for an important challenge to climate policy-making. The involved conditions also have to be intuitively compelling. As the example above hints at, these conditions might or might not be sufficiently compelling depending on what one takes to be the main intuition behind classical

unbounded conditions. Hence, the results we get when restricting population ethics to a bounded domain raises new and important questions that need to be further investigated: Is the implication of Bounded Repugnant Conclusion II sufficiently counterintuitive to work as an adequacy condition for a satisfactory population ethics? Might it even capture the main intuition behind the counterintuitiveness of the original Repugnant Conclusion? Or is unboundedness an essential part of the counterintuitiveness of it such that this negative feature is not preserved in the bounded domain?

More broadly, this result suggests asking why exactly the Repugnant Conclusion is counterintuitive. Is the quantity-quality trade-off involved in the Bounded Repugnant Conclusion II sufficiently similar to a general quality-quantity trade-off problem for every aggregative axiology (see Budolfson & Spears (2018c), discussed above) to make it unsuitable as a condition on theory choice with respect to aggregative axiologies?

Ultimately, we need to more carefully scrutinize the source of the counterintuitiveness of the original Repugnant Conclusion to know whether it will carry over to the bounded domain. Moreover, could the force of bounded impossibility theorems be weakened by finding good reasons to restrict the upper bound on the domain further? And will the further assumptions that seem to be need for bounded theorems, such as assumptions regarding the possible size of the involved populations, the difference between very high and very low positive welfare, and the measurement of welfare (in the above example we assumed a ratio scale which isn't necessary for the unbounded theorems) open up for ways of escaping the theorems that are not available in the unbounded domain? This is an important but neglected area of research in population ethics which the second deflationary response puts focus on.

5. Conclusion

Policy analysis requires an axiology, population dynamics are important to climate change, and there is radical disagreement among experts about population axiology (Arrhenius (forthcoming), (2000a), (2000b), (2001), (2011)). Does this state of affairs limit our ability to know how to respond to climate change? Although several prominent voices have voiced this Worry, we suggested that it is not obviously well-founded, and we have highlighted two possible deflationary responses. In the first, we noted that many important policy questions are likely to be subject to simple, cross-theoretical dominance resolutions, as illustrated by a corner solution to an optimization problem. In

the second deflationary response, we observed that the intuitions that support the axioms that lead to the Repugnant Conclusion also support the axioms in the bounded case while avoiding the Repugnant Conclusion. Because any real-world policy question is a question about a bounded population domain (even if potentially very large in quantity), we can adopt these axioms for purposes of policy in their modified bounded form.

We also noted some important limitations and possible problems for these deflationary strategies. Regarding the first deflationary response, we noted that the climate policy menu under consideration may not yield one dominating option. Moreover, there could be additional considerations, such as bounded political capital, which could complicate the issue such that it cannot be settled by the suggested dominance-identification procedure, or could simply the issue by further reducing the practical space of policy options to those in which many axiologies agree.

Regarding the second deflationary response, there is the worry that the impossibility theorems might reappear over a bounded domain when the classical adequacy conditions are appropriately adjusted. An important challenge highlighted by considering the Repugnant Conclusion on a bounded domain is the need to identify exactly what constitutes the main counterintuitiveness of the Repugnant Conclusion and whether it carries over from the unbounded to the bounded domain (or, perhaps, to any other domains). This is a neglected but important area for further research in light of the impossibility theorems in population axiology on unbounded domains and the possibility theorem above on bounded domains.

In the meantime, we need not overstate the *practical* importance of the Repugnant Conclusion and other challenging problems in population ethics as we seek to cope with important challenges for the future of humanity. As we have tried to show, scepticism and paralysis are not yet warranted, as there are promising deflationary responses to the impossibility theorems and strategies for gaining consensus given disagreement for practical policymaking. Policy analysis may not need to wait for greater consensus in population ethics.[27]

References

Anglin, W. (1977). The Repugnant Conclusion. *Canadian Journal of Philosophy*, *7*(4), 745–754.

[27] Thanks to Krister Bykvist, Tim Campbell, Diane Coffey, Melissa LoPalo, Kevin Kuruc, Tristram McPherson, Josh Petersen, Sangita Vyas, and audiences at Paris School of Economics and the Australian National University.

Arrhenius, G. (forthcoming). *Population Ethics: The Challenge of Future Generations*. Oxford University Press.

Arrhenius, G. (2000a). An Impossibility Theorem for Welfarist Axiologies. *Economics and Philosophy*, *16*(02), 247–266.

Arrhenius, G. (2000b). *Future Generations: A Challenge for Moral Theory*. Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2:170236

Arrhenius, G. (2001). What Österberg's Population Theory Has in Common With Plato's. In *Omnium-gatherum. Philosophical Essays Dedicated to Jan Österberg on the Occastion of his Sixtieth Birthday* (Vol. 50, pp. 29–44). Uppsala: Department of Philosophy, Uppsala University: Uppsala Philosophical Studies.

Arrhenius, G. (2011). The Impossibility of a Satisfactory Population Ethics. In H. Colonius & E. N. Dzhafarov (Eds.), *Descriptive and Normative Approaches to Human Behavior, Advanced Series on Mathematical Psychology* (pp. 1–26). World Scientific Publishing Company.

Arrhenius, G., Ryberg, J., & Tännsjö, T. (2014). The Repugnant Conclusion. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014). Retrieved from http://plato.stanford.edu/archives/spr2014/entries/repugnant-conclusion/

Asheim, G. B., & Zuber, S. (2014). Escaping the repugnant conclusion: Rank discounted utilitarianism with variable population. *Theoretical Economics*, *9*(3), 629–650. https://doi.org/10.3982/TE1338

Beckstead, N. (2013). *On the Overwhelming Importance of Shaping the Far Future*. New Brunswick, NJ.

Blackorby, C., Bossert, W., & Donaldson, D. (1995). Intertemporal Population Ethics: Critical-Level Utilitarian Principles. *Econometrica*, *63*(6), 1303–1320. https://doi.org/10.2307/2171771

Blackorby, C., Bossert, W., & Donaldson, D. J. (2005). *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. New York: Cambridge University Press.

Blackorby, C., & Donaldson, D. (1984). Social Criteria for Evaluating Population Change. *Journal of Public Economics*, *25*(1–2), 13–33. https://doi.org/10.1016/0047-2727(84)90042-2

Bostrom, N. (2013). Existential Risk Prevention as Global Priority. *Global Policy*, *4*(1), 15–31. https://doi.org/10.1111/1758-5899.12002

Broad, C. D. (1979). *Five Types of Ethical Theory* (1 edition). London: Routledge.

Broome, J. (1992). *Counting the Cost of Global Warming*. Cambridge: The White Horse Press.

Broome, J. (2004). *Weighing Lives*. Oxford: Oxford University Press.

Broome, J. (2010). The most important thing about climate change. *Why Ethics Matters*, 101.

Broome, J. (2012a). *Climate Matters*. Norton.

Broome, J. (2012b). *Climate Matters: Ethics in a Warming World*. Retrieved from https://books.google.com/books?hl=en&lr=&id=RjrYYEk8GYQC&oi=fnd&pg=PA1&dq=%22to+their+clearest+and+most+compelling+essences.+Our+hope+is%22+%22the+Cost+of+Global%22+%228:+The+Future+versus+the%22+%22series+will+broaden+the+set+of+issues+taken+up+by+the+human%22+&ots=ctUEq1iVaP&sig=F6Y0B7587dZBLxnbtvRb9DPZUXc

Budolfson, M., & Spears, D. (2018a). *An impossibility result for decision-making under normative uncertainty*. mimeo.

Budolfson, M., & Spears, D. (2018b). *Methods for quantifying animal wellbeing and estimating optimal tradeoffs against human wellbeing — and lessons for axiology, including new arguments for separability*. mimeo.

Budolfson, M., & Spears, D. (2018c). *Why the Repugnant Conclusion is Inescapable*. mimeo.

Bykvist, K. (2017). Moral Uncertainty. *Philosophy Compass*, *12*(3), 1–8.

Bykvist, K., MacAskill, W., & Ord, T. (2019). *Moral uncertainty*. Oxford: Oxford University Press.

Cowen, T. (1996). What Do We Learn From the Repugnant Conclusion? *Ethics*, *106*, 754–775.

Fleurbaey, M., & Hammond, P. J. (2004). Interpersonally Comparable Utility. In S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of Utility Theory* (2nd ed., pp. 1179–1285). Boston: Springer.

Fleurbaey, M., & Tungodden, B. (2010). The tyranny of non-aggregation versus the tyranny of aggregation in social choices: a real dilemma. *Economic Theory*, *44*(3), 399–414.

Greaves, H. (2017). Population Axiology. *Philosophy Compass*, *12*(11), 1–15.

Greaves, H., & Ord, T. (2017). Moral Uncertainty About Population Axiology. *Journal of Ethics and Social Philosophy*, *12*(2).

Gustafsson, J. E. (forthcoming). Our Intuitive Grasp of the Repugnant Conclusion. In Arrhenius, Gustaf, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics*. Oxford: Oxford University Press.

Hare, R. M. (1988). Possible People. *Bioethics*, *2*(4), 279–293.

Hedden, B. (2016). Does MITE make right? In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vols. 1–11, pp. 102–135). Oxford University Press.

Hsiung, W., & Sunstein, C. R. (2006). Climate Change and Animals. *University of Pennsylvania Law Review*, *155*(6), 1695–1740.

Huemer, M. (2008). In Defence of Repugnance. *Mind*, *117*(468), 899–933. https://doi.org/10.1093/mind/fzn079

Mackie, J. L. (1985). Parfit's Population Paradox. In J. Mackie & P. Mackie (Eds.), *Persons and Values* (pp. 242–248). Oxford: Oxford University Press.

McMahan, J. (1981). Review: Problems of Population Theory. *Ethics*, *92*(1), 96–127.

McTaggart, J. M. E. (1927). *The Nature of Existence*. Cambridge.

Narveson, J. (1967). Utilitarianism and New Generations. *Mind*, *76*(301), 62–72.

Ng, Y.-K. (1989). What Should We Do About Future Generations? *Economics and Philosophy*, *5*(02), 235–253. https://doi.org/10.1017/S0266267100002406

Norcross, A. (1997). Comparing Harms: Headaches and Human Lives. *Philosophy & Public Affairs*, *26*(2), 135–167.

Pachauri, R. K., Mayer, L., & Intergovernmental Panel on Climate Change (Eds.). (2015). *Climate change 2014: synthesis report*. Geneva, Switzerland: Intergovernmental Panel on Climate Change.

Parfit, D. (1982). Future Generations: Further Problems. *Philosophy & Public Affairs*, *11*(02), 113–172.

Parfit, D. (1984). *Reasons and Persons* (1991st ed.). Oxford: Clarendon.

Parfit, D. (1986). Overpopulation and the Quality of Life. In P. Singer (Ed.), *Applied Ethics* (1 edition, pp. 145–164). Oxford: New York: Oxford University Press.

Parfit, D. (2016). Can We Avoid the Repugnant Conclusion? *Theoria*, *82*, 110–127.

Scovronick, N., Budolfson, M. B., Dennig, F., Fleurbaey, M., Siebert, A., Socolow, R. H., … Wagner, F. (2017). Impact of population growth and population ethics on climate change mitigation policy. *PNAS*, *114*(46), 12338–12343.

Shiell, L. (2008). The Repugnant Conclusion and Utilitarianism under Domain Restriction. *Journal of Public Economic Theory*, *10*(6), 1011–1031.

Sider, T. R. (1991). Might Theory X Be a Theory of Diminishing Marginal Value? *Analysis*, *51*(4), 265–271.

Sidgwick, H. (1907). *The Methods of Ethics*. London: Macmillan.

Tännsjö, T. (2002). Why We Ought to Accept the Repugnant Conclusion. *Utilitas*, *14*(03), 339–359.

Weitzman, M. L. (2009). On Modeling and Interpreting the Economics of Catastrophic Climate Change. *The Review of Economics and Statistics*, *91*(1), 1–19.

Appendix: A Smoothness Axiom and a New Argument for Total Utilitarianism

One response to the argument in Section 4 of the paper would be to agree that the modified axioms in their bounded versions capture *some* of our important intuitions, but not *all* of them, because there is a specific intuition that is omitted: that axiology is infinitely continuous. Consider the case in which a family of axiology is chosen, based on axioms some bounded and some unbounded, such that a social welfare function of form W is chosen, with the additional properties that:

- Bounded separability is assumed in social evaluation, so that the social welfare function can be written as a function of two variables: $\widehat{W} = g(\bar{n})h(\tilde{x})$, where $\bar{n}$ is the expected size of the population and $\tilde{x}$ is the expectation of $f(x)$. Then, *g* and the other functions are functions of all real numbers (not just counting numbers).
- *f* and *h* are both identity functions, as in total or average utilitarianism or Theory X', so the expression simplifies to: $\widehat{W} = g(\bar{n})\ \bar{x}$, where $\bar{x}$ is average utility.
- *g* is the identity function on the bounded set, as in total utilitarianism, and is any non-decreasing function outside of the bounded set, so the Repugnant Conclusion is not logically entailed (and therefore may or may not be avoided).

This is the sort of family of social welfare functions that section 4 highlights as possible, but extended for illustration to the case of expectations, in order to cover real numbers (and not only counting numbers of people); this will not appeal to advocates of non-expected social evaluations.

Now consider the intuition that axiology should be infinitely continuous – an intuition that may appear as an experience of unease about the boundedness of axioms. We can formalize this axiom as:

> **Smoothness:** *g* is C^{∞}, which is mathematical notation for the property of a function in which each derivative is continuous everywhere.

For real-valued functions, the Smoothness axiom would imply that they are polynomials. Therefore, g must be the identity function everywhere, because it is the identity function in the bounded set. The upshot is that the bounded assumptions above *plus the Smoothness axiom* imply that $\widehat{W}$ is expected Total Utilitarianism.[28]

The Smoothness axiom – and the intuitive response to the boundedness proposal that it captures – is therefore a new, constructive argument for Total Utilitarianism. With the smoothness axiom, $\widehat{W}$ implies the Repugnant Conclusion. Therefore, the Smoothness axiom introduces a new theoretical cost of avoiding the Repugnant Conclusion, in the context of the bounded axioms of $\widehat{W}$. If you find boundedness distasteful because you find infinite continuity to be a plausibly compelling property of axiology, then that intuition – in combination with other axioms – is a new argument counting in favour of Total Utilitarianism and acceptance of the Repugnant Conclusion. Of course, it can also be taken as a new impossibility theorem for those who accept smoothness, the bounded assumptions above, but not the Repugnant Conclusion.

[28] Thanks to Kevin Kuruc for suggesting consideration of this argument.