



ARTICLE

# What calibrating variable-value population ethics suggests

Dean Spears<sup>1</sup>  and H. Orri Stefánsson<sup>2</sup> 

<sup>1</sup>Economics Department and Population Research Center, University of Texas at Austin; Economics and Planning Unit, Indian Statistical Institute – Delhi Centre; IZA; Institute for Future Studies, Stockholm, Sweden; r.i.c.e. ([www.riceinstitute.org](http://www.riceinstitute.org)) and <sup>2</sup>Department of Philosophy, Stockholm University; Swedish Collegium for Advanced Study, Uppsala; Institute for Future Studies, Stockholm, Sweden

**Corresponding author:** H. Orri Stefánsson; Email: [orri.stefansson@philosophy.su.se](mailto:orri.stefansson@philosophy.su.se)

(Received 17 August 2021; revised 18 November 2023; accepted 18 December 2023)

## Abstract

Variable-Value axiologies avoid Parfit’s *Repugnant Conclusion* while satisfying some weak instances of the *Mere Addition* principle. We apply calibration methods to two leading members of the family of Variable-Value views conditional upon: first, a very weak instance of Mere Addition and, second, some plausible empirical assumptions about the size and welfare of the intertemporal world population. We find that such facts calibrate these two Variable-Value views to be nearly totalist, and therefore imply conclusions that should seem repugnant to anyone who opposes Total Utilitarianism only due to the Repugnant Conclusion.

**Keywords:** Repugnant Conclusion; variable-value views; utilitarianism; calibration

## 1. Introduction

Much research in population ethics is motivated by the quest to avoid what Parfit (1984) called the *Repugnant Conclusion*, one version of which states that:<sup>1</sup>

**The Repugnant Conclusion.** For any perfectly equal population of very well-off people, there is a better population consisting entirely of lives that are barely worth living.

Total Utilitarianism (TU), according to which a population is better the greater sum of welfare it contains, is widely recognized to entail the Repugnant Conclusion. No matter how well-off people are in some population *A*, and independently of

---

<sup>1</sup>Parfit’s own formulation of the Repugnant Conclusion states that: ‘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’ (Parfit 1984: 388). Our formulation is closer to Arrhenius’s (Forthcoming). Spears and Budolfson (2021) have argued that formalizations of the Repugnant Conclusion should be broader – including, for example, additions to unaffected, intersecting populations – but for this paper we ignore that proposal and focus on what they call a ‘restricted’ formalization.

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

A's size, there is some (potentially much bigger) imaginable population *Z* that contains a greater sum of welfare than *A* does – even though people in *Z* have lives that are each barely worth living (understood as having barely positive welfare).

Most paths to avoiding the Repugnant Conclusion begin by abandoning what Parfit called the *Mere Addition principle*, which can be stated thus:

**Mere Addition.** By adding any life worth living to any population, without making anyone else worse off, we do not make the population worse.

Total Utilitarianism implies the Mere Addition principle. But this principle is violated by *Average Utilitarianism* (AU), according to which a population is better the greater average welfare it contains. And Average Utilitarianism avoids the Repugnant Conclusion: Population *Z*, whose members all have lives that are barely worth living, contains lower average welfare than *A*. So *A* is better than *Z*, according to Average Utilitarianism.

Somebody who abandons the Mere Addition principle thinks that adding a life worth living, without making anyone worse off, can make a population worse. But what about adding a life *well* worth living? Consider merely adding a person who lives a very good life by modern standards: say, a happy professor living in a developed country in 2023. Surely by adding a person like that to any population, without thereby making anyone else worse off, we have not made the population worse? Not according to Average Utilitarianism. To see this in an absurd example: adding our professor to a single-person ‘population’ whose member is only a tiny bit better-off than the professor makes the resulting population *worse*, according to Average Utilitarianism. In fact, if the future of humanity is as long and wonderful as some hope (see e.g. Ord 2020), then adding a person like this to the *actual* intertemporal world population<sup>2</sup> might make the resulting population worse, according to Average Utilitarianism. This anti-natalist implication of Average Utilitarianism violates even a very weakened version of the Mere Addition principle.

In light of the above counterintuitive implications of on, the one hand, Total Utilitarianism (implying the Repugnant Conclusion) and, on the other hand, of Average Utilitarianism (violating even a highly weakened Mere Addition principle), some theorists have been attracted to a family of views that are often called *Variable-Value views*.<sup>3</sup> Some views within this family avoid the Repugnant Conclusion altogether while capturing the intuition that adding a well-off person to a *small* population always makes the resulting population better. More specifically, these views hold that the quantity that added persons (with fixed levels of welfare) contribute towards the overall value of a population decreases as the size of the population increases, *cumulatively* contributing only a bounded amount, which is how such views escape the Repugnant Conclusion.

<sup>2</sup>By ‘intertemporal world population’ we mean the totality of humanity throughout history. We focus on *human* populations. This is not because we think that the welfare of animals is unimportant. But how precisely to integrate animal welfare into population ethics is far from evident. Therefore, for the sake of simplicity, we focus on human welfare and populations of people.

<sup>3</sup>Hurka (1983) coined the term, and was probably the first to suggest such a view in response to Parfit’s Repugnant Conclusion, but views in this family have since been proposed or investigated by Ng (1989), Sider (1991), Asheim and Zuber (2014), and Pivato (2020), although not all of these authors endorsed the Variable-Value axiology that they identified or explored.

Various versions of Variable-Value views have been rigorously formalized. These formalizations and the ensuing analysis have focused on *qualitative* properties of Variable-Value views: with which *axioms* do they comply? However, there has not been a similar focus on the *quantitative* implications of Variable-Value views. In particular, one might wonder *how fast* the quantity that an added person contributes towards the overall value of a population diminishes as the size and average welfare of the population increases, and what implications that will have for various trade-offs between increasing the size and the average welfare of a population. Similarly, one might wonder precisely which weakenings of the Mere Addition principle these views can accommodate without implying seemingly repugnant instances of the Repugnant Conclusion.

Our aim in this paper is to offer some suggestive examples that fit in this quantitative gap in the population ethics literature.<sup>4</sup> So, our paper has more in common with the population ethics sub-literature that argues using striking, carefully chosen examples than with the sub-literature that provides qualitative, axiomatic proofs. In particular, we shall, first, conditionally assume that a very weak and intuitively compelling version of the Mere Addition principle holds even under what we take to be plausible (optimistic) empirical predictions about the future. Next, we calibrate two leading members of the Variable-Value family – namely, Number-Dampened Generalized Utilitarianism (NDGU) and Rank-Discounted Generalized Utilitarianism (RDGU) – to this assumption. Informally, the weak Mere Addition principle that we assume ensures that merely adding people who are very well-off by modern standards, such as happy professors in the developed world, does not make the population worse. The empirical prediction is that the future of humanity is long and prosperous, such that, in particular, the average welfare of the future population is even higher than the welfare of someone who is very well-off by modern standards.

We should stress two things from the start. First, for our argument to succeed we do not have to assume that the future *must* (or even *will*) be prosperous and populous. We only need the conditional claim that *if* the future is prosperous and populous, you would nevertheless support a mere addition of a happy present-day professor. Second, from the point of view of the universe there is of course nothing special about relatively well-off professors in 2023 when evaluating the full intertemporal population. We write of this case merely because we expect that both you, the reader, and we, the authors, have some intuition about it.

Our main observation is that, when combined with the above two assumptions, these two leading members of the Variable-Value family calibrate to be nearly

---

<sup>4</sup>Our aim is not to examine *all* Variable-Value views. In particular, because we are principally investigating the usefulness of weakening Mere Addition in response to the Repugnant Conclusion, we shall not be concerned with those Variable-Value views that satisfy the strong version of Mere Addition (i.e. the version entailed by Total Utilitarianism), such as the theory examined in Sider (1991). Instead, the aim is to examine those views that (unlike Average Utilitarianism) imply some weak instance of Mere Addition, without implying the strong version of Mere Addition.

We note also that a normative reason for excluding from our examination the view in Sider (1991) is that it implies what Arrhenius's (Forthcoming) calls 'The Very Anti Egalitarian Conclusion', and which states that for any perfectly equal population of at least two persons with positive welfare, there is a population which has the same number of people, lower average (and thus lower total) welfare and inequality, which is better. In fact, Sider himself rejects the view due to implications like this (Sider 1991: 270).

totalist, in the sense that they come quite close to agreeing with Total Utilitarianism. In particular, they imply countless instances of the Repugnant Conclusion. (By an ‘instance’ of the Repugnant Conclusion, we mean the judgement that some particular population consisting only of lives that are barely worth living is better than some particular perfectly equal population of very well-off people.) Of course, these Variable-Value views do not imply the *qualitative* Repugnant Conclusion stated above – which holds for *all* populations of very well-off people. But the aforementioned implications, we argue, should nevertheless seem repugnant to those who oppose the Repugnant Conclusion.<sup>5</sup> So, these results seem problematic for any argument that the Repugnant Conclusion requires us to reject Total Utilitarianism in favor of either NDGU or RDGU.

## 2. Variable-Value Axiologies and their Implications

Before discussing in detail the previously mentioned implications of Variable-Value axiologies, let us explain what types of axiologies we have in mind.

An infinity of population axiologies could have value vary, for instance, by whether population size is odd or even. Here we follow the population ethics literature which has understood the term ‘Variable-Value axiologies’ to refer to a particular structure of well-behaved families of social welfare functions that are designed to respond to the tension between, on the one hand, satisfying (some version of) Mere Addition and, on the other hand, avoiding the Repugnant Conclusion. In his first paragraph on Variable-Value, Arrhenius ([Forthcoming](#): 88) summarizes: ‘These principles are sometimes called “compromise theories” since a Variable-Value Principle can be said to be a compromise between Total and Average Utilitarianism. With small populations enjoying high welfare, a Variable Value Principle behaves like Total Utilitarianism and assigns most of the value to the total sum of welfare. For large populations with low welfare, the principle mimics Average Utilitarianism and assigns most of the value to average welfare.’ In the context of Ng’s (1989) trilemma among Mere Addition, Non-Antiegalitarianism<sup>6</sup> and avoiding the Repugnant Conclusion, we interpret the core of the Variable-Value idea to be a principled approach to rejecting Mere Addition in favour of the other two.<sup>7</sup> So, we are interested in complete and transitive families of social welfare functions that:

- are well-behaved in the sense of satisfying (*ex post*) Pareto, Extended Egalitarian Dominance, Non-Antiegalitarianism, and other non-controversial principles in the literature;<sup>8</sup>

<sup>5</sup>In fact, according to the principle of ‘unrestricted instantiation’ (Tännsjö 2020), these implications *must* be seen as repugnant if the Repugnant Conclusion is to be an argument against Total Utilitarianism.

<sup>6</sup>Non-Antiegalitarianism says that a perfectly equal population is better than a population with the same number of people, inequality and lower average welfare. (See e.g. Arrhenius 2000.)

<sup>7</sup>For this reason, we disregard Sider’s (1991) example of Geometrism, which rejects Non-Antiegalitarianism, and any other candidate axiology that does so; we are unaware of any author in the population ethics literature (including Sider in fact) who defends any Antiegalitarian Variable-Value proposal as plausibly the true population axiology.

<sup>8</sup>Since we will not make direct use of these principles, it suffices to define them informally. Pareto says that if every person is at least as well-off in population A as in population B, then A is at least as good as B.

- avoid the Repugnant Conclusion by rejecting Mere Addition;
- compromise between Total Utilitarianism for small populations and a population-size-insensitive alternative for large populations;<sup>9</sup> and
- can be calibrated by a parameter that quantifies the distance between Total Utilitarianism and the size-insensitive alternative as two convergent endpoints.

For clarity, we restrict our attention to the very large set of axiologies that, for perfectly equal populations of size  $n$ , reduce to  $g(\bar{x})f(n)$  for some increasing  $g$  and some concave, bounded  $f$ , where  $\bar{x}$  is the average welfare in the population. (In section 4, we however illustrate these axiologies' implications for a highly unequal intertemporal population.) In what follows, we ask quantitative questions about  $f$ . In particular, we use a calibration method (informally described in the next section) which determines, for instance, how quickly  $f$  approaches its bound as population size increases. A totalist  $f$  would be linear. We ask quantitatively: how totalist does a *plausible* Variable-Value view have to be? The answer, we argue, turns out to be rather totalist.

Two well-known examples of Variable-Value axiologies that satisfy all of the above discussed properties are Number-Dampened Generalized Utilitarianism (with an appropriate choice of functional form), and Rank-Discounted Generalized Utilitarianism. We will investigate these views in detail in section 4 after first informally presenting our general argument in the next section.

### 3. Our Calibration Argument, Informally

To clarify: it is not our own view that the Repugnant Conclusion is indeed repugnant or must be avoided by the correct theory of population ethics (Zuber *et al.* 2021). But many population ethicists have held that view, and we respond to them here. The population ethics literature is unclear on what exactly it is that is supposed to make the Repugnant Conclusion repugnant. We do not intend to take a stance on this question. In the rest of this paper, we will informally describe a judgement as being 'repugnant' when it involves preferring a much larger, much worse-off population over a large but much smaller, much better-off perfectly equal population. We mean 'repugnant in the sense of the Repugnant Conclusion'. Perhaps no such judgement is actually repugnant at all, but if repugnance is to be found in the Repugnant Conclusion, we propose that it needs to be found *consistently* in any similarly 'repugnant' judgements (and hereafter without the quotation marks).

But let's say that you do find the Repugnant Conclusion repugnant; that you find Non-Antiegalitarianism unrejectable; and that you therefore abandon the strong Mere Addition principle in favour of a Variable-Value axiology. Have you then escaped repugnance? We suggest that it depends upon the calibration of the resulting partially totalist view.

---

Extended Egalitarian Dominance says that if there is perfect equality in population A which is of greater size than population B, and every person in A has higher positive welfare than every person in B, then A is better than B. (See e.g. Arrhenius Forthcoming.) Non-Antiegalitarianism is defined in fn 6.

<sup>9</sup>Recall the previous citation to Arrhenius (Forthcoming: 88).

Here is why. We conjecture that you, the reader, have a strong conviction that the mere deletion of your life, or ours, or that of any other very well-off person by present-day standards, would not make the intertemporal world population better (even if it could be accomplished by magic, the ‘mere’ deletion having no effect on the welfare of anyone else) than the actual intertemporal world population in which this very well-off person by present-day standards in fact exists. A great life judged by the standards of our times does not in and of itself make the world worse.<sup>10</sup>

A natural question is why we are emphasizing our times, your life, or ours. We are evaluating the actual intertemporal population, after all, from the point of view of timeless population ethics. From the point of view of the universe, why is today special? The answer – as briefly mentioned in the Introduction – is that of course it is not. But we are writing to you, the reader. And you live in the present and come to our argument with beliefs and intuitions and, in the case of your life, hedonic experiences. We think lives like yours and ours are easy, readily available, and informative test cases for the present purposes. In fact, we think that test cases like these illustrate that there is at least one population and one welfare level such that adding a life at that welfare level to that population does not make matters worse: namely, adding a good life, like a professor’s in our times, to the actual intertemporal world population.

So why is agreement on this case important? Because it disciplines the calibration of any Variable-Value axiology, the combination of totalism on the one hand and, on the other hand, the number-insensitive counterpart (such as AU or leximin). We propose that two facts are probably true of the actual intertemporal world population. But more importantly, we assume that even if these two facts are not actually true, you would find that, if these facts *were true*, then the mere addition of a very well-off person by present-day standards would still not make the intertemporal world population worse. The two facts are:

- The future is vast: the actual intertemporal world population is enormous.
- The future is splendid: the actual intertemporal world population is full of lives much better than ours, i.e. much better than that of a happy present-day professor.

These facts, plus our judgement about the mere addition of a very well-off person by present-day standards, bound our calibration of quantitatively how non-totalist a Variable-Value axiology can be. In steps:

1. Because the population is enormous, we will be making decisions like an averagist (or otherwise like a non-totalist, depending on the details of the particular Variable-Value view), unless the tuning parameter is calibrated to move away from totalism only very slowly.

---

<sup>10</sup>Underlying our suggestion is the assumption that a typical professor in a developed country, and more generally people in the top part of the current global distribution of welfare, have lives that are *very* well worth living. While some may question this assumption (e.g. Benatar 1997), we hope that it will strike most readers as innocent.

2. Because someone like us is relatively badly-off compared to the splendid full distribution, adding such a person pulls down the average, disadvantages the lexical ladder, or otherwise looks undesirable to the non-totalist part of the axiology.
3. Because we judge that adding someone like us is nevertheless not a worsening, it must be the case that the tuning parameter is calibrated to move away from totalism only very slowly, so that the totalist benefits of the addition outweigh (in this case) the non-totalist costs of adding a relatively badly-off person.

And this brings us to the implications for the Repugnant Conclusion. If the tuning parameter is such that the Variable-Value axiology is, in the end, calibrated to be quite close to totalism, then it will often agree with totalism about how to rank populations. And that means that it will make many repugnant judgements where it prefers larger, worse-off populations to smaller, better-off ones, agreeing with totalism even in many quantitatively extreme cases. The universally quantified Repugnant Conclusion is escaped, but *repugnance* is not. So the spectre of repugnance seems hardly a reason to choose a Variable-Value approach. The next section makes this general argument quantitatively precise.

Some readers will have recognized that our argument presents an application of a familiar logic in decision theory: calibration of variable-value objective functions to reveal tensions between intuitions for large-quantity decisions and intuitions for small-quantity decisions. The leading result in this literature is Rabin's (2000) celebrated argument about expected utility theory. Formally, we extend Rabin's argument about choice under risk to analogous functional forms in population ethics.

Rabin established that an expected utility maximizer can only be moderately risk averse when relatively small sums of money are involved – e.g. always turning down 50-50 gambles between losing \$100 and winning \$105 – if she is extremely risk averse when larger sums of money are involved – e.g. turning down 50-50 gambles between losing \$2000 and winning any (including infinite) amount of money. So, the lesson of Rabin's argument is that an expected utility maximizer is either surprisingly risk averse when stakes are large or surprisingly risk neutral when stakes are small. Similarly, the lesson of our calibration exercise is that Variable-Value views are either surprisingly totalist or surprisingly strongly anti-natalist, assuming a vast and splendid future.<sup>11</sup>

#### 4. Calibrated Examples Using Two Leading Variable-Value Axiologies

To proceed, let us stipulate that there is a welfare level beyond which lives at that level are excellent by the standards of 21st-century developed countries. For concreteness, we shall occasionally assume that a typical happy professor in a present-day developed country is at that level – we assume that most readers of this

<sup>11</sup>Nebel and Stefánsson (2023) apply a similar logic to inequality averse views about how to order populations of a fixed size, in particular, to Prioritarianism and Generalized-Gini Egalitarianism, and find that such views can only be moderately inequality averse when small differences in welfare are at stake if they are extremely inequality averse when larger welfare differences are at stake.



paper will have familiarity with such a life. When we state two Variable-Value axiologies algebraically, we will assume that the lifetime welfare of such excellent lives can be represented by some positive real number. The particular number we use to represent the welfare of excellent lives does not matter, so long as we can assume that these lives are indeed excellent; that excellent lives are clearly better than lives barely worth living; and that there could be lives that the axiology values as even better than these excellent lives (such as in a long, prosperous future). This last stipulation prevents the social evaluation of lifetime welfare from being too sharply bounded above, an assumption which some might deny. (We briefly return to this issue in the next section.)

For precision and reproducibility, we have posted at the journal website an editable Excel spreadsheet that conducts all of the computations in this calibration exercise.<sup>12</sup>

We adopt the notation that finite vectors of real numbers are populations, where these number represent each person's lifetime welfare. For instance,  $\mathbf{x} = (x_1, \dots, x_n)$  is a population.  $\mathbf{X} = \{\mathbf{x}, \mathbf{y}, \dots\}$  is the set of all populations. The natural-number length of the vector (e.g.  $\mathcal{N}(\mathbf{x})$ ) is the size of the population.  $\mathbf{x} \preceq \mathbf{y}$  means that  $\mathbf{y}$  is at least as good as  $\mathbf{x}$ . The first view in the Variable-Value family that we consider can be stated as follows:<sup>13</sup>

**Number-Dampened Generalized Utilitarianism (NDGU).** *There is a concave (and increasing) real-valued function  $f$  such that for any  $\mathbf{x}, \mathbf{y} \in \mathbf{X}$ :*

$$\mathbf{x} \preceq \mathbf{y} \Leftrightarrow \bar{x} f(\mathcal{N}(\mathbf{x})) \leq \bar{y} f(\mathcal{N}(\mathbf{y}))$$

To our knowledge, this family was introduced to the literature implicitly in the diagrams of Hurka (1983). The concavity of  $f$  means that NDGU reduces the value of additions to the population as population size grows. Moreover, if (but only if)  $f$  is bounded, then NDGU avoids the universally quantified Repugnant Conclusion.

For illustrative purposes, we assume in our example that

$$f(\mathcal{N}(\mathbf{x})) = 1 - e^{-\frac{\mathcal{N}(\mathbf{x})}{\alpha}}$$

for some  $\alpha > 0$ . The parameter  $\alpha$  is the crucial parameter of calibration that tunes how quickly, as population size increases,  $f$  transitions from a TU-like gradient for small populations to an AU-like nearly constant 1 for large populations. Larger  $\alpha$  is more totalist,<sup>14</sup>  $\alpha$  nearer to zero is more averagist.

Now let  $\mathbf{x}_{\square} = (x_{[1]}, \dots, x_{[j]}, \dots, x_{[m]})$  be the nondecreasing reordering of  $\mathbf{x}$ , that is, an ordering of the welfare levels in  $\mathbf{x}$  such that for any  $i$ ,  $x_{[i]} \leq x_{[i+1]}$ . The second Variable-Value view that we shall consider can then be stated as follows:

<sup>12</sup>In addition, since inequality in the distribution of welfare makes a difference according to RDGU, the Excel sheet contains a calibration exercise for RDGU with inequality. As can be seen there, however, the introduction of inequality does not substantially affect our result.

<sup>13</sup>Recall from section 2 that  $\bar{x}$  is the average welfare in population  $x$ . In the below calibration, we however allow that it may be average *transformed* (or scaled) welfare.

<sup>14</sup>To see this, consider multiplying a population size,  $n$ , by  $\lambda$ . How does  $\frac{f(\lambda n)}{f(n)}$  behave as  $\alpha$  gets large? Using L'Hôpital's rule, the limit of the ratio is described by  $\lambda e^{-\frac{n(1-\lambda)}{\alpha}}$ , which goes to  $\lambda$  for any  $n$  as  $\alpha$  gets large.



**Rank-Discounted Generalized Utilitarianism (RDGU).** *There is a  $\beta \in (0, 1)$  such that for any  $x, y \in X$ :*

$$x \approx y \Leftrightarrow \sum_r \beta^r g(x_{[r]}) \leq \sum_r \beta^r g(y_{[r]})$$

where  $g$  is increasing and weakly concave.

This view was introduced and characterized by Asheim and Zuber (2014). It avoids the (universally quantified) Repugnant Conclusion because  $\beta^1 + \beta^2 + \beta^3 \dots$  is a convergent series, which ensures that the aggregated value of a perfectly equal population is bounded and remains finite, no matter how large it becomes. In other words, the (universally quantified) Repugnant Conclusion does not follow from RDGU.

Like  $\alpha$  tunes NDGU,  $\beta$  tunes RDGU. Asheim and Zuber (2014) prove that as  $\beta$  approaches 1, RDGU approaches totalism,<sup>15</sup> and as  $\beta$  approaches 0, RDGU approaches a variable-population version of leximin.

RDGU does not satisfy the strong Mere Addition principle that Total Utilitarianism entails. This is because adding a life to a population lowers the weights of any otherwise-existing better-off lives, which may worsen the population by more than the additional life improves it. However, RDGU must satisfy *some* weakened Mere Addition principle, since  $\beta > 0$ , which means that *some* mere additions are valuable. And, in fact, the closer  $\beta$  is to 1, the closer RDGU comes to agreeing with totalism. We want to examine what RDGU implies if we assume that it satisfies a particular, plausible and very weak instance of the Mere Addition principle.

With these definitions, we are ready to calibrate. What does it imply for NDGU's  $\alpha$  and RDGU's  $\beta$  if adding a present-day excellent life to the intertemporal world population does not make the population worse?

The first step is to state a plausible case for the mere addition of an excellent life not making the intertemporal world population worse. Table 1 describes two populations differentiated by one excellent life. The past population size of 120 billion is chosen to match demographers' estimates; the future population size is chosen (as a conditional hypothetical) to reflect optimistic longtermist hopes. If either NDGU or RDGU are used in a scaled or transformed version, then 'average utility' refers to those transformed values; otherwise it is simply average lifetime welfare.

The next step is to find the parameter values. If the population with the added excellent life is at least as good as the population without it, then NDGU's  $\alpha \geq 2.34639 \times 10^{12}$  and RDGU's  $\beta \geq 0.999999999999777$ . In other words, we have now calibrated NDGU and RDGU to what we propose is a plausible judgement about a particular mere addition, given the above realistic (but of course not unquestionable) empirical assumptions.

Finally, we can apply these calibrations from a real-world case to the theoretical case of repugnant-style conclusions. In Parfit's canonical formulation of the Repugnant Conclusion, the small population consists of 10 billion people, 'all with very high quality of life' (Parfit 1984: 388). With the above calibrations, how large

<sup>15</sup>In fact, it approaches Critical-Level Generalized Utilitarianism, but for simplicity we are ignoring non-zero critical levels here.

**Table 1.** Two hypothetical populations: a case of weak mere addition

Subpopulations	Number of lives	Average utility
Without added excellent life		
past	$1.2 \times 10^{11}$	1
excellent lives	$10^9$	2
wonderful future	$10^{12}$	10
With added excellent life		
past	$1.2 \times 10^{11}$	1
excellent lives	$10^9 + 1$	2
wonderful future	$10^{12}$	10

would a large population have to be in order to be judged as better, if (potentially scaled) per-person welfare were only 1% as great as in the small population?

For Total Utilitarianism, the answer is immediate: At least a trillion people, that is, 100 times as large as the small population. NDGU, calibrated as above, however recommends the large population if it contains at least 1.30 trillion people. In other words, according to this calibrated version of NDGU, a perfectly equal population of 10 billion very well-off people is worse than a population of 1.30 trillion people where each person's (potentially scaled) welfare is only 1% as great as in the small population. RDGU however recommends the small population if it contains at least 1.13 trillion people. In other words, according to calibrated RDGU, a perfectly equal population of 10 billion very well-off people is worse than a population of 1.13 trillion people where each person's welfare is only 1% as great as in the small population.

So, these views, once calibrated to recommend the mere addition in Table 1, differ from Total Utilitarianism in this repugnant-like judgement by respectively 30% and 13%. These are not radical, qualitative differences. We propose that anyone worried about the repugnance of totalism should find the same repugnance in these judgements. So, NDGU and RDGU, calibrated to a plausible, real-world judgement, substantively agree with Total Utilitarianism in this theoretically prominent (but practically unrealistic) choice between large and small populations.

## 5. Lesson and Concluding Remarks

Recall that the intuitive appeal of Variable-Values views was supposed to be that they could avoid the Repugnant Conclusion while satisfying at least some weak instances of the Mere Addition principle. We have now seen, however, that if two leading members of the family of Variable-Values views satisfy what we take to be a very plausible, and certainly weak, instance of Mere Addition – and moreover do so when we make plausible empirical assumptions about the intertemporal world population but don't assume that the contributive value of welfare is bounded above – then these two members have implications that, we suspect, those who oppose

Total Utilitarianism due to the Repugnant Conclusion will find repugnant.<sup>16</sup> The results of this paper could thus be interpreted as suggesting, along with a growing literature – including Spears and Budolfson (2021) on additions to an unaffected population and Arrhenius and Stefánsson (Forthcoming) on risky choice between uncertain populations – that the repugnance of Total Utilitarianism can be found in even those theories that were designed to avoid such repugnance.

Our results in this paper are merely suggestive. They are offered to provide perspective based on quantitative examples, not qualitative proofs. Here is what they suggest to us: First, a wish to avoid repugnance may not be a good reason to favour the aforementioned members of the Variable-Value family over Total Utilitarianism. Second, perhaps it is time to learn to live with such repugnance.

Before concluding, it is worth acknowledging that those who are sufficiently attracted to the Variable-Value views we have discussed might draw a radically different conclusion from our results, namely, that they show that one should put an upper bound on the contributive value of a person's welfare.<sup>17</sup> In our result for RDGU, we assumed that the  $g$  function is concave but not bounded; whereas in our NDGU result we assume that, for any given population size, the contributive value of a person's welfare is proportional to their (potentially transformed or scaled but unbounded) level of welfare. Those who are sufficiently committed to either of these views may thus take our paper as an argument that the contributive value of a person's welfare is bounded above. Such bounding will of course have implications that many will find unattractive. In particular, it implies that some arbitrarily great welfare gains – enjoyed by people who are already very well off – result in no gain in moral value. Still, we acknowledge that some may find this to be a bullet worth biting, and others may even find this to be independently plausible.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0266267124000026>.

**Acknowledgements.** We are grateful for very useful written comments from Jake Nebel, Christian Tarsney, and three referees for this journal; and for helpful suggestions and questions from the audience at Climate Ethics Workshop, Institute for Futures Studies, in April 2020.

**Funding statement.** Spears' research is supported by NICHD grants K01HD098313 and P2CHD042849. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Stefánsson's research is supported by Riksbankens Jubileumsfond through a *Pro Futura Scientia* fellowship.

## References

- Arrhenius G. 2000. An impossibility theorem for welfarist axiologies. *Economics & Philosophy* 16, 247–266.
- Arrhenius G. Forthcoming. *Population Ethics: The Challenge of Future Generations*. Oxford: Oxford University Press.
- Arrhenius G. and H.O. Stefánsson Forthcoming. Population ethics under risk. *Social Choice and Welfare*.
- Asheim G.B. and S. Zuber 2014. Escaping the repugnant conclusion: rank-discounted utilitarianism with variable population. *Theoretical Economics* 9, 629–650.

<sup>16</sup>Recall also fn. 5.

<sup>17</sup>We are grateful to the referees for making us see the need to acknowledge this.

- Benatar D.** 1997. Why it is better never to come into existence. *American Philosophical Quarterly* **34**, 345–355.
- Hurka T.** 1983. Value and population size. *Ethics* **93**, 496–507.
- Nebel J.M. and H.O. Stefánsson** 2023. Calibration dilemmas in the ethics of distribution. *Economics and Philosophy* **39**, 67–98.
- Ng Y.-K.** 1989. What should we do about future generations?: Impossibility of Parfit's theory x. *Economics & Philosophy* **5**, 235–253.
- Ord T.** 2020. *The Precipice: Existential Risk and the Future of Humanity*. New York, NY: Hachette Books.
- Parfit D.** 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Pivato M.** 2020. Rank-additive population ethics. *Economic Theory* **69**, 861–918.
- Rabin M.** 2000. Risk aversion and expected-utility theory: a calibration theorem. *Econometrica* **68**, 1281–1292.
- Sider T.R.** 1991. Might theory x be a theory of diminishing marginal value? *Analysis* **51**, 265–271.
- Spears D. and M. Budolfson** 2021. Repugnant conclusions. *Social Choice and Welfare* **57**, 567–588.
- Tännsjö T.** 2020. Why Derek Parfit had reasons to accept the Repugnant Conclusion. *Utilitas* **32**, 387–397.
- Zuber S., D. Spears, J. Gustafsson et al.** 2021. What should we agree on about the Repugnant Conclusion? *Utilitas* **33**, 379–383.

**Dean Spears** is an Associate Professor in the Economics Department and Population Research Center at the University of Texas, where he is Director of the Population Wellbeing Initiative. His research studies children in developing countries, the demography of low birth rates, and population ethics. Email: [dean@riceinstitute.org](mailto:dean@riceinstitute.org). URL: <https://sites.utexas.edu/pwi/>

**H. Orri Stefánsson** is Professor of Practical Philosophy at Stockholm University, Pro Futura Scientia Fellow at the Swedish Collegium for Advanced Study, and researcher at the Institute for Futures Studies. His current research focuses on decision-making under extreme uncertainty, ethics of risk, distributive ethics, climate ethics and catastrophic risk. URL: <https://orristefansson.is/>