Why the Repugnant Conclusion is Inescapable

Mark Budolfson* Dean Spears†

December 2018

Working Paper
Princeton University
Climate Futures Initiative in Science, Values, and Policy

Abstract

The spectre of the repugnant conclusion and the search for a population axiology that avoids it has endured as a focus of population ethics. This is in part because the repugnant conclusion is often interpreted as a defining problem for totalism, while the implications of averagism and related views are taken to illustrate the theoretical cost of avoiding the repugnant conclusion. However, we show that this interpretation cannot be sustained unless one focuses only on a special case of the repugnant conclusion: namely, the subset of instances of the repugnant conclusion where there is no portion of the population unaffected by the choice between population outcomes (as in Derek Parfit's original illustration). To avoid an inappropriate focus on only this proper subset of instances of the repugnant conclusion, we formulate a general characterization of the repugnant conclusion. We then prove formally that all leading welfarist axiologies imply this conclusion, including averagism and Ng's Theory X′, including probabilistic and 'very repugnant' variants that involve the addition of negative lives. We then prove that the full range of axiologies considered by population ethics each imply an extended version of the repugnant conclusion, including axiologies that are incomplete, intransitive, rank-dependent, person-affecting, and/or pluralist. The upshot is that the repugnant conclusion does not ultimately tell against any approach to axiology, and the methodological requirement to avoid the repugnant conclusion should be dropped from population axiology.

*Harvard University, Princeton University CFI, and University of Vermont.

†University of Texas at Austin, Princeton University CFI, Indian Statistical Institute-Delhi Centre, IZA Institute of Labor Economics, and Institute for Future Studies, Stockholm. corresponding author: dspears@utexas.edu. We are grateful for comments from and discussion with Gustaf Arrhenius, Geir Asheim, Walter Bossert, John Broome, Krister Bykvist, Tim Campbell, Erik Carlson, Diane Coffey, Marc Fleurbaey, Mike Geruso, Johan Gustafsson, Nicholas Lawson, Melinda Roberts, Orri Stefánsson, Larry Temkin, Stéphane Zuber, participants in the UT-Austin population ethics group, and seminar participants at the Institute for Future Studies.

1 Introduction

Following Parfit (1984), the repugnant conclusion is often formulated as the claim that, for any population of very well-off people, there is an imaginable larger population whose existence would be better, even if everyone in the larger population has lives that are barely worth living. The literature assumes that the repugnant conclusion must be avoided, and this has been one of the central motivations of the population ethics literature since Parfit introduced it.

The repugnant conclusion is often taken to be a devastating problem to totalism. In addition, Arrhenius (2000, 2009, n.d.) proves an impossibility theorem showing that no axiology can vindicate a set of intuitive judgments about population ethics while also avoiding the repugnant conclusion. Arrhenius notes that one response to his theorem could be a thoroughgoing skepticism or paralysis. However, he is much more enthusiastic about the possibility of a deflationary response (although he does not identify such a response himself): namely, to "try to find a way to explain away the relevance of the [repugnant conclusion and associated impossibility] theorem for moral justification."

In this paper we take up this deflationary project, and take it one step further by proving that all leading population axiologies imply the repugnant conclusion, including averagism, Ng's Theory X′, and indeed all axiologies in the population ethics literature. Our proof therefore refutes the assumption that the repugnant conclusion is a special problem for totalism or any special class of axiologies, and it refutes the idea that the guiding principle of population axiology should be to avoid the repugnant conclusion.

The plan of the paper is as follows. In the next section we offer a general characterization of the repugnant conclusion. This is an important first step toward our results, as previous formal work[1] has relied on a characterization of the repugnant conclusion that is restricted to a proper subset of all possible instances of the repugnant conclusion — namely, instances of the repugnant

[1]Including foundational results by Arrhenius, upon which we build.

conclusion in which there is no *base population* of individuals who are entirely unaffected by the choice between population outcomes, to which additional people are added in the "good lives" vs. "larger number of repugnantly bad lives" outcomes that are compared. In contrast, we identify a more fundamental and fully general characterization of the repugnant conclusion that does not artificially rule out any instances. In the next section, we make this mathematically precise.

We then prove (Theorems 1 and 1*) that instances of the repugnant conclusion are implied by all leading welfarist axiologies. Thus, we show that it is an illusory property of welfarist axiologies such as average utilitarianism and Ng's Theory X′ that they avoid the repugnant conclusion, where this illusion arises from too-narrow a focus on a mere subset of instances of the repugnant conclusion. We next prove a similar theorem (Theorem 2) for all 'probabilistic' social welfare functions that derive from the welfarist axiologies covered by Theorem 1 and weight uncertain outcomes by their probability.

We then define an extended version of the repugnant conclusion and prove it is implied by all axiologies in the population ethics literature (Theorems 3, 3*, and 4), including axiologies that are incomplete, intransitive, rank-dependent, person-affecting, and/or pluralist. Ultimately, this shows that the repugnant conclusion is a problem for every axiology, and so repugnance is not a special problem for any of the leading families of axiologies.[2]

We then turn to a deeper explanation of these results, marshalling a number of observations made in the literature outside of population ethics. We note that a common theme emerges, which is that any axiology that aggregates over unbounded spaces will have repugnant implications. This is the fundamental mechanism that our proofs exploit. In light of our formal results, we take the upshot to be that if the repugnant conclusion is unavoidable, then we should not try to avoid it. So, the repugnant conclusion does not ultimately tell against any axiology, and

[2]Thus, our response contrasts with many other responses in the literature such as Ng (1989), Ryberg (1996a), Portmore (1999), Tännsjö (2002), Huemer (2008), and Gustafsson (2018) who each argue in favor of accepting the repugnant conclusion even on the assumption (which they take for granted, by focusing on the restricted subset with no base population) that other axiologies can avoid it. Our demonstration that the very repugnant conclusion is inevitable need not deny that it is deeply regrettable.

the methodological requirement to avoid the repugnant conclusion should be dropped from population axiology. More generally, having counterintuitive implications over an unbounded space does not tell against any view — a fact that we believe should emerge as a guiding insight to axiology, both in population theory and more generally.[3]

2 General Characterization of the Repugnant Conclusion

Our first step is to provide a general characterization of the repugnant conclusion (RC):

> **Repugnant Conclusion (RC)**: For any:
>
> - Arbitrarily large number of arbitrarily high utility people: $n^h > 0$, $u^h > 0$, and
> - Arbitrarily small positive value of life: $\varepsilon > 0$,
>
> There exists:
>
> - A number of small-positive-value lives: n^ε,
> - A (possibly empty) set of base population lives with number $n^0 \geq 0$ and utility u^0,
>
> such that it is better to add to the base population the small-positive-value lives than to add the high-utility lives. The "base population" lives are lives that occur regardless of whether the small-positive-value[4] lives or the high-utility lives are chosen.

This general characterization of the RC is needed, because Parfit's original example is only one of many possible instances of the RC:

> **Restricted RC**: **Parfit's Original Example**. For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger

[3] This conclusion complements prior work about axiology in unbounded spaces (Cowen, 1996; Norcross, 1997, 1998; Arrhenius, 2000; Fleurbaey and Tungodden, 2010; Arrhenius, n.d.; Bossert, 2017).

[4] We take no position on what would constitute a small-positive-value life: whether it must be bland, or good could barely outweigh bad, or it could be excellent but very short, or could be common among non-human animals, all of which the literature has considered (Parfit, 1984; Portmore, 1999; Tännsjö, 2002; Arrhenius, n.d.).

> 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, p. 388).

Parfit's example is the canonical illustration of the RC, but it is not the only possible instance. Of particular importance for the proofs that follow, we note that in Parfit's example the base population is empty ($n^0 = 0$). In our terminology, this assumption that $n^0 = 0$ is the defining feature of a *restricted* RC. In contrast, consider the following *unrestricted* RC:

> **Unrestricted RC**: **Analog to Parfit's Example**. For any possible addition to the actual human population where the total future addition is at least ten billion more people, all with a very high quality of life, there must be some much larger total future addition to the actual human population whose addition, if other things are equal, would be better even though its members have lives that are barely worth living.

Parfit's original, restricted example and this unrestricted analog are both instances of the RC. Parfit's example succeeds as a succinct initial illustration of the RC, while the unrestricted analog succeeds in describing a more detailed and more realistic choice situation in which the RC looms. Neither of these instances of the RC is more fundamental or more repugnant than the other — i.e., it is irrelevant whether $n^0 = 0$ or $n^0 > 0$ to the degree of repugnance of these conclusions, and to their status as instances of the RC. This is because both include the essential characteristic of the RC: forgoing many very-high-welfare lives in favor of low-welfare lives. The general characterization of the RC offered above correctly classifies both of these as instances of the RC, and thus avoids the illegitimate implication that *all* instances of the RC *must* be restricted RCs.

In contrast, previous work on the RC focuses on the restricted RC, especially in proving formal results.[5] For example, Arrhenius (2000, n.d.) defines the RC in a way that excludes any

[5]The proofs below concern the unrestricted VRC, in which low-positive-value lives are accompanied by arbitrarily many arbitrarily negative lives. Neither Arrhenius' review of population axioms nor any other source in the population ethics literature considers our unrestricted VRC, which we show is implied not only by totalism and averagism, but also by every leading welfarist axiology and, in its extended form, by all leading population axiologies. At the same time, our general approach is different from Arrhenius approach and complementary to it —

non-restricted RC as an instance of the RC, and proves that a range of plausible axiologies each imply the restricted RC. Having noted that the restricted RC is only a proper subset of the RC, and having characterized the set of all instances of the RC, we are now in a position to prove in the next section that all leading axiologies imply the RC, including averagism, Ng's Theory X′, and indeed more.

Before turning to the proofs, we note that in what follows we prove results about the logically stronger *very* repugnant conclusion (VRC) (Arrhenius, 2003, n.d.), in which the many low-positive-value lives of the RC are also accompanied by many extremely negative lives full of suffering. This is useful because the VRC is taken to be even more repugnant than the RC, and so by proving that all leading population axiologies have the logically stronger implication of the VRC, we can provide an even stronger demonstration that there is ultimately no special problem for totalism or any other axiology provided by anything like the repugnant conclusion. And because being subject to the VRC entails being subject to the RC, the theorems that follow immediately entail results about the repugnant conclusion. The following provides a formal characterization of the VRC:

Very repugnant conclusion (VRC): For any:

- Arbitrarily large number of arbitrarily high utility people: $n^h > 0$, $u^h > 0$,
- Arbitrarily large number of arbitrarily negative utility people: $n^\ell \geq 0$, $u^\ell < 0$, and
- Arbitrarily small positive value of life: $\varepsilon > 0$,

There exists:

indeed, we take our results to build on his own, which provided most of the initial fundamental results in population axiology. Arrhenius' general approach is to show that sets of weak conditions are mutually inconsistent; we instead study the many axiologies which imply versions of one single condition that is slightly stronger than avoiding the original restricted VRC. Therefore, one way to understand our formal proofs in comparison with Arrhenius' results is that our unrestricted VRC is a strengthening of Arrhenius' original restricted (and therefore weaker) VRC, and we show that this strengthening is sufficient to apply to a much wider range of axiologies than the related, weaker versions that Arrhenius considers. Note that our statement of the unrestricted RC is similar to what Arrhenius (n.d.) calls the Strong Quality Addition Principle (see also Anglin (1977)).

- A number of small-positive-value lives: n^{ε},
- A (possibly empty) set of base population lives: $n^0 \geq 0$, u^0,

such that it is better to add to the base population the negative-utility and small-positive-value lives than to add the high-utility lives.

The very repugnant conclusion, as we have formalized it, implies the repugnant conclusion, which is the special case in which $n^{\ell} = 0$. Table 1 summarizes four categories of repugnant conclusions, according to the characterizations we have offered.[6]

Table 1: Four Categories of Repugnant Conclusion

	$n^0 \geq 0$	$n^0 = 0$
$n^{\ell} \geq 0$	Very Repugnant Conclusion	Restricted Very Repugnant Conclusion
$n^{\ell} = 0$	Repugnant Conclusion	Restricted Repugnant Conclusion

Again, the VRC is widely regarded to be even more repugnant than the repugnant conclusion, because it chooses very negative lives over very positive lives.[7] As Table 1 shows, the only distinguishing factor of the unrestricted case is the existence of unaffected the base population, which is irrelevant to repugnancy.[8] Thus, by proving that all leading axiologies imply the (unrestricted) VRC in what follows, we show that there is ultimately no special problem for totalism or any other axiology provided by anything like the repugnant conclusion.

[6] As a technical note, under the existence independence axiom (a property of any population axiology with an additively separable social welfare function, including totalism, prioritarianism, and CLGU), each restricted conclusion is equivalent to its unrestricted counterpart.

[7] As Arrhenius (2003) argues, "we might, for example, accept the Repugnant Conclusion but not the Very Repugnant Conclusion because we give greater moral weight to suffering than to positive welfare."

[8] One way to see the irrelevance of the base population to repugnancy is to realize that it could contain just one member, or members that live for an arbitrarily short time, or be of a different species than the added population members, or exist at a different place and time (such that, at the time when the added population lives and forever after, only perfectly equally well-off people at u^h or people at ε and u^{ℓ} live).

Table 2: Nine categories of welfarist axiologies

	$f(x) = h(x) = x$	f concave, $h(x) = x$	f concave, $h = f^{-1}$
g constant ($g' = 0$)	average utilitarianism	average prioritarianism	average egalitarianism
g linear ($g' = 1$)	total utilitarianism	total prioritarianism	total egalitarianism
g concave ($0 \leq g' \leq 1$)	Ng's (1989) X'	variable-value prioritarianism	variable-value egalitarianism

3 Theorem 1: All Leading Welfarist Axiologies Imply the Very Repugnant Conclusion

In this section, we prove that the VRC is true of every leading anonymous, aggregative welfarist axiology. In later sections we prove extended results about all other axiologies that are discussed in population ethics.

To proceed formally with an algebraic illustration, we begin by specifying the set of leading welfarist axiologies:

> **Leading welfarist axiologies**: Let n represent the size of a population ($n \geq 0$) and u real-number-valued lifetime utility levels. All leading welfarist axiologies take the form:
>
> $$W = g(n)\, h\Big(\overline{f(u)}\Big).$$
>
> g need not be defined for non-natural numbers. If $n > m$ then $g(n) \geq g(m)$. $f(0) = h(0) = 0$. $g(n) > 0$ if $n > 0$. f and h are strictly increasing and continuous.[9]

What does this category rule out and what does it rule in? Table 2 notes that this definition

[9]This formalization is related to a similar family of functional forms used by Greaves and Ord (2017). They interpret this as encompassing a broad tent of population ethics views. Our formulation differs from theirs only in that it is more inclusive. For example, W takes an average over transformed utilities ($\overline{f(u)}$ rather than $f(\bar{u})$), to permit prioritiarianism, which Adler (2009) defines to require additive separability.

yields commonly-held welfarist views as special cases. Many other functional forms are also included, but we highlight those in Table 2 partly because averagism and Ng's (1989) Theory X' are taken in the literature to avoid the repugnant conclusion, given that the literature has focused only on the subset of the RC involving the restricted RC, and that it is true that those two axiologies avoid the restricted RC.[10]

Two requirements are implied by the definition of W above (see section 3.1 and appendix A.1 for discussion of weaker, more fundamental conditions that also imply the VRC):

- **Same-number generalized welfarism**. For any fixed population size, the functional form is characterized by continuity, anonymity, strong Pareto, and same-number independence axioms (Blackorby et al., 2005).[11] However (as the example of averagism demonstrates) different-number independence need not be satisfied.

- **Extended egalitarian dominance**. If population A is perfectly equal-in-welfare and 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 (compare Arrhenius, n.d.).

Same-number generalized welfarism produces an additive $\sum f(u_i)$ structure for a given n. Extended egalitarian dominance is what requires g to be weakly increasing.

With this definition, we state our first theorem:

> **Theorem 1**. The *very repugnant conclusion* is true for all W of the form of *leading welfarist axiologies*; for any ε, n^h, u^h, n^ℓ and u^ℓ, we can find n^ε, n^0 and u^0 such that:
>
> $$g\left(n^0+n^h\right)h\left(\frac{n^0 f\left(u^0\right)+n^h f\left(u^h\right)}{n^0+n^h}\right)<g\left(n^0+n^\ell+n^\varepsilon\right)h\left(\frac{n^0 f\left(u^0\right)+n^\ell f\left(u^\ell\right)+n^\varepsilon f\left(\varepsilon\right)}{n^0+n^\ell+n^\varepsilon}\right).$$
>
> *Proof.* The proof will select n^ε, n^0, and u^0 by construction:

[10] Similar remarks apply to the views outlined in Hurka (1983), although they are not highlighted in Table 2.

[11] Blackorby et al. (2005) describe this as "generalized utilitarian", which is consistent with both prioritarian and egalitarian functional forms – see Table 2.

Case I: g is constant. $\exists c$ such that $g(n) = c$, for all n.

In the inequality, the g terms cancel. Pick $u^0 < 0$ and n^0 such that $n^0 f(u^0) < -n^h f(u^h)$. Then, since $g(.) > 0$, $h(0) = 0$ and h is strictly increasing, the LHS is negative. By choosing a large enough n^ε, the RHS can be brought arbitrarily close to $h(f(\varepsilon)) > 0$, and the inequality would hold.

Case II: g is (at least weakly) increasing but bounded. $\exists b$ such that $g(n) < b$, for all n.

Let $\tilde{b}$ be the least upper bound. Choose a small $\delta > 0$. Let n^* be such that $\tilde{b} - g(n) < \delta$ for all $n > n^*$. Now, for any choice of $n^0 > n^*$ we would have $\frac{g(n^0+n^\ell+n^\varepsilon)}{g(n^0+n^h)} > \frac{\tilde{b}-\delta}{\tilde{b}}$. As in case i, choose $n^0 > n^*$ and $u^0 < 0$ such that $n^0 f(u^0) < -n^h f(u^h)$, and a sufficiently large n^ε. Then we would have

$$h\left(\frac{n^0 f(u^0)+n^h f(u^h)}{n^0+n^h}\right) < \frac{\tilde{b}-\delta}{\tilde{b}} h\left(\frac{n^0 f(u^0)+n^\ell f(u^\ell)+n^\varepsilon f(\varepsilon)}{n^0+n^\ell+n^\varepsilon}\right) < \frac{g(n^0+n^\ell+n^\varepsilon)}{g(n^0+n^h)} h\left(\frac{n^0 f(u^0)+n^\ell f(u^\ell)+n^\varepsilon f(\varepsilon)}{n^0+n^\ell+n^\varepsilon}\right)$$

Case III: g is (at least weakly) increasing and unbounded. $\forall b$, $\exists n$ such that if $m > n$ then $g(m) > b$.

Set $u^0 = \varepsilon$ and choose any n^0. For large enough n^ε, the argument of the h function on the LHS would be lower than the argument of the h function on the RHS, so that the ratio of the h functions on the LHS is < 1. Also, since g is unbounded, for large enough n^ε we have $g(n^0 + n^\ell + n^\varepsilon) > g(n^0 + n^h)$. Choose n^ε large enough so that both of these are true, and the inequality is obtained. □

Having proved Theorem 1, we highlight some of its limitations, and state and preview some additional results.

First, Blackorby and Donaldson's (1984) critical-level generalized utilitarianism (CLGU) with a positive critical level does not satisfy extended egalitarian dominance. However, our definition of a tradeoff-making social welfare function includes CLGU if the critical level and repugnant conclusion are understood as in Broome (2004), as Broome "standardizes" the functional form of

f so that zero is the "neutral" level, his term for the critical level. Therefore, under Broome's standardized interpretation, CLGU implies the VRC, because it is included in Theorem 1. Moreover, as Broome shows, on this understanding CLGU implies the restricted RC, which Broome finds initially unintuitive but ultimately acceptable.[12] Further theorems in section 5 extend our main results to include non-standardized CLGU.

Second, our definition of leading welfarist axiologies does not include rank-dependent views (Sider, 1991; Asheim and Zuber, 2014), because they are not same-number generalized welfarist, as they violate same-number independence. Independence is attractive in same-number, risk-free cases, especially for ethical decision-making. Because ranking full populations includes the far future and far past, rank-dependent views require knowing the well-being of unaffected people in the far future and far past to make present-day policy and ethical choices, which is normatively implausible and epistemically infeasible. Despite the exclusion from Theorem 1, rank-dependent axiologies and indeed all other welfarist axiologies discussed in the literature are susceptible to an extended version of the very repugnant conclusion, which we prove in section 5 (Theorem 3).

3.1 Non-algebraic statement of Theorem 1

In appendix A.1, we prove a more general version of Theorem 1, which uses properties of axiologies rather than the algebraic functional form of the definition of W. We show that the VRC is implied by any axiology that satisfies transitivity, extended egalitarian dominance, and a property that we define formally in the appendix called "convergence in signs." We interpret convergence in signs to be a requirement of minimal tradeoff-making across people, or

[12] See also Bykvist (2007). Arrhenius (n.d.) formalizes this as the "weak repugnant conclusion," arguing that the neutral level and the zero level should be considered distinct. Under that interpretation, CLGU with a positive general level escapes the repugnant conclusion at the price of implying the "sadistic conclusion," the "weak repugnant conclusion," and related conditions (Arrhenius, 2000; Bossert, 2017), where the sadistic conclusion is a mirror image of the repugnant conclusion that similarly produces an unintuitive consequence in an unbounded space. Understood in that way, CLGU would still imply a repugnant outcome, but would be excluded from our definition for violating egalitarian dominance.

equivalently, a very weak continuity:

> **Convergence in signs (informal statement)**. If enough identical lives, at a utility level u, are added to any base population, eventually (possibly in a very large population) the result is a combined population that is overall just as good as some perfectly equal population of the same size as the combined population, in which every person has a utility of the same sign as u.

Every axiology of form W (including averagism, totalism, and Theory X′) satisfies transitivity, extended egalitarian dominance, and convergence in signs, which we interpret to be basic requirements for a plausible welfarist axiology.

Other social welfare functions also are consistent with extended egalitarian dominance and convergence in signs. One example is from unpublished notes by Partha Dasgupta, which can be generalized to the following functional form:

$$h\left(\overline{f(u)}\right) - \frac{\alpha}{n},$$

where all elements are as defined above, and α is a positive constant. This family of functional forms is not plausible for social evaluation, but it can avoid the restricted RC. However, it is consistent with extended egalitarian dominance and convergence in signs, and therefore implies the VRC.

> **Theorem 1***. Any transitive axiology that satisfies *extended egalitarian dominance* and *convergence-in-signs* implies the *very repugnant conclusion.*
>
> *Proof.* See appendix A.1.

See appendix A.1 for technical details. Informally, note that Theorem 1* does not require a complete social ordering and does not require real-valued utilities. This is important because some axiologists — for a notable recent example, Chang (2016) — have argued that incompleteness may be a way to avoid the RC; Theorem 1* shows that incompleteness offers no escape. Moreover,

as the appendix explains, extended egalitarian dominance can be replaced with *priority for lives worth living* (which is an alternative principle that any population containing only people with positive welfare is better than any population containing only people with negative welfare) and yield the same result. Finally, other axiologists have proposed lexical or non-Archimedean axologies as a response to the RC. In the appendix, section A.2 describes how our proofs extend to both of these approaches.

4 Theorem 2: All Leading Expected Social Welfare Functions Imply the Probabilistic Very Repugnant Conclusion

The frontier of research in population axiology has recently incorporated innovative extensions to risky cases where possible future people exist with probability between zero and one (Voorhoeve and Fleurbaey, 2016; Roberts, 2018a; Arrhenius and Stefánsson, 2018; Nebel, forthcoming). Consider a social welfare function that ranks probabilistic distributions, rather than certain outcomes; a natural extension of the leading welfarist axiologies characterized in the previous section to such an expected social welfare function is:

$$E[W] \equiv \sum_s \pi_s W_s = \sum_s \pi_s g(n_s)\, h\left(\overline{f(u_s)}\right),$$

where s indexes states, π_s is the probability of state s, and the same restrictions apply to g, h, and f. Arrhenius and Stefánsson (2018) and [removed for review] call these "population prospects." Blackorby et al. (1998) and Fleurbaey and Zuber (2015) have explored utilitarian and egalitarian, respectively, expected social welfare functions of this form.

In a probabilistic setting, choosing low-positive-value lives over high-positive-value lives perhaps can be made to seem even more repugnant, by assigning a low probability to the outcome in which the benefits of the low-positive-value lives are even realized. Arrhenius and

Stefánsson (2018), in a series of important and novel results that to our knowledge were the first to consider the repugnant conclusion in a probabilistic setting, document some consequences of expected population axiologies. Among other important results, they show that expected average utilitarianism results in the "risky repugnant conclusion," which in our terminology is a probabilistic version of the restricted RC. They also show that expected total utilitarianism implies the "risky very sadistic conclusion," which is the probabilistic version of the conclusion that for any population with negative welfare, there is a population with positive welfare which is worse. These results are important, because the literature assumes that average utilitarianism avoids the repugnant conclusion, while total utilitarianism avoids the very sadistic conclusion.

Beyond these important results, it is possible to extend our first theorem from the previous section to a probabilistic setting with the following characterization of the probabilistic very repugnant conclusion, which is not investigated by Arrhenius and Stefánsson (2018):

Probabilistic Very Repugnant Conclusion: For any:

- Arbitrarily large number of arbitrarily high utility people: $n^h > 0$, $u^h > 0$,
- Arbitrarily large number of arbitrarily negative utility people: $n^\ell \geq 0$, $u^\ell < 0$,
- Arbitrarily small positive value of life: $\varepsilon > 0$,
- Arbitrarily small probability: $p > 0$

There exists:

- A number of small-positive-value lives: n^ε,
- A (possibly empty) set of base population lives: $n^0 \geq 0$, u^0,

such that it is better to add to the base population the negative-utility lives with certainty and the small-positive-value lives with probability p than to add the high-utility lives with certainty.

It can be shown that all probabilistic versions of the axiologies in Theorem 1 imply the probabilistic very repugnant conclusion:

> **Theorem 2.** The *probabilistic very repugnant conclusion* is true of any expected social welfare function of the form $E[W]$ (i.e. is true of the expected version of every axiology covered by Theorem 1).
>
> *Proof.* The proof similar to in the non-probabilistic case of Theorem 1, and is based on the linearity of the expectation operator and the ability to choose n^{ε} and n^{0} to be very large. In particular, we proceed in cases of g. If g is unbounded, then choose $u^{0} > \left|u^{\ell}\right|$ and $n^{0} > n^{\ell}$, so h is positive, and then let n^{ε} go to infinity, so pgh will be larger than any positive number. If g is bounded, then make u^{0} very negative, and make n^{0} large enough that g is arbitrarily close to the bound and the negative-addition and positive-addition populations without the ε lives are arbitrarily close to one another in value. Then make n^{ε} much larger than n^{0}. □

Note that, in the repugnant outcome, the very bad lives are certainly added, but the small-positive-value lives are added only with a very small probability. This demonstrates that the first theorem is robust to considering a probabilistic setting, rather than a deterministic setting; this is unsurprising because of the continuous, linear nature of the expectation operator.

5 Theorem 3: All Welfarist Axiologies in the Literature Imply the Extended Very Repugnant Conclusion, Including Rank-Dependent and Critical Level Axiologies, as well as All Leading Intransitive and Incomplete Axiologies

This section extends Theorem 1 to include the full range of welfarist axiologies discussed in the literature, including non-standardized critical level generalized welfarism (CLGU) and rank-dependent axiologies, including maximin and maximax, as well as leading approaches to rejecting completeness and transitivity.

We begin with rank-dependent (RD) axiologies of the sort proposed by Sider (1991) and developed and characterized by Asheim and Zuber (2014). Both Sider (1991) and Asheim and Zuber (2014) believe rank-dependent axiologies are of special interest because they escape the restricted RC. At the same time, Sider (1991) and many others argue that rank-dependent axiologies are implausible on the grounds that they do not satisfy even same-number separability or independence. This is analogous to the Stone Age or Egyptology objection to average utilitarianism (McMahan, 1981), but stronger because AU satisfies independence in same-number cases.

One family of rank-dependent axiologies developed in a series of important papers by Asheim and Zuber (2014) is rank-discounted generalized utilitarianism (RDGU), which focuses on the functional form $\sum_i \beta^{i-1} u_i$, where $0 < \beta < 1$ and i is a rank order of increasing utilities. RDGU avoids the restricted RC by discounting the importance of the most-well-off lives in a population. However, while RDGU avoids the restricted RC, it violates the "mere addition" principle even for very high utility levels, because it implies that for any high level of positive utility v and for any β, it is a worsening to add one v-person to a population consisting of one living person with utility above $\frac{v}{1-\beta}$. Sider (1991) notes the existence of Geometrism, a similar rank-dependent

axiology where the ranking is done separately within the two sets of negative and non-negative lives, and the rank within each set is decreasing in absolute value. Geometrism satisfies the mere addition principle.

Both views escape the restricted RC in the same way that a maximin social welfare function would: by valuing the well-being of some people much more than others, depending upon their rank within the population. In particular, on both views a large number of lives with small positive value makes only a bounded contribution to social welfare. But, by discounting in this way, both Geometrism and RDGU are vulnerable to the following conclusion:

> **Very Repugnant Dictator Conclusion (First Variant).** For any very small utility increment $\varepsilon > 0$, for any terrible negative quality of life, and for any large number of lives, there exists a population such that it would be considered an improvement to add the large number of terrible, negative quality lives, while also increasing the well-being of one existing person (who can be any number of ranks above the worst-off[13]) by ε.

The proof follows the exact method of proof by which rank-dependence avoids the restricted RC: because a geometric series with $0 < \beta < 1$ has a finite sum, and because this sum can be discounted to be less than any positive number by making its starting point late enough in the rank, any number of lives (of positive or negative quality) can be made of arbitrarily small importance in absolute value by positing an unaffected starting population to which they are added in which the added lives have the highest value.

Therefore, both Geometrism and RDGU have repugnant consequences. Geometrism also has further counterintuitive consequences: Sider (1991), and every other writer on the topic of which we are aware, rejects Geometrism because of its anti-egalitarianism, even in same-number cases, because it would transfer well-being from low-weight people with low positive utility to high-weight people with high positive utility. RDGU avoids such anti-egalitarianism, but as a

[13]In this and the following conclusions for rank-dependent axiologies, we note that the conclusion does not depend on the person being worst-off, as some rank-dependent views, such as maximin, are highly attentive to the worst-off person.

consequence implies the following conclusion:

> **Nearly Anti-Dominance Conclusion**. for any very small utility decrement $\varepsilon < 0$, for any very high quality of life, and for any large number of lives, there exists a population such that it would be considered a *worsening* to add the large number of very high quality lives, while also decreasing the well-being of one very-well-off person (who can be any number of ranks above the worst-off) by ε.

RDGU also implies a conclusion that combines the two mechanisms:

> **Very Repugnant Dictator Conclusion (Second Variant)**. For any very small utility increment $\varepsilon > 0$, for any very high quality of life and any large number of high-quality lives, and for any very low quality of life and any large number of low-quality lives, there exists a population such that it is better to add the large number of very low-quality lives, while also increasing the well-being of one person (who can be any number of ranks above the worst-off) by ε, instead of adding the large number of very high-quality lives.

This result resonates with Theorem 1 above: although rank-dependent axiologies avoid the restricted Repugnant and Sadistic Conclusions (i.e. where there are no additions to an independent base population), they nonetheless entail repugnant consequences when adding to base populations.

With this is mind, we can take the first step towards extending Theorem 1 to rank-dependent axiologies by considering a motivating example given RDGU. Asheim and Zuber's (2014) RDGU was not included in Theorem 1's set of leading population axiologies, but RDGU does imply the second variant of the Very Repugnant Dictator Conclusion, which together with Theorem 1 combines to make an extended, disjunctive conclusion, that applies to RDGU and to all W of the form in Theorem 1: For any very small, positive utility quantity $\varepsilon > 0$, for any very high quality of life and any large number of high-quality lives, and for any very low quality of life and any large number of low-quality lives, there exists a population such that, instead of adding the large

number of very high-quality lives, it would be better to both (a) add the large number of very low-quality lives, and also (b) do either of:

- increase the well-being of one person (who can be any number of ranks above the worst-off) by ε, or
- add some number of new lives, each of value ε.

What this conclusion demonstrates is that there are two ways to change the properties of some person in a population by ε. The Dictator Conclusions above consider *improvements* in the quality of existing lives by ε; Theorem 1 considers *adding new lives* of value ε. These are comparably small changes in the well-being of a full population; it is useful to introduce a definition that captures both of these types of changes. So, let us define an ε-change as a change that makes a difference in either one of these two ways:

> ε-**change**: Let $\varepsilon > 0$ represent any small, positive quantity of well-being. An ε-change either:
>
> - increases the well-being of one person by ε, or
> - adds one new life of well-being ε.
>
> One or more ε-changes can be part of an overall package of changes to a population, but to qualify as an ε-change, a change must be the only change that a particular person receives.

For example, an ε increase could involve slightly improving a tiny headache. One way to see that a ε increase could be very repugnant is to recall Portmore's (1999) suggestion that ε lives in the restricted RC could be "roller coaster" lives, in which there is much that is wonderful, but also much terribly suffering, such that the good ever-so-slightly outweighs the bad. Here, one admitted possibility is that an ε-change could substantially increase the terrible suffering in a

life, and also increase good components; such a ε-change is not the only possible ε-change, but it would have the consequence of increasing the total amount of suffering.

With this definition of ε-change in hand, we can now characterize the extended very repugnant conclusion:

Extended very repugnant conclusion (XVRC): For any:

- Arbitrarily large number of arbitrarily high utility people: $n^h > 0$, $u^h > 0$,
- Arbitrarily large number of arbitrarily negative utility people: $n^\ell \geq 0$, $u^\ell < 0$, and
- Arbitrarily small positive quantity of well-being: $\varepsilon > 0$,

There exists:

- A number of ε-changes: n^ε, and
- A (possibly empty) set of base population lives,

such that it is better to both add to the base population the negative-utility lives and cause n^ε ε-changes than to add the high-utility lives.

The XVRC extension from the VRC retains all of the repugnance of choosing many terrible lives over many wonderful lives for merely ε-benefits to other people. Moreover, if ε-changes are of the "roller coaster" form, they could increase deep suffering considerably beyond even the arbitrarily many u^ℓ lives, and in fact could require everyone in the chosen population to experience terrible suffering.

We have seen above that the XVRC conclusion is true for rank-dependence. Recall from the discussion of W in section 3 that CLGU is included in W and therefore in Theorem 1 under Broome's "standardized" interpretation, but not under Arrhenius' "weak repugnant conclusion" interpretation. However, even under Arrhenius' interpretation, CLGU implies the extended very repugnant conclusion, as do many other axiologies:

Theorem 3. The *extended very repugnant conclusion* is true for all W of the form in Theorem 1, and also for RDGU, for Geometrism, for axiologies that attend only to a finite set of ranks (such as maximin or maximax), for Necessitarianism[14], for Presentism, and for both standardized and non-standardized CLGU.

Proof. The conclusion for Theorem 1's W, including standardized CLGU, is implied by that Theorem. The conclusion for RDGU is implied by the Very Repugnant Dictator Conclusion (Second Variant). For non-standardized CLGU, $n^h f\left(u^h\right) - n^\ell f\left(u^\ell\right)$ is finite and positive; the conclusion is satisfied for any necessarily-existing population large enough that $n^0 \geq n^\varepsilon > \frac{n^h f\left(u^h\right) - n^\ell f\left(u^\ell\right)}{\varepsilon}$, with n^ε people receiving ε-changes. For maximin and maximax, construct a preexisting population such that the additional very good and very bad lives are neither the best or the worst; make an ε-change improving the worst or best, respectively. For Geometrism, construct a pre-existing population with many bad lives, each of which is worse than u^ℓ, and many good lives, each of which is better than u^h; then by increasing the number of pre-existing very bad and good lives, the additional good and bad lives can be made arbitrarily unimportant, and less valuable than the opportunity to increase the well-being of the pre-existing best and worst lives by ε. For Necessitarianism and Presentism, assign ε improvements to presently-existing people. □

As in the case of Theorem 1, instead of referring to axiologies in the literature, we can alternatively provide conditions that are sufficient for the Extended Very Repugnant Conclusion. Here is one example, which is of particular interest because it applies to axiologies that do not

[14] Greaves (2017) includes Necessitarianism and Presentism in a catalog of population axiologies with the following definition: "Presentism holds that the only persons who matter are persons who presently exist (and, in particular, not those who might or will exist in the future). … Necessitarianism holds that the only persons who matter, in a situation of deciding between A and B, are those who exist both A and in B (i.e., those who, for the purposes of the present decision situation, exist 'necessarily' — who exist regardless of how the decision is settled — and not those whose very existence is contingent on one's current decision)." Theorem 3 also trivially applies to what Greaves calls "Actualism" and "Harm-minimization theories," as presented there.

require a complete or transitive axiology, which shows that rejecting completeness or transitivity should not be seen as a way of avoiding repugnant conclusions:

> **Theorem 3***. Any axiology implies the *extended very repugnant conclusion* in the same-sized-addition[15] case where $n^h = n^\ell$, if the axiology endorses the principle that:
>
> "For two same-sized populations A and B, it is sufficient for A to be better than B if all of the following are true:
>
> - Mean utility in A is greater than mean utility in B;
> - For every increasing, strictly concave function ϕ, mean ϕ-transformed utility in A is greater than B (second-order stochastic dominance);
> - For some fixed number n, which can be set arbitrarily large, the worst-off person in A is better-off than the worst-off person in B; the second-worst-off person in A is better off than the second-worst-off person in B; and so on up to the n^{th}-worst-off pair; and,
> - A is perfectly equal and B contains inequality."
>
> *Proof.* The proof shows how a comparison can be constructed such that the combined population with the u^ℓ lives is A and the combined population with the u^h lives is B. Set $u^0 = u^\ell - \varepsilon$. The ε-changes will be increases of the base population from u^0 to u^ℓ. This satisfies the equality of A and inequality of B. Choose $n^0 > n$. Increase n^0 until both mean conditions are met (the ϕ-transformed condition follows from Jensen's inequality). Then $A > B$, and the extended very repugnant conclusion is fulfilled. □

In other words, to deny the same-sized-addition case of the XVRC is to deny that, in a same-number comparison of two populations meeting *all* of those criteria, A would be better than

[15] If the reader wishes, the XVRC could be modified to hold that for any n^h and any n^ℓ, there is a $\tilde{n}^h > n^h$ and a $\tilde{n}^\ell > n^\ell$ for which the XVRC as stated above holds. Then $\tilde{n}^h = \tilde{n}^\ell$ could be chosen, and the repugnance would only be increased by increasing the number of wonderful and terrible lives.

B.

We do not believe that a serious candidate for a welfarist axiology can deny this same-number condition. Because the comparison between A and B is a same-number comparison, the axiology need not be *complete* over different-number cases. Perhaps even more strikingly, because only two populations are compared, Theorem 3* does not require *transitivity*.[16] Because $n^h = n^\ell$ can be arbitrarily large, this same-sized-addition constraint does not impact the repugnance of the condition. Therefore, this result rebuts the suggestions of Chang (2016), Temkin (2012), and others that denying completeness or transitivity might be seen as a way of avoiding repugnant conclusions, and this rebuts the suggestion by Temkin (2012) and others that the repugnant conclusion provides reason to deny the transitivity of the better-than relation.

6 Theorem 4: All Expected Social Welfare Functions in the Literature Imply the Probabilistic Extended Very Repugnant Conclusion

Just as Theorem 2 extended Theorem 1 to expected social welfare functions corresponding to the axiologies covered by Theorem 1, so too in this section we extend Theorem 3 to the expected social welfare functions corresponding to the axiologies covered by Theorem 3. To begin, consider any social welfare function V and any set of probabilities over states $\{\pi_s\}$; a natural extension of the axiologies characterized in the previous section is: $E[V] = \sum_s \pi_s V_s$, where V_s is the social welfare function's value of state s.

Probabilistic Extended Very Repugnant Conclusion: For any:

- Arbitrarily large number of arbitrarily high utility people: $n^h > 0$, $u^h > 0$,

[16] Pummer (2018) makes a similar argument that potentially repugnant spectrum conclusions such as Hangnails for Torture can be reached without transitivity.

- Arbitrarily large number of arbitrarily negative utility people: $n^{\ell} \geq 0$, $u^{\ell} < 0$,
- Arbitrarily small positive quantity of well-being: $\varepsilon > 0$, and
- Arbitrarily small probability: $p > 0$,

There exists:

- A number of ε-changes: n^{ε}, and
- A (possibly empty) set of base population lives,

such that it is better to both add to the base population the negative-utility lives with certainty and cause n^{ε} ε-changes with probability p than to add the high-utility lives with certainty.

Theorem 4. The probabilistic extended very repugnant conclusion is true of every social welfare function of the form $E[V]$ (i.e. is true of the expected version of every axiology covered by Theorem 3).

Proof. Theorem 3 already showed this for Theorem 1's W; for expected maximin, maximax, Necessitarianism, and Presentism it is obvious; for RDGU, Geometrism, and non-standarized CLGU, reducing p is a change additively separable from the ε increases in the utility of pre-existing lives (once the pre-existing populations are constructed as in Theorem 3), so reducing p simply functions as reducing ε, which can be compensated for by increasing the number of ε-changes. □

Therefore, a very large set of axiologies (all social welfare functions in the population ethics literature of which we are aware, including more than every one considered by Greaves and Ord's (2017) thorough and leading survey[17]) imply the very repugnant conclusion that it is better

[17] They explain their menu of population axiologies: "Our list includes every actually-advocated theory we are aware of that is both (i) sufficiently precisely specified for us to know what the corresponding value function is, and (ii) consistent with the structural limitations that we laid out in [their] section 2. While we don't explicitly discuss it here, our results also hold for Geometrism (Sider, 1991) — a theory that was described but never seriously advocated."

(in expectation) certainly to create an arbitrarily large number of arbitrarily bad lives rather than certainly to create an arbitrarily large number of arbitrarily excellent lives, in order to additionally have, along with the new bad lives, an arbitrarily tiny probability of some arbitrarily tiny benefits.

7 Corollaries: All leading pluralist and person-affecting axiologies imply the repugnant conclusion

Theorem 1 showed that all leading welfarist axiologies imply the Very Repugnant Conclusion, and Theorem 3 showed that all other welfarist axiologies considered in the literature imply the Extended Very Repugnant Conclusion; Theorems 2 and 4 established analogous results in a probabilistic context with expected social welfare functions.

Others have conjectured that the repugnant conclusion could be avoided by pluralist or person-affecting axiologies (Temkin, 2012; Roberts, 2015). In this section, we rehearse results from the existing literature that demonstrate that this is not so for the leading axiologies of these types. In particular, we discuss pluralism (the view that welfare is only one among several dimensions of goodness)[18] and person-affecting axiologies (which reject the impersonal aggregation of standard approaches to welfarism).

In a series of important papers, Gustaf Arrhenius has responded to both pluralist and person-affecting axiologies, and has shown that both imply the restricted RC. In the following sections 7.1 and 7.2, we highlight and rehearse Arrhenius' demonstrations and draw implications in the context of our theorems in this paper. We discuss pluralism and person-affecting axiologies in turn, although we note that they could be combined (e.g., into person-affecting pluralism of

[18] In the literature, pluralism is often understood to include views that are welfarist, but prioritarian or egalitarian rather than utilitarian (Temkin, 2012). Because we have already covered all these welfarist views with Theorems 1 and 2 above, we find it useful to use "pluralism" to refer to the set of non-welfarist pluralist views. Nothing important turns on this terminological matter of convenience.

which the welfare of persons is only one of the dimensions) without changing the force of these observations. We also note, in section 7.3, that Voorhoeve and Fleurbaey's (2016) Conditional on Existence View would imply the VRC.

7.1 Pluralist Axiologies

Theorems 1 through 4 have assumed that all relevant axiologies are welfarist axiologies. Yet, some philosophers have conjectured that the RC could be avoided by a pluralist axiology that includes welfarism as only one of multiple dimensions of value. However, Arrhenius proves an important corollary to his Theorems which also functions as a corollary to our Theorems. The intuition behind Arrhenius' corollary is that all pluralist axiologies remain vulnerable to the restricted RC because the non-welfarist dimensions of value can be held constant. We can state Arrhenius' corollary formally:

> **Arrhenius' Pluralist Corollary** For any pluralist axiology in which any welfarist dimension of value is combined with any number of non-welfarist dimensions of value, and for any axiologically-relevant difference in the welfare properties of any set of populations, a possible set of populations exists which (a) matches the welfare properties of the set of populations, and (b) holds equal the non-welfarist dimensions of value. Therefore, for every theorem in this paper and any pluralist axiology in which any welfarist dimension of value is combined with any number of non-welfarist dimensions of value, the theorem applies to the pluralist axiology, with non-welfarist dimensions of value held equal.

The upshot is that pluralism offers no escape from the repugnant conclusion.

7.2 Person-Affecting Axiologies

Person-affecting axiologies hold that for a population to be better or worse than another, it must be better or worse for someone. (*e.g.* Roberts, 2015, 2018b). The population ethics literature

Table 3: Parfit's original RC: Not resolved by person-affecting axiology

option A	10	10	$\cdots$	10	$*$	$*$	$*$	$\cdots\cdots$	$*$	$*$
option Z	$*$	$*$	$\cdots$	$*$	ε	ε	ε	$\cdots\cdots$	ε	$*$

Table 4: Person-affecting axiology implies Extended RC (XRC)

option A'	10	10	$\cdots$	10	$*$	$*$	$*$	$\cdots\cdots$	$*$	ε
option Z'	$*$	$*$	$\cdots$	$*$	ε	ε	ε	$\cdots\cdots$	ε	2ε

contains suggestions that a person-affecting axiology could avoid the RC. However, consider the example in Table 3. In the table, each column is a potential person, a number is a level of well-being, and an asterisk ($*$) indicates that a person does not exist under that option. If we also assume that A and Z are the only accessible options and (without loss of generality) that 10 represents a very high quality of life, the choice between A and Z is the same choice as in Parfit's original restricted RC. However, person-affecting axiologies fail to classify Z as worse than A, because no person exists in both A and Z, so neither is worse for anyone who exists in both,[19] so Z is not worse than A according to a person-affecting axiology. In this way, a person-affecting axiology fails to resolve the RC because it fails to imply that the larger population that is full entirely of lives that are barely worth living is worse.[20] The fact that person-affecting axiology fails to resolve the RC is in addition to the well-known, fundamental theoretical cost of person-affecting views, which is the non-identity problem.

Moreover, consider the case in Table 4, where again A' and Z' are assumed to be the only two accessible options. Here, Z' is strictly better than A' on a person-affecting view; this means that person-affecting axiologies imply what we called in Theorem 3 the *extended* repugnant conclusion, given that the rightmost person in the table receives an ε-change.

Arrhenius (2015) offers a formal proof that the repugnant conclusion is implied by all person-

[19]Roberts (2018b) offers an extended person-affecting axiology that also attends to the well-being that potential people might achieve in other accessible outcomes C, but for this example, A and B are the only accessible outcomes, so this theoretically important revision does not make a difference in this case.

[20]On a pluralist extension of person-affecting axiology, where person-affecting ties are broken by aggregate welfare, Z would be strictly preferred.

affecting axiologies that endorse the Subjunctive Person-Affecting Restriction, which following Holtug (2004) he defines as "if an outcome A is better than B, then A would be better than B for someone that would exist if either A or B were to obtain" (p. 114). Nebel (forthcoming) argues that person-affecting axiologies that do not endorse the Subjunctive Person-Affecting Restriction are implausible because they offer no reason not to create a person whose life will only be full of terrible suffering in every outcome where the person exists. Formally, Arrhenius proves the following:

> **Arrhenius' Person-Affecting Theorem**. There is no population axiology which satisfies the Subjunctive Person-Affecting Restriction, Egalitarian Dominance, and Inequality Aversion, and avoids the restricted Repugnant Conclusion.[21]

In other words, Arrhenius' impossibility theorems apply to any person-affecting approach that endorses the Subjunctive Person-Affecting Restriction. The implication of these formal results, including the examples in Tables 3 and 4, is that person-affecting axiologies cannot escape the RC.

7.3 The Conditional on Existence View

Voorhoeve and Fleurbaey (2016) have recently proposed the Conditional on Existence View as an approach to egalitarianism under social risk. According to this view, the currency of distributive ethics — whether utilitarian or otherwise — should combine an individual's final well-being and expected well-being conditional on existence. We will call this c_i for person i's currency.

Voorhoeve and Fleurbaey only consider same-number cases, and do not consider cases of different-number aggregation. To address such cases, the currency of distributive ethics would

[21] Arrhenius defines Egalitarian Dominance as the condition that if A is a perfectly equal population of the same size as B and if every person in A is better off than every person in B, then A is better than B. He defines Inequality Aversion formally as an extremely weak egalitarian condition, namely that for any welfare level of the best off and worst off, and for any number of best off lives, there is a (much) greater number of worst off lives such that it would be at least as good to have an equal distribution of welfare on any level higher than the worst off, other things being equal.

have to be aggregated by any of the aggregation rules considered in this paper (such as taking the total or the average of the currency). Without such an aggregation rule, the Conditional on Existence View does not imply anything about different number cases and does not form a complete welfarist axiology. An aggregation rule could take any of the forms from Theorem 1, but now applied to currency:

$$W^c = g(n)h\left(\overline{f\left(c_i\right)}\right).$$

As we have shown, all of these aggregation rules imply the VRC, so the Conditional on Existence View does, as well. For a simple example of the restricted repugnant conclusion, consider two outcomes: one in which very many people each receive a small amount of the currency, and another in which a smaller number of people each receive a large amount of the currency. If currency is totaled, then this is a currency-denominated version of the restricted repugnant conclusion. More broadly, for any view W^c, a currency-denominated non-restricted repugnant conclusion is entailed by Theorem 1. The upshot is that while the conditional on existence view might be a plausible understanding of egalitarianism, it does not avoid the repugnant conclusion. [22]

Other philosophers have considered versions of the conditional on existence view outside of axiology. For example, Frick (2018) considers that we may have moral reasons to promote a person's wellbeing only conditional on their existence. On this view, *reasons* for action are divorced from traditional axiological questions about the *betterness* of outcomes, and axiological questions are set aside, in contrast to the views considered here that assume a tight connection between reasons for action and traditional axiology. Because our focus in this paper is axiology only, we do not consider this important further range of views. (For further discussion, see chapter 12 of Arrhenius (n.d.).)

[22] Thanks to Marc Fleurbaey for helpful correspondence on these issues.

8 Intuition for the Theorems: All Axiologies are Exposed to Repugnance Over Unbounded Spaces

Totalism is willing to tradeoff the wellbeing of some individuals for a greater total sum of wellbeing among an unbounded number of other individuals. This exposes totalism to the RC, as Parfit's classic example shows. However, the fundamental mechanism of aggregation over unbounded spaces that creates exposure to repugnant conclusions is not a unique feature of totalism, as aggregation over an unbounded number of people is a feature of all leading welfarist axiologies. This is the fundamental explanation why all leading welfarist axiologies have the repugnant implications demonstrated by the theorems in the previous sections. This section presents intuition for that fact, by using a small number of examples to focus attention on this fundamental mechanism. Whenever aggregation is done over an unbounded space, repugnant outcomes inevitably occur. The proofs of the previous sections establish this; the examples in this section illuminate the fundamental mechanism.

Our method in this section is to compare Parfit's classic version of the RC with a small number of other cases to illustrate how aggregation over unbounded spaces create repugnant conclusions for non-totalist versions of welfarism even in same number cases. We rely on important examples identified by prior authors in other, same-number contexts. To begin, consider again:

> **Restricted RC: Parfit's Original Example**. 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, p. 388).

Now compare the following two classic examples that also involve aggregation over unbounded space that yield "repugnant" conclusions even in same-number cases:

> **Utility Monster Example**. For any possible population of at least ten billion people, all with a very high quality of life, there must be some equal-sized population whose

existence, if other things are equal, would be better even though its members have lives that are barely worth living except for one individual who has a sufficiently large level of utility (Nozick, 1974; Ryberg, 1996a).

This is a "repugnant"-like implication of averagism as well as totalism, given that the same number of individuals exist in both alternatives. In fact, the Utility Monster Example would apply similarly to both averagism and totalism in a "very repugnant" version where all lives but one have arbitrarily negative utility, instead of being barely worth living.

Consider also:

Tiny Headaches Example. For some imaginable population, all with a very high quality of life, there must be some equal-sized population whose existence, if other things are equal, would be better even though one individual has a life of torture while the others have lives that are equally good except for the avoidance of a tiny headache in each person's life (Norcross, 1997).

Again, this is a "repugnant" implication of averagism as well as totalism, given that the same number of individuals exist in both alternatives.[23]

More generally, these examples illustrate that the mechanism of aggregation over unbounded spaces creates exposure to repugnant conclusions. Repugnant implications are neither unique to totalism, specifically, nor to different-number cases, generally. Because all plausible axiologies permit aggregation over unbounded spaces, this means that all plausible axiologies are exposed to repugnant conclusions, as the theorems above demonstrate. Here we have intentionally used examples identified by prior authors in other same-number contexts to illuminate this fundamental mechanism (see also Cowen, 1996; Fleurbaey and Tungodden, 2010).

The intuition common to Theorems 1 through 4 is that aggregating consequences for an unbounded number of people is bound to create repugnant implications: either something that

[23]Appendix A.3 proves a further example of a repugnant conclusion for averagism, by a different method of proof than Theorem 1.

seems important will be outweighed by an unbounded number of initially unimportant-seeming matters, something that initially seems unimportant will unduly shape the outcome, or both. Thus such seemingly-disparate axiologies as maximin and classical total utilitarianism have in common that they are both prepared to accept the cost of many arbitrarily negative lives and forgo the benefits of many arbitrarily positive lives, for the right arrangement of infinitesimal tweaks.

9 Conclusion

The results above reveal the full extent of the repugnant conclusion: no leading axiology can avoid the Extended Very Repugnant Conclusion, which is just as repugnant as the Very Repugnant Conclusion. Theorem 1 shows that the VRC cannot be avoided by any leading welfarist axiology despite prior consensus in the literature to the contrary, and Theorem 3 shows that the XVRC cannot be avoided by any other welfarist axiology in the literature. Together with the results in section 7, this shows that every plausible axiology has repugnant implications, and thus that the repugnant conclusion does not tell against any axiology. In light of these results, the idea that the repugnant conclusion must be avoided cannot remain the leading methodological principle in population axiology. If repugnance is unavoidable, then we should not try to avoid it.

The fundamental explanation of these results is that all axiologies have repugnant consequences.[24] This complements the work of others who have noted unintuitive consequences of axiologies over an unbounded domain in same-number cases (Cowen, 1996; Norcross, 1997, 1998; Arrhenius, 2000; Fleurbaey and Tungodden, 2010; Arrhenius, n.d.; Bossert, 2017). We interpret these papers as contributing to the same general insight as our own: repugnant implications are

[24]Fleurbaey and Tungodden (2010) suggest a similar conclusion, studying same-number problems, outside of population ethics: "we believe that one should be cautious when criticizing maximin, (generalized) utilitarianism or any other social ordering on the basis of how they perform in extreme cases. The assessment of the various possible social ordering functions should be more comprehensive and, maybe, more focused on cases that are directly relevant to actual policy issues" (p. 410).

an inevitable feature of any plausible axiology. If repugnance cannot be avoided, then it should not be. We believe this should be among the guiding insights for the next generation of work in value theory.[25]

A Additional results

A.1 Non-algebraic version of Theorem 1

Here we present a version of Theorem 1 which uses statements of conditions, rather than the algebraic functional form of W, to describe the axiologies of interest. Following Blackorby et al. (2005), in this appendix we use notation $u\mathbf{1}_n$ to mean a population of n people, each with utility u; the symbol $\sim$ means "exactly as good as." A key characteristic of the W functional form is convergence in equivalence:

> **Convergence in equivalence.** Let P be any base population. Let u be any utility level and $\delta > 0$ any small difference. Then there exists some number n, which may depend on P, δ, and u, such that, for all $n' > n$, there exists v such that $v\mathbf{1}_{|P|+n'} \sim (P \cup u\mathbf{1}_{n'})$ and $|u - v| < \delta$.

In other words, if you add enough lives of utility u to any base population, eventually it becomes similar in social value to a same-size population in which each life is of value arbitrarily close to u. This is an implication of same-number generalized utilitarianism, but may also be satisfied by other axiologies. Notice that $v\mathbf{1}_{|P|+n'}$ is constructed to be the same size as $P \cup u\mathbf{1}_{n'}$, so this is a same-number comparison (a weakening of same-number generalized utilitarianism).

Maximin does not satisfy convergence in equivalence, because it only attends to the worst-off person. For the same reason, RDGU $\left(\text{a generalized maximin discussed in section 5: } \sum_i \beta^{i-1} u_i\right)$ also does not satisfy convergence in equivalence. Consider $\beta = 0.5$, a base population consisting of one life of value 0, and $u = 1$. As many u lives are added, the social value converges to 1, which would be equivalent, in the limit, to a population in which everybody had utility 0.5, not 1. In other words, a population with a billion-and-one people with utility 0.51 would be better

[25] We do not draw conclusions about what axiology results from dropping the axiomatic requirement to avoid the repugnant conclusion. Anglin (1977), Tännsjö (2002), Huemer (2008), and more argue that the result is a form of total utilitarianism; see also Broome (2004) and Bykvist (2007). However, a reader could accept our conclusion without accepting totalism, for example by choosing an alternative axiology as illustrated by Table 2, which highlights some of the available axiologies in a welfarist framework depending on choices about the shape and degree of curvature of f, g, and h. We also do not consider the further question whether it is *regrettable* that repugnance is inevitable.

than a population with one person with utility 0 and a billion people with utility 1.

For theorem 1*, what we actually require is something weaker than convergence in equivalence:

> **Convergence in signs**. Let P be any base population. Let u be any utility level. Then, for all n, there exists $\tilde{n} > n$, such that there exists v such that $v\mathbf{1}_{|P|+\tilde{n}} \sim (P \cup u\mathbf{1}_{\tilde{n}})$ and v is positive if and only if u is positive.

In other words, if you add enough lives worth living (or not living) to any base population, the result is eventually, at least for some population size larger than any finite population size, a population that is overall just as good as a perfectly equal population in which each person has a utility with the same sign as the added life. We interpret convergence in signs to be a property of population axiologies that are minimally willing to make tradeoffs among people. Totalism, averagism, and every form of W satisifies convergence in signs. However, RDGU also fails this condition. With $\beta = 0.5$, a population with a trillion-and-one people with utility -0.2 would be better than a population with one person with utility -0.75 and a trillion people with utility 0.25. Better than either of these, according to such a social welfare function, would be for nobody to exist at all.

A final condition that could be used in place of extended egalitarian dominance is priority for lives worth living (compare Carlson, 1998; Blackorby et al., 2005):

> **Priority for lives worth living**. Any population containing only people each with positive utility is better than any population containing only people each without positive utility.

Note that neither priority for lives worth living nor extended egalitarian dominance implies the other. RDGU satisifies priority for lives worth living, but CLGU does not.

> **Theorem 1***. Any transitive population axiology that satisfies *convergence in signs* and either *extended egalitarian dominance* or *priority for lives worth living* implies the very repugnant conclusion.

> *Proof.* Choose a base population with negative utility, large enough that the base population with the very good lives is equivalent to a same-size perfectly equal population with negative utility, by convergence in signs. Also by convergence in signs, choose n^{ε} large enough that the base population with the very bad lives and with the $\varepsilon > 0$ lives both is equivalent to a same-size perfectly equal population with positive utility (like ε) and is larger than the base population with the very good lives. Then apply either extended egalitarian dominance or priority for lives worth living on the equivalent populations, and the result follows with transitivity. □

Notice that this proof does not require completeness of the axiology. Utility levels need not be real numbers and need not be continuously divisible. What is required is utility levels that are ranked and signed, where signs are (positive, non-positive), and that there is at least one positive and one non-positive level of utility. Signs are used by both extended egalitarian dominance and convergence-in-signs. Note that without a notion of "positive," a small-positive-value cannot be stated, and neither can the repugnant conclusion.

A.2 Lexical Views, Non-Archimedeanism, and categories of utility

Several population axiologists have proposed a response to the RC which separates utility into categories that cannot be traded-off (*e.g.* Carlson, 2017).[26] For example, there may be lower pleasures and higher pleasures, such that no increment of lower pleasures is as good as any small increment of higher pleasures. Parfit (1984), for example, considers a version of this that he calls the "Lexical View," and concludes that it, too, has repugnant implications: "These conclusions are less repugnant and less absurd. But they are both implausible," (see also section 6.2 of Arrhenius, n.d.). However, Parfit does not test the Lexical View with conclusions as repugnant as our VRC and Extended VRC. This appendix, building upon Arrhenius and others, does so for two possible versions. Section A.2.1 considers the possibility that better-off people are qualitatively different; section A.2.2 considers the possibility that higher pleasures are qualitatively different. In general, because lexical views still must aggregate across people, they remain subject to repugnance.

A.2.1 Different categories of people: The better-off are different

Each person is described by a utility quantity u_i. However there is a threshold $\tau > 0$: utility levels above the threshold ($u_i > \tau$) are lives that are qualitatively better than lives below the threshold. Lives, then, are sorted into two categories: those above the threshold and those below.

What this view has not yet specified is how utility is aggregated across people within these categories. But there must be some aggregation. Let W^τ be aggregate welfare among people above the threshold and W_τ be aggregate welfare below the threshold, for any of the functional forms W (or, for extensions including RDGU and maximin, within the categories). To preserve the lexical structure, it must be that the social ordering is increasing in W^τ, with W_τ breaking ties (otherwise, increases in W_τ could compensate for decreases in W^τ).

[26] Arrhenius (2005) offers compelling arguments to doubt that such superiority exists, because of the possibility to pollute an excellent life with marginal degrees of pain. See also Ryberg (1996b).

Now consider the choice offered by the Extended Very Repugnant Conclusion. Let the base population lives be above τ, and let the ε-changes be to those lives. Then, enough ε-changes to enough base lives will increase total and average utility above τ, and will therefore increase W^{τ} (to simultaneously increase the number of people living above τ, use ε-changes to move lives above τ). Because W^{τ} increases, the option with ε-changes is better, no matter how much the u^{ℓ} lives reduce W_{τ} (moreover, by also having negative-utility base-population lives, an option can be constructed that increases average and total utility among the lives below τ). So, average utility, total utility, and any W increase among both lives above and lives below τ, and the number of people living above τ increases, so the option with very negative lives and ε-changes is chosen. This conclusion is the Extended VRC.

A.2.2 Different categories of well-being: Higher versus lower pleasures

Consider a ranked set of utility dimensions $d \in D$. Dimensions are different categories of pleasure, and no amount of a lower pleasure is worth as much to a person's well-being as any amount of a higher pleasure. Then, person i's well-being is described by a vector (u_i^d). Lower numbers d are lexically more important categories. These properties are sufficient to rank populations consisting of only one person: rank by the first dimension, then break ties with the second, and so on. However, these properties do not speak to how aggregation is done *across people*, which is the question at the heart of population axiology. In other words, merely proposing a lexical understanding of well-being has not yet proposed a population axiology.

To preserve the lexical structure, dimensions of well-being must first be aggregated across people, within dimensions; then dimensions can be aggregated lexically. Therefore, within any category, including the first category, a lexical view faces the same sort of uni-dimensional aggregation of person-specific utility scalars as do non-lexical views (such as in Theorem 1). A lexical view could have within-dimension sub-principles of any form W (totalist, averagist, X'), or any other form in the paper. But all of these, as the theorems show, imply repugnant conclusions. Therefore, all lexical views of this dimensional type, however they aggregate within-dimensions, are subject to within-dimension repugnant conclusions, such as in Theorem 3*'s weak same-number, non-transitive condition, and each of the rest.

In a lexical repugnant conclusion, ε would represent a tiny, perhaps imperceptible increment of the best category. For example, for ε close to zero, and for three dimensions, we can imagine $(\varepsilon, 0, 0)$. Like in the single-dimension repugnant conclusion, ε could be arbitrarily small; like in the single-dimension repugnant conclusion, however small it is there are an infinity of cases between it and zero: $(\frac{\varepsilon}{n}, 0, 0)$, for all the natural numbers n.

To Parfit, this might represent listening to Mozart for a millisecond longer before it is replaced with Haydn. Another possibility is a "roller coaster" life in which the highest pleasures barely outweigh the highest suffering. Several authors have observed that small quantities of excellent goods is one way to create the infinitesimal, ε utility increments that repugnant conclusions require (Tännsjö, 2002; Huemer, 2008). Portmore (1999) emphasizes a short duration: "lives which are qualitatively identical to those in [the high-quality world] but very short-lived." Such tiny differences in excellence are imaginable (our lives, even if excellent, might seem only ε long to imaginable long-lived aliens), and therefore they are within the domain of repugnant conclusions.

A.3 Additional repugnant conclusion for average utilitarianism

This appendix provides another example of how even in the different number cases characteristic of population axiology, averagism other non-totalist views can be exposed to repugnant implications over an unbounded space. Further examples exist in the literature (Anglin, 1977). For example, Hurka (1982) presents a version of a repugnant conclusion for time-period-specific average utilitarianism, which Greaves (2017) calls "time-integrated instantaneous averagism."

> **Another repugnant conclusion for generalized averagism**. For any opportunity to bring 10 billion people (who exist under either option) to a very high level of wellbeing, there is an imaginable large number of extremely brief lives with no suffering, but with very low positive value that would — if additionally created, in a distant corner of the universe, as a further consequence of the improvement — cause an average utilitarian to forgo the option of improving the 10 billion lives.

Proof. Let u^h_{10b} and u^ℓ_{10b} be positive real numbers, representing the average well-being of the 10 billion people with and without the improvement, respectively. Let n^0 and u^0 (both positive real numbers) be the number and average well-being of the unaffected population, who would otherwise exist. Finally let $\epsilon > 0$ be the tiny positive average well-being and n^ϵ the quantity of new ϵ-people created.

We can state Repugnant Consequences 7 formally as: for any improvement from u^ℓ_{10b} to u^h_{10b}, there exists a threshold of new people $n^{\epsilon *}$ and a small positive ϵ such that if $n^\epsilon > n^{\epsilon *}$ then an average utilitarian would not prefer the addition-and-improvement:

$$\frac{10^{10} \times u^\ell_{10b} + n^0 \times u^0}{10^{10} + n^0} > \frac{10^{10} \times u^h_{10b} + n^0 \times u^0 + n^\epsilon \times \epsilon}{10^{10} + n^0 + n^\epsilon}.$$

This condition will be true if

$$n^{\epsilon} > n^{\epsilon *} \equiv \frac{10^{10}\left(10^{10}+n^{0}\right)\left(u_{10b}^{h}-u_{10b}^{\ell}\right)}{10^{10}\times u_{10b}^{\ell}+n^{0}\times u^{0}-\epsilon\left(10^{9}+n^{0}\right)},$$

which must be positive and bounded above by the numerator because $\left(u_{10b}^{h}-u_{10b}^{\ell}\right)$ is positive and ϵ can be chosen to be arbitrarily close to zero. □

Notice the points of similarity with the repugnant conclusion for total utilitarians: an opportunity to make ten billion people arbitrarily better off is forgone because of the unbounded possibility of creating many, many low-value lives, even though nobody in the scenario ever has a life of suffering or a life not (at least barely) worth living, and even though the many lives need never exist.

References

Adler, Matthew D (2009) "Future Generations: A Prioritarian View," *George Washington Law Review*, Vol. 77, p. 1478.

Anglin, Bill (1977) "The repugnant conclusion," *Canadian Journal of Philosophy*, Vol. 7, pp. 745–754.

Arrhenius, Gustaf (2000) "An impossibility theorem for welfarist axiologies," *Economics & Philosophy*, Vol. 16, pp. 247–266.

——— (2003) "The Very Repugnant Conclusion," *Uppsala Philosophical Studies*, Vol. 51.

——— (2005) "Superiority in value," *Philosophical Studies*, Vol. 123, pp. 97–114.

——— (2009) "One more axiological impossibility theorem," *Logic, Ethics, and All That Jazz. Uppsala Philosophical Studies*, Vol. 57.

——— (2015) "Existential Question and the Person Affecting Restriction," *Weighing and Reasoning: Themes from the Philosophy of John Broome*, p. 110.

——— (n.d.) *Population Ethics: The Challege of Future Generations*: unpublished typescript (July 2017).

Arrhenius, Gustaf and H. Orri Stefánsson (2018) "Population ethics under risk," working paper, Institute for Futures Studies.

Asheim, Geir B and Stéphane Zuber (2014) "Escaping the repugnant conclusion: Rank-discounted utilitarianism with variable population," *Theoretical Economics*, Vol. 9, pp. 629–650.

Blackorby, Charles, Walter Bossert, and David Donaldson (1998) "Uncertainty and critical-level population principles," *Journal of Population Economics*, Vol. 11, pp. 1–20.

Blackorby, Charles, Walter Bossert, and David J Donaldson (2005) *Population issues in social choice theory, welfare economics, and ethics*, No. 39: Cambridge University Press.

Blackorby, Charles and David Donaldson (1984) "Social criteria for evaluating population change," *Journal of Public Economics*, Vol. 25, pp. 13–33.

Bossert, Walter (2017) "Anonymous welfarism, critical-level principles, and the repugnant and sadistic conclusions," working paper, University of Montréal.

Broome, John (2004) "Weighing lives," *Oxford*.

Bykvist, Krister (2007) "The good, the bad, and the ethically neutral," *Economics & Philosophy*, Vol. 23, pp. 97–105.

Carlson, Erik (1998) "Mere addition and two trilemmas of population ethics," *Economics & Philosophy*, Vol. 14, pp. 283–306.

——— (2017) "On Some Impossibility Theorems in Population Ethics," working paper, Uppsala University.

Chang, Ruth (2016) "Parity, imprecise comparability and the repugnant conclusion," *Theoria*, Vol. 82, pp. 182–214.

Cowen, Tyler (1996) "What do we Learn from the Repugnant Conclusion?" *Ethics*, Vol. 106, pp. 754–775.

Fleurbaey, Marc and Bertil Tungodden (2010) "The tyranny of non-aggregation versus the tyranny of aggregation in social choices: a real dilemma," *Economic Theory*, Vol. 44, pp. 399–414.

Fleurbaey, Marc and Stéphane Zuber (2015) "Discounting, beyond utilitarianism," *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 9, pp. 1–52.

Frick, Johann (2018) "Conditional Reasons and the Procreation Asymmetry," working paper, Princeton.

Greaves, Hilary (2017) "Population axiology," *Philosophy Compass*, Vol. 12.

Greaves, Hilary and Toby Ord (2017) "Moral uncertainty about population axiology," *Journal of Ethics and Social Philosophy*, Vol. 12, pp. 135–167.

Gustafsson, Johan E. (2018) "Our Intuitive Grasp of the Repugnant Conclusion," working paper, IFFS.

Holtug, Nils (2004) "Person-affecting moralities," in *The Repugnant Conclusion*: Springer, pp. 129–161.

Huemer, Michael (2008) "In defence of repugnance," *Mind*, Vol. 117, pp. 899–933.

Hurka, Thomas (1982) "Average utilitarianisms," *Analysis*, Vol. 42, pp. 65–69.

——— (1983) "Value and population size," *Ethics*, Vol. 93, pp. 496–507.

McMahan, Jeff (1981) "Problems of Population Theory," *Ethics*, Vol. 92, pp. 96–127.

Nebel, Jacob M. (forthcoming) "An intrapersonal addition paradox," *Ethics*.

Ng, Yew-Kwang (1989) "What Should We Do About Future Generations?: Impossibility of Parfit's Theory X," *Economics & Philosophy*, Vol. 5, pp. 235–253.

Norcross, Alastair (1997) "Comparing harms: headaches and human lives," *Philosophy & Public Affairs*, Vol. 26, pp. 135–167.

——— (1998) "Great harms from small benefits grow: how death can be outweighed by headaches," *Analysis*, Vol. 58, pp. 152–158.

Nozick, Robert (1974) *Anarchy, State, and Utopia*, New York: Basic Books.

Parfit, Derek (1984) *Reasons and Persons*: Oxford.

Portmore, Douglas W (1999) "Does the total principle have any repugnant implications?" *Ratio*, Vol. 12, pp. 80–98.

Pummer, Theron (2018) "Spectrum arguments and hypersensitivity," *Philosophical Studies*, Vol. 175, pp. 1729–1744.

Roberts, Melinda A. (2015) "Population axiology," *The Oxford Handbook of Value Theory*, pp. 399–423.

——— (2018a) "The Better Chance Puzzle and the Value of Existence: A Defense of Person-Based Consequentialism," working paper, The College of New Jersey.

——— (2018b) "Modal Ethics and Moral Value," working paper, The College of New Jersey.

Ryberg, Jesper (1996a) "Is the repugnant conclusion repugnant?" *Philosophical Papers*, Vol. 25, pp. 161–177.

——— (1996b) "Parfit's repugnant conclusion," *The Philosophical Quarterly*, Vol. 46, pp. 202–213.

Sider, Theodore R (1991) "Might theory X be a theory of diminishing marginal value?" *Analysis*, Vol. 51, pp. 265–271.

Tännsjö, Torbjörn (2002) "Why we ought to accept the repugnant conclusion," *Utilitas*, Vol. 14, pp. 339–359.

Temkin, Larry (2012) *Rethinking the Good*: Oxford University Press.

Voorhoeve, Alex and Marc Fleurbaey (2016) "Priority or equality for possible people?" *Ethics*, Vol. 126, pp. 929–954.