Economics & Philosophy (2026), **42**, 645–673
doi:10.1017/S0266267125100631

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Repugnance without Mere Addition

Tomi Francis

Université de Fribourg, Fribourg, Switzerland
Email: t.m.francis123@googlemail.com

(Received 11 June 2023; revised 11 May 2025; accepted 23 July 2025; first published online 15 January 2026)

Abstract

This paper concerns the difficulty of avoiding an additive version of the Very Repugnant Conclusion. An impossibility theorem is provided which shows that we cannot avoid this version of the Repugnant Conclusion even if we deny the Mere Addition Principle and closely related principles which place limits on the badness of adding happy people, such as "Dominance Addition" and additive "Non-Sadism" conditions. I argue that the impossibility theorem shows that the additive version of the Very Repugnant Conclusion cannot reasonably be avoided by population-ethical means alone. One must instead either deny structural conditions such as acyclicity, adopt a radically unorthodox fixed-population axiology, or accept this version of the Very Repugnant Conclusion.

Keywords: Population ethics; population axiology; repugnant conclusion; mere addition; impossibility theorem

1. Introduction

Most people are inclined to deny Derek Parfit's

> **Repugnant Conclusion** For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living (Parfit 1984: 388).

Doing so in a plausible way, however, has turned out to be rather difficult. Some of the difficulties have been evident from the beginning: along with the statement of the Repugnant Conclusion, Parfit provided his celebrated Mere Addition Paradox.[1] Since then, philosophers and economists have contributed yet stronger arguments and impossibility theorems, revealing the problem of avoiding the Repugnant Conclusion to be even harder than Parfit first thought.[2] In this paper, I provide

[1]See Parfit (1984: 419–442).

[2]See Ng (1989), Carlson (1998), Arrhenius (2000, 2003, 2009, 2011, n.d.), Blackorby *et al.* (2003, 2005), Nebel (2019) and Spears and Budolfson (2021).

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

another argument in this vein. For ease of reference, let's call it the Additive Impossibility Result.

Parfit's Mere Addition Paradox relies heavily on the Mere Addition Principle, which says that additions of good lives to the world cannot make an outcome worse. This principle is plausible, but questionable. The main point of the Additive Impossibility Result is to do without it and its close cousins, such as the "Weak Non-Sadism Condition" appearing in Gustaf Arrhenius's favoured Sixth Impossibility Theorem. That condition says that additions of good lives (no matter how many) cannot make an outcome worse than adding a large number of very bad lives.[3]

Instead, the Additive Impossibility Result appeals to what I call the *Non-Additive Weak Non-Sadism Condition*. This principle is non-additive in the following sense: while the Weak Non-Sadism Condition holds that *adding* a large number of bad lives to any unaffected background population always results in a worse outcome than adding any number of lives worth living, the Non-Additive Weak Non-Sadism Condition only requires that a large population of bad lives would be worse *by itself* than any number of lives (perhaps barely) worth living.

Arrhenius's Sixth Theorem uses the additive Weak Non-Sadism Condition to derive a non-additive version of the Repugnant Conclusion.[4] The Additive Impossibility Result switches this around. It uses the Non-Additive Weak Non-Sadism Condition to derive an additive version of the Repugnant Conclusion, which says (simplifying things a little) that *adding* a large number of people with lives barely worth living can be better than adding a smaller number of people with a very high quality of life.[5]

This switch has a philosophical payoff. As I shall argue in §2, Weak Non-Sadism might fail if we take seriously the idea that additions of good lives can make an otherwise excellent population much worse: that is, that the Mere Addition Principle is false in such a way that, by rejecting it, we can avoid the original Mere Addition Paradox. (Of course, one might just take this to illustrate the high cost of

[3]See Arrhenius (2009, 2011, n.d.).

[4]Strictly speaking, the Repugnant Conclusion avoidance condition used in Arrhenius's sixth theorem is additive, but in a weaker sense: the size of the added population of excellent lives can depend on the choice of unaffected background population. However, Arrhenius's sixth theorem can be modified to incorporate a non-additive avoidance condition for the Repugnant Conclusion, provided we strengthen the General Non-Extreme Priority condition in a way that I believe does not significantly affect its plausibility. This modified theorem, which I call the Non-Additive Impossibility Result, is proved in the Appendix.

[5]Spears and Budolfson (2021) have also provided an impossibility theorem which, like the Additive Impossibility Result, derives an additive version of the Repugnant Conclusion from a principle like the Non-Additive Weak Non-Sadism Condition and other premises. However, Spears and Budolfson make structural assumptions which, while standard in the welfare economics literature, are controversial for population axiology. These include completeness, transitivity, option set independence, anonymity, and the representability of wellbeing levels by real numbers. The Additive Impossibility Result makes none of these assumptions (but does assume acyclicity in place of transitivity). Incompleteness in the context of population ethics is defended by many authors, including Blackorby *et al.* (1996), Qizilbash (2007a, b, 2018), Rabinowicz (2009), Chang (2016), Parfit (2016), Gustafsson (2020) and Nebel (2022). Anonymity is less controversial, though some authors reject it, including Temkin (1987, 2012) and Roberts (2011). Intransitivity is defended by Temkin (1987, 1996, 2012) and Rachels (1998, 2001, 2004). Option-set-dependent betterness is defended by Frick (2014, 2022) and, on some readings, by Temkin (2012) (according to Cusbert 2017).

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

denying the Mere Addition Principle.) Weak Non-Sadism is compelling, but the *improvement* it offers over the Mere Addition Principle is, I shall suggest, smaller than it first seems.

In contrast, the Non-Additive Weak Non-Sadism Condition is compelling regardless of whether the Mere Addition Principle is true. Moreover, it is *more* plausible than avoidance of the additive version of the Repugnant Conclusion. Yet the additive version of the Repugnant Conclusion still seems "repugnant" in the same sort of way as the non-additive version. Consider a future for humanity in which our descendants live excellent lives, contrasted with a drab future in which a much larger population has lives barely worth living. It is "repugnant" for the drab future to be better. Suppose now that regardless of which future we choose, some (perhaps large) number of alien persons will also come to exist in a causally isolated part of the universe. Does the presence of these aliens change our intuition that the "drab future" would be worse? I think not. If the drab future is better overall, given certain facts about the numbers and wellbeing levels of the unaffected aliens, this is an instance of the "additive" Repugnant Conclusion.

The Additive Impossibility Result therefore answers the question of whether a reasonable solution to the problem of avoiding the Repugnant Conclusion can be confined to the variable-population case. The answer is: No. If the Repugnant Conclusion is to be avoided in a reasonable way, one must either reject certain plausible principles for comparing same-person populations, or reject foundational structural assumptions such as Acyclicity or what I shall call "Finite Fine-Grainedness".

2. Mere Addition and Non-Sadism

Let us begin by recalling Parfit's Mere Addition Paradox (1984: Ch. 19). The principle at its centre is, unsurprisingly, the

> **Mere Addition Principle** If X is a population, and Y is a population consisting solely of good lives, then $X + Y$ is not worse than X.

The argument runs as follows. We begin with an arbitrary population A, consisting of many excellent lives. We then add an extremely large number of lives barely worth living to A; call the resulting population A^+. By the Mere Addition Principle, A^+ is not worse than A. By repeatedly applying plausible principles for comparing same-person populations, we find that A^+ is worse than some population Z, in which all of the A^+ people have a life which is barely worth living, and just a little better than the additional lives in A^+. (The gains to the A people in population A^+ are outweighed by the losses to the remaining people.) By transitivity, Z is not worse than A. This is a version of the Repugnant Conclusion.

How should we respond to the Mere Addition Paradox? One attractive option is to simply deny the Mere Addition Principle. While it seems implausible that additions of good lives could make the world worse, we might find the Repugnant Conclusion still more implausible. One significant advantage to this response to the

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

Mere Addition Paradox is that it confines our population-ethical difficulties to variable-population cases.[6]

To be sure, it is hard to see *why* the Mere Addition Principle should be false. But perhaps some explanation could be offered. One might, for instance, appeal to the badness of inequality in order to argue that A^+ is worse than A. Or perhaps we could say that the move from A to A^+ makes things worse in a holistic sense, as we would move from a situation in which everyone has an excellent life to one in which almost everyone has a life barely worth living. While it is difficult to account for judgements like these in person-affecting terms (since a mere addition isn't bad for anyone), intrapersonal arguments for the Repugnant Conclusion provide another line of evidence that we might need to appeal to impersonal considerations to avoid the Repugnant Conclusion.[7]

Later impossibility theorems are widely thought to close the door on the possibility of escaping the Repugnant Conclusion by denying the Mere Addition Principle. In particular, Gustaf Arrhenius has shown that the Mere Addition Principle can be replaced by other, supposedly more compelling, different-number principles. These are

Dominance Addition If A is a population, and B is a population consisting only of good lives, and A^+ consists of the A-people with higher wellbeing levels than they enjoy in A, then

$$A \preceq A^+ + B$$

Non-Sadism If A is a population consisting of good lives, and B is a population consisting of bad lives, then for any unaffected background population I,

$$I + B \prec I + A$$

Weak Non-Sadism There is some bad wellbeing level b, and some number of lives at this level, such that if B is a population consisting of at least this number of lives at level b, or at some worse level, and A is any population consisting of good lives, and I is any unaffected background population, then

$$I + B \prec I + A$$

However, these conditions are not particularly compelling *conditional* on the Mere Addition Principle being false in the way it would need to be for us to avoid the Mere Addition Paradox. Consider first the principle of Dominance Addition. This says that adding good lives while at the same time making existing people better off results in an outcome which is at least as good. But if Mere Addition is false, adding good lives can be bad. While it is clearly a good thing for existing people to be better off, there is no obvious reason to expect that the good thing must always outweigh the bad thing. There is thus no real reason to expect that it cannot be worse for existing people to be made better off, and additional good lives to be added at the

[6]We obviously cannot say the same about denying plausible same-person principles, and it also seems unlikely on its face that intransitivity could be confined to variable-population cases.

[7]See Nebel (2019: 342–343) and Francis (2023: 107).

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

same time. This claim is plausible, but only because the Mere Addition Principle is plausible. Dominance Addition is plausible *because* adding good lives does not appear to be bad.

The same point applies to Non-Sadism. Although it would clearly be a bad thing for an additional bad life to be added to a population, there is not much reason to expect that this must always be worse than an addition of good lives, if we think that an addition of good lives can also be a bad thing.[8] Again: Non-Sadism is plausible *because* adding good lives does not appear to be bad (that is: it is plausible *because* the Mere Addition Principle seems to be true).

Consider finally the Weak Non-Sadism condition. This principle seems more compelling than Non-Sadism because even if adding good lives could be bad, surely it could not be *so bad* as to be worse than an addition of a large number of very bad lives. The problem here, however, is that if we get out of the Mere Addition Paradox by denying the Mere Addition Principle, we must think that adding good lives can be very bad indeed, in the right circumstances.

To see this, let B be some population which is so bad that, by Weak Non-Sadism, adding it is worse than adding any population of good lives. We can choose, on the basis of B, a population A in which every person is many times better off than the people in B are badly off, and which contains vastly more people. Those who wish to deny Mere Addition to avoid the Repugnant Conclusion will think that adding a suitably large population of barely good lives Z^- to A can render the latter population worse than Z, a population consisting of lives barely worth living. But is it obvious that $A + B$ is likewise worse than Z? It is not. While it would be a tragedy for the B lives to exist, there could be trillions of excellent A lives for every B life. Such a population might plausibly be better (or at least not worse) than any population consisting only of lives barely worth living. If that is right, then, by transitivity, $A + B$ must not be worse than $A + Z^-$. And this is exactly what it takes for Weak Non-Sadism to be false.

The lesson to draw here is that the plausibility of the Dominance, Non-Sadism and Weak Non-Sadism conditions is closely tied to the plausibility of the Mere Addition Principle itself. The impossibility theorems involving these conditions bring out the costs of this way of avoiding the Repugnant Conclusion, but they do not show that it is unacceptable. To close off this route completely, one would need to provide an impossibility theorem which replaces the Mere Addition Principle with a principle which remains compelling even on the assumption that the Mere Addition Principle is false. Ideally, this replacement should also be clearly more plausible than avoidance of the Repugnant Conclusion. This can indeed be done, as we shall now see.

[8]Some population axiologies, such as Average Utilitarianism, say that it can be better to add bad lives than not to add any lives. This is implausible, but as Carlson (1998: 302–304) notes, axiologies which violate Non-Sadism need not imply this implausible claim. As a toy example, consider the view that ranks populations with negative total wellbeing according to their total wellbeing levels, but ranks populations with non-negative total wellbeing according to their average wellbeing levels (with non-negative wellbeing populations ranked above negative ones).

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

3. The Additive Impossibility Result

3.1 Framework: Wellbeing and Populations

We assume that there are infinitely many possible people, and that we have at our disposal some set of wellbeing levels.[9] A population is any logically possible assignment of finitely many people to wellbeing levels. If X is a set of possible persons (hereafter a "group") $X[w]$ denotes the population where the X people exist at wellbeing level w and nobody else exists. When populations (or groups) X and Y are disjoint, $X + Y$ denotes their set theoretic union: the population (group) consisting of the X people and Y people, who are (in the case of populations) at their respective wellbeing levels in X and Y.[10] An option set is any finite non-empty set of populations; let $\mathcal{C}$ be the set of all option sets. A *population axiology* $\succeq$ is a three-place relation on $P \times P \times \mathcal{C}$, where P is the set of all possible populations, and $\succeq$ is reflexive in the sense that we have $\succeq (X, X, C)$ for any population X and any option set C. Other relations, such as $\succ$, and $\prec$, are defined from $\succeq$ in the standard way. I shall often use "at-least-as-good-as", "better" and so on to refer to a population axiology and its derived relations.

We will need to make some fairly minimal assumptions about the structure of wellbeing. We will need to assume the existence of a prudential at-least-as-good-as relation on wellbeing levels. I shall denote this relation by $\geq$, with other symbols standing for derived relations in the obvious way, and "betterness" talk referring to the relevant derived relations. The relation $\geq$ is assumed to be transitive and reflexive.[11] For technical reasons, we need to assume the Directedness Property: for any wellbeing levels w_1 and w_2, there exists some w_3 such that $w_3 \geq w_1, w_2$ and some w_4 such that $w_4 \leq w_1, w_2$.[12] The Directedness Property is weaker and less controversial than the completeness requirement (which we shall not assume), according to which for any wellbeing levels and w', either $w \geq w'$ or $w \leq w'$.

We shall assume that lives and wellbeing levels can be categorized as being *good* or as *bad*. These correspond to the levels of a life worth living and of a life worth not living respectively. Good lives are better than bad ones. A *neutral* life (or wellbeing level) is one which is neither good nor bad.[13] A *barely good* life is one at a wellbeing level which is good, and which is close to not being good.

[9]I will not discuss populations involving non-human animals in this paper.

[10]Any principles which contain this notation should be taken to implicitly quantify only over disjoint populations when the unions of such populations are taken; I shall generally omit such qualifications in order to improve readability.

[11]One might worry that the assumption of transitivity here undermines the motivation to drop it as a requirement for population axiology, since authors who reject transitivity as a requirement on population axiology usually reject it as a requirement on prudential betterness (e.g. Rachels 1998; Temkin 2012). We can get around this worry by restricting our attention to a transitive subset of the set of all wellbeing levels.

[12]Alternatively, as in the case of transitivity, it is sufficient to restrict our attention to an upwards- and downwards-directed subset of the set of all wellbeing levels.

[13]The category of neutral lives, as I have defined it, includes what Gustafsson (2020) calls "undistinguished" lives, which are neither good nor bad, yet may not be close to being bad. Gustafsson thinks that repugnant conclusions are not repugnant when they involve lives that are "barely good" in my sense, but are not close to being bad (instead being nearly undistinguished). I disagree, but it would take us too far afield to treat this issue thoroughly.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

3.2 Finite Fine-Grainedness

We shall take "closeness" to be a primitive binary relation on wellbeing levels. Its intended interpretation is exactly what it sounds like: two wellbeing levels are "close" if they are close together, on some appropriate specification of this notion. For instance, we might say that w and w' are close if the addition or omission of a few pinpricks of pain can make the difference between w being better or worse than w'. The notion of closeness is crucial to our first assumption about the structure of wellbeing, namely

> **Finite Fine-Grainedness** For any wellbeing levels $w > u$, there exists a finite chain of wellbeing levels $w = w_0 > w_1 > \ldots > w_n = u$ such that each w_i is close to w_{i+1}.

Thomas (2018) argues convincingly that Finite Fine-Grainedness is not a principle which is *automatically* true of the structure of wellbeing. *Pace* Arrhenius (2009, 2011), mathematical models purporting to represent the structure of wellbeing can violate Finite Fine-Grainedness. But mathematical possibility is one thing, and philosophical plausibility is another. It seems to me that no such model succeeds in faithfully representing the structure of wellbeing.

Here's why.[14] Suppose that we take any life, and either make an existing second of it slightly more painful, or extend the life by a second. Both modifications result in only a small difference in wellbeing. Yet sufficiently many such modifications can turn any arbitrarily good life into an arbitrarily long life of constant, agonizing torture. Given, as seems plausible, that any finite life is better than a sufficiently long life of agonizing torture, for any lives $w > u$ we can find a finite consecutively close chain from w towards some life w_n which is worse than u. This argument, if it succeeds, justifies Finite Fine-Grainedness.[15]

One might resist arguments of this sort.[16] Erik Carlson (2022: 215–216) suggests that small additions of pain might correspond to large differences in wellbeing. He argues that adding or worsening pain episodes might bring about a threshold or holistic effect, whereby a slight amount of added pain corresponds to a sharp drop in wellbeing. I can see two reasons why this might be the case. First, certain levels of pain might be incompatible with certain important goods. A small increase in pain may therefore remove such goods from a person's life, thereby indirectly leading to a large loss of wellbeing. Second, the badness of pain itself may increase in a greater-

[14]A similar argument is offered by Arrhenius (2016: 171–172). Another similar argument is suggested by Thomas (2018).

[15]Strictly speaking, this argument supports only the weaker principle of **Weak Finite Fine-Grainedness**. For any wellbeing levels $w > u$, there exists a finite chain of wellbeing levels $w = w_0 > w_1 > \ldots > w_n$ such that each w_i is close to w_{i+1}, and w_n is worse than u. But it is hard to see how Finite Fine-Grainedness could be false if Weak Finite Fine-Grainedness is true. In any case, if the objection that Finite Fine-Grainedness is logically stronger were to be pressed, the problem could be avoided in another way: we could simply restrict our attention to a chain of wellbeing levels whose consecutive members differ only by pinpricks. Given Weak Finite Fine-Grainedness, such a set can span between any arbitrarily good wellbeing level and any arbitrarily bad wellbeing level, and so the premises of the Additive Impossibility Result will remain compelling when our attention is restricted in this way.

[16]I am grateful to an anonymous reviewer for pressing this point.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

than-linear fashion, so that n pain episodes may sometimes be more than n times as bad as a single episode.[17]

I am inclined to think, however, that neither of these considerations should lead us to reject Finite Fine-Grainedness.[18] While it strikes me as plausible that it is not possible for a person to enjoy certain important goods like knowledge or friendship during a sufficiently intense episode of pain, I doubt that these goods disappear in a discontinuous rather than continuous way. If agonizing pain is distracting enough to completely overpower one's capacity to enjoy the good of friendship, then surely slightly-less-agonizing pain will mostly overpower it. Similarly, while the badness of pain *might* increase in a greater-than-linear fashion, it is implausible that it should increase in a discontinuous fashion. Sixty-four seconds of pain at intensity sixty-four could not plausibly be twice as bad, let alone a million times as bad, as sixty-three seconds of pain at intensity sixty-three.

It is also worth noting that the validity of the Additive Impossibility Result will not depend on the interpretation of the "closeness" relation. While "closeness", on my intended interpretation, is a relation which holds between two wellbeing levels when they are evaluatively very similar, we might instead interpret "closeness" as a relation which holds between two wellbeing levels just in case it intuitively seems that one is not lexically superior to the other. The General Non-Elitism and General Non-Extreme Priority conditions we shall discuss later derive some of their intuitive force from the former, intended, interpretation of "closeness". But they may still be defensible on the latter, less stringent, interpretation of "closeness". Finite Fine-Grainedness, on the other hand, will be much harder to deny on this lax interpretation of "closeness". Yet the Additive Impossibility Result will hold however we interpret "closeness", provided we interpret it the same way for each premise.

3.3 Option-Set-Dependence and Acyclicity

Let us say that A is *sequentially worse* than B, which we denote by $A \triangleleft B$, if and only if there exists some sequence with at least two members $A_1, A_2, \ldots, A_n$, with $A_1 = A$ and $A_n = B$, such that for all $i < n$, $A_i \prec A_{i+1}$ in every option set containing A_i and A_{i+1}. Given transitivity, sequential worseness is equivalent to strict worseness in all option sets. Since we are not assuming transitivity, sequential worseness is a logically weaker notion than strict worseness in all option sets. We shall use it as a substitute for strict worseness. There are two reasons to appeal to sequential worseness. First, sequential worseness comparisons can be made without the need to worry about option set dependence. Second, and more importantly, the sequential worseness relation is automatically transitive, even if the ordinary worseness relation is not. That is, we have the

[17]Klocksiem (2016: 1324) similarly suggests that a pain episode is much worse if it counts as a "genuine pain".

[18]It is also at least questionable whether Finite Fine-Grainedness is even needed for something like the Additive Impossibility Result to go through. Thornley (2021) has shown that we can make do without this principle in Arrhenius's Sixth Theorem if we use probabilistic versions of General Non-Elitism and General Non-Extreme Priority. This move could be adapted for the Additive Impossibility Result. (I have avoided doing so in order to keep the proof relatively simple.)

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

Transitivity Lemma Let X, Y and Z be any populations. Suppose $X \triangleleft Y$ and $Y \triangleleft Z$. Then $X \triangleleft Z$. A proof of this fact may be found in the Appendix.

The premises of the Additive Impossibility Result will be stated in terms of the sequential worseness relation. Since this relation is a little artificial, this could make it difficult to determine how plausible these premises are. It is therefore worth keeping in mind that the relation of being worse in *all* option sets is logically weaker than sequential worseness. This is because, if X is worse than Y in all option sets, then there is a sequence, namely X, Y, such that each member of the sequence is worse than its successor (if it has one) in all option sets containing the two. We can therefore imagine the premises of the Additive Impossibility Result (except for Acyclicity) as being stated in terms of the "worseness in all option sets" relation, rather than in terms of sequential worseness. As these reinterpreted premises would be logically stronger, imagining them in this way will not make them seem more plausible than they are. The reason I have stated the premises in terms of sequential worseness rather than worseness in all option sets is simply that this simplifies the proofs, as it allows us to appeal to the Transitivity Lemma.

Those qualifications aside, the sequential worseness notation allows us to easily state our second structural premise. This will be an acyclicity condition which we shall use in place of transitivity:

Acyclicity For every population X, it is not the case that $X \triangleleft X$.

3.4 Premises

The Additive Impossibility Result has four non-structural premises. Our avoidance condition for the Repugnant Conclusion shall be the

Additive Anti-Repugnance Condition For any barely good wellbeing level z, there exists some good wellbeing level $a > z$ and bad wellbeing level $b' < z$, and numbers of lives n and m, such that for any groups A and B consisting of n and at least m lives respectively, any group Z, any unaffected background population I, and any bad level $b \leq b'$,

$$I + B[b] + Z[z] \triangleleft I + A[a]$$

The Additive Anti-Repugnance Condition, illustrated by Figure 1, requires that an A-population must always make a better addition than a Z-population, but only when some arbitrarily bad population B is also added along with the Z-population. In other words, it is an avoidance condition for an additive version of the Very Repugnant Conclusion, which differs from the Repugnant Conclusion in that it compares a population of excellent lives to an extremely large number of lives barely worth living, *plus* a smaller number of very bad lives.[19]

[19]See Arrhenius (2003, 2009, 2011).

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

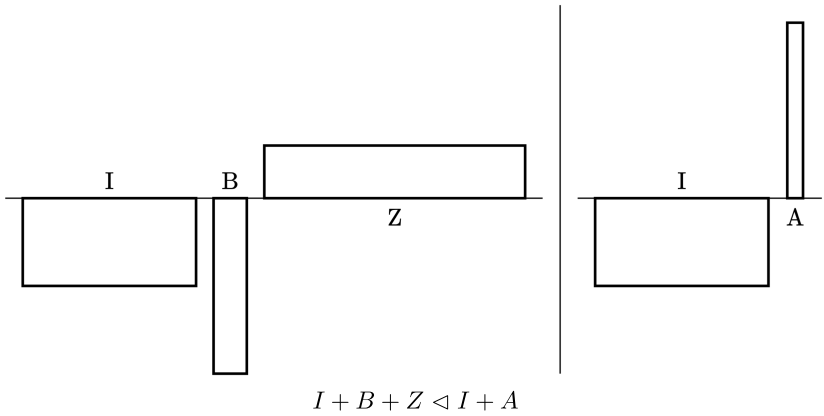


Figure 1. The Additive Anti-Repugnance Condition.

Our other different-number condition shall be the

Non-Additive Weak Non-Sadism Condition[20] There exists a bad wellbeing level b which is not minimal,[21] and some number n, such that for any group B of size at least n, any $b^- \leq b$, and any population A containing only good lives,

$$B[b^-] \lhd A$$

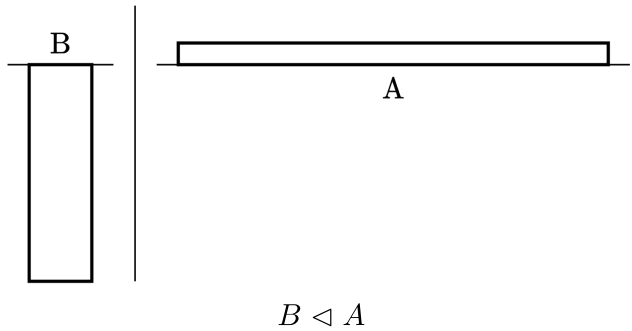


Figure 2. The Non-Additive Weak Non-Sadism Condition.

The Non-Additive Weak Non-Sadism Condition, illustrated by Figure 2, requires that there is *some* population which is so bad that it is worse than any population consisting solely of good lives. While some of the premises of our impossibility theorem might be up for debate, I believe that this one is not, *provided* one

[20]This principle is closely related to what Blackorby *et al.* (2005: 135) call "Priority for Lives Worth Living".

[21]A wellbeing level w is minimal just in case there does not exist any $w^- < w$.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

countenances any different-number comparisons whatsoever.[22] The Non-Additive Weak Non-Sadism Condition is not satisfied by all proposed population axiologies; it is not, for instance, satisfied by positive critical level views, such as those discussed by Blackorby *et al.* (2005: Ch. 5). But such views are to be rejected precisely because they do not satisfy the Non-Additive Weak Non-Sadism Condition.[23]

Our same-person principles shall be analogues of the ones appealed to by Arrhenius (2009, 2011, n.d.). They are more or less the same as the same-named principles in these works, with the main difference being that their claims are in terms of the $\triangleleft$ relation, rather than the worseness relation.[24] Apart from this, they differ only notationally, and in one or two unimportant respects.[25] They are as follows.

General Non-Elitism For any wellbeing levels $a > e > c$, where a is close to e, there exists a number n such that for any single-person group A, any group C of at least n people, and any unaffected background population I,

$$I + A[a] + C[c] \triangleleft I + A[e] + C[e]$$

General Non-Elitism, illustrated by Figure 3, says that rather than providing a small benefit to a single better off person, it would be better to instead provide benefits of a fixed size to a sufficiently large number of worse-off people.

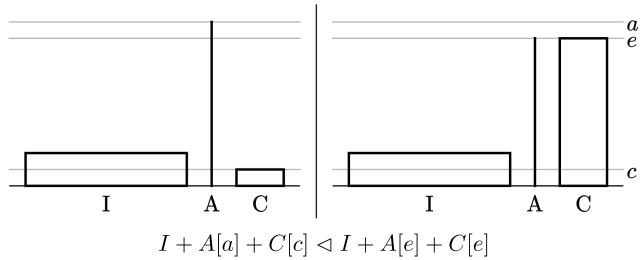


Figure 3. General Non-Elitism.

[22]Bader (2022a, b) notably demurs from this assumption, claiming instead that all different-number populations are incomparable. I am inclined to reject his view precisely on the grounds that it does not validate the Non-Additive Weak Non-Sadism Condition. In any case, Bader's view also fails to satisfy the Additive Anti-Repugnance Condition.

[23]The usual rationale offered for critical level views is precisely that they allow for avoidance of the Repugnant Conclusion. But since avoidance of the (additive or non-additive) Repugnant Conclusion is less intrinsically plausible than the Non-Additive Weak Non-Sadism Condition, this kind of rationale is insufficient to justify these views.

[24]Since Arrhenius (2009, 2011) assumes transitivity and implicitly assumes option set independence, the two relations are equivalent in his framework.

[25]My version of General Non-Extreme Priority differs in that because of a reordering of some of the quantifiers it is compatible with the view that priority given to the worse-off increases without a bound as people move down the wellbeing scale. Additionally, my versions of General Non-Elitism and General Non-Extreme Priority are formulated as fixed-population principles; in contrast, Arrhenius (2009, 2011) states them as same-number principles. The final difference is that many variables in my statements of the conditions are formulated as inequalities: such-and-such holds for any population of at least this number of lives, at at least/at most this level, rather than for any population of exactly this many lives at exactly this level. The strengthenings in this respect are harmless in terms of their impact on the plausibility of the premises.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

General Non-Extreme Priority For any wellbeing level w, and any barely good level z, there is a good wellbeing level $a' > z$, and a number n, such that for any wellbeing levels $a \geq a'$ and $w^+ > w$, where w^+ is close to w, any group A of size at least n, any single-person group C, and any unaffected background population I,

$$I + A[w^+] + C[z] \lhd I + A[w] + C[a]$$

General Non-Extreme Priority, illustrated by Figure 4, says that rather than benefiting a single person who is perhaps badly off to some small extent, it would be better to instead lift some sufficiently large number of people up from a barely good wellbeing level to some sufficiently good wellbeing level. Both principles are very plausible.

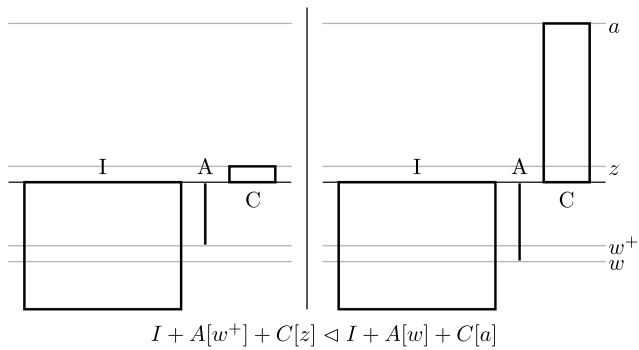


Figure 4. General Non-Extreme Priority.

3.5 The Additive Impossibility Result

That is enough groundwork to state the main result of this paper:

The Additive Impossibility Result There is no population axiology which satisfies all of the following conditions:

(1) Finite Fine-Grainedness.
(2) Acyclicity.
(3) The Additive Anti-Repugnance Condition.
(4) The Non-Additive Weak Non-Sadism Condition.
(5) General Non-Elitism.
(6) General Non-Extreme Priority.

3.6 Lemmas

To prove the Additive Impossibility Result, we shall need to appeal to three lemmas. We shall need to show that our premises imply "Inequality-Averse Addition", "Sufficient Trade-Offs" and "Axiological Aggregation". Roughly, according to Inequality-Averse Addition, large improvements to a first group of people are outweighed by small improvements to a sufficiently large second group of people

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

who are worse off than the first group. According to Sufficient Trade-Offs, rather than having a first and second group of people, all at some barely good wellbeing level, it would be better if instead the first group were at some very good wellbeing level, and the second group were at some very bad wellbeing level, provided the first group is sufficiently larger than the second. According to Axiological Aggregation, rather than having a first and second group of people, all at some barely good wellbeing level, it would be better if the first group were at some slightly better level, and the second group were at some bad level, provided the first group is sufficiently larger than the second. Each of these principles applies in the presence of any unaffected background population. More precisely, the three lemmas we need, which are illustrated by Figures 5, 6 and 7 respectively, are as follows:

The Inequality Aversion Lemma The General Non-Elitism Condition and Finite Fine-Grainedness imply **Inequality-Averse Addition**. For any wellbeing levels $a > e > c$, and any number n, there is a number m such that if A and C are groups containing n and at least m lives respectively, and I is any unaffected background population,

$$I + A[a] + C[c] \lhd I + A[e] + C[e]$$

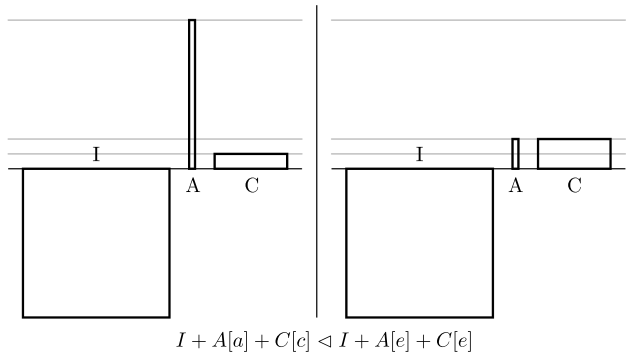


Figure 5. Inequality-Averse Addition.

The Sufficient Trade-Offs Lemma The General Non-Extreme Priority condition and Finite Fine-Grainedness imply **Sufficient Trade-Offs**. For any barely good z, any wellbeing level $b < z$, and any number n, there is a good wellbeing level $a' > z$ and a number m such that for any $a \geq a'$, any group A of size at least m, any group C of size n, and any unaffected background population I,

$$I + A[z] + C[z] \lhd I + A[a] + C[b]$$

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

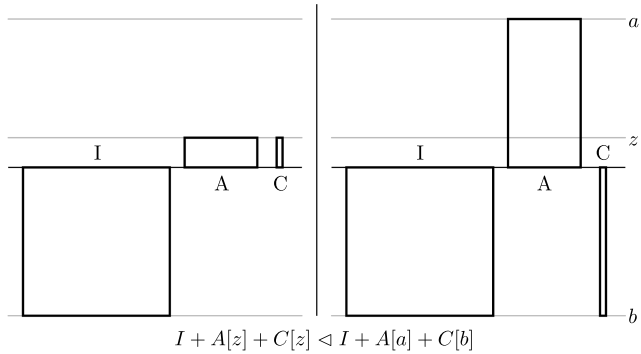


Figure 6. Sufficient Trade-Offs.

The Axiological Aggregation Lemma The General Non-Elitism Condition, General Non-Extreme Priority and Finite Fine-Grainedness imply **Axiological Aggregation**. For any barely good z, any $z^+ > z$ which is not maximal, any $b < z$, and any number of lives n, there is a number of lives m such that if B and Z are groups of n and m people respectively, and I is any unaffected background population, then

$$I + B[z] + Z[z] \lhd I + B[b] + Z[z^+]$$

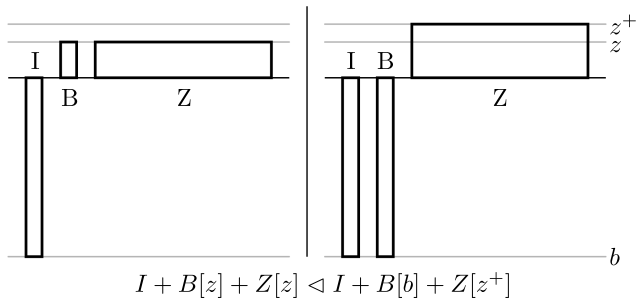


Figure 7. Axiological Aggregation.

Proofs of all three lemmas may be found in the Appendix. The ideas of the proofs of the first two lemmas are fairly simple, and are exactly analogous to the proofs of Arrhenius's (2011) Lemmas 1.1 and 1.2 respectively. Both Inequality-Averse Addition and Sufficient Trade-Offs result from applying General Non-Elitism and General Non-Extreme Priority finitely many times. If a trade-off can be made between one person and m people, then, by applying such a trade-off n times, one can show that the same kind of trade-off can be made between n people and $n \cdot m$ people. Similarly, we can drop the restriction that only small differences in wellbeing can be traded-off by repeatedly applying such many-person trade-offs along a finite chain of consecutively close wellbeing levels between any two wellbeing levels.[26]

[26]A chain w_i of wellbeing levels is consecutively close just in case each w_j is close to w_{j+1}.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

Such a chain always exists by the Finite Fine-Grainedness condition. Essentially, the proofs of the first two lemmas just consist in showing that both things can be done. Axiological Aggregation may then be proved by applying both lemmas finitely many times (that is, by induction).

3.7 Proof of the Additive Impossibility Result

The Additive Impossibility Result, illustrated by Figure 8, is proved by constructing groups B, B', Z and A, and wellbeing levels $a > z > z^- > b > b^-$, in such a way that the populations illustrated in Figure 8 bear the $\triangleleft$ relations as stated in the caption of the same figure. In this construction, a is a good wellbeing level, z and z^- are barely good levels, and b and b^- are bad wellbeing levels.

$$(i) \triangleleft (ii) \triangleleft (iii) \triangleleft (iv) \triangleleft (i)$$

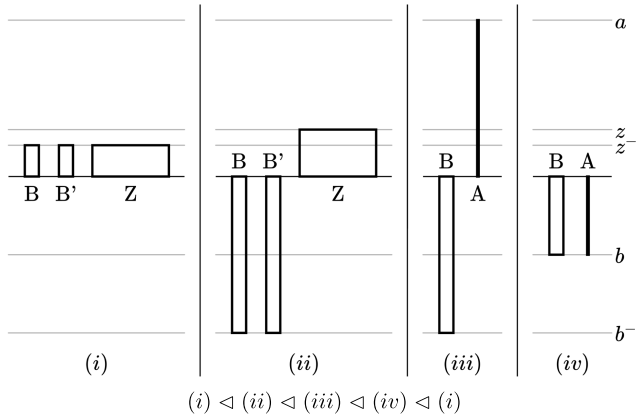


Figure 8. Proof of the Additive Impossibility Result.

We choose the sizes of groups A and B', and wellbeing levels a and b^-, in such a way that the claim that $(ii) \triangleleft (iii)$ is an instance of the Additive Anti-Repugnance Condition, with B the unaffected background population. We additionally stipulate that the size of group B is sufficiently large for us to obtain that $(iii) \triangleleft (iv)$, by Inequality-Averse Addition. We obtain $(iv) \triangleleft (i)$ by the Non-Additive Weak Non-Sadism Condition.[27] Since nothing said so far turns on the size of group Z, we can ensure that Z is chosen to be large enough that Axiological Aggregation implies that $(i) \triangleleft (ii)$. Finally, applying the Transitivity Lemma, we obtain that $(i) \triangleleft (i)$, contradicting Acyclicity.

4. Which Premise Should be Rejected?

Since the premises of the Additive Impossibility Result are mutually inconsistent, at least one of them has to go. But which? For reasons stated earlier, I believe that we

[27]That is, in constructing B, we ensure that it is large enough to the appropriate role in the application of both Inequality-Averse Addition and the Non-Additive Weak Non-Sadism Condition.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

should not reject either Finite Fine-Grainedness or the Non-Additive Weak Non-Sadism Condition. The remaining premises are Acyclicity, General Non-Elitism and General Non-Extreme Priority. Alternatively, we can reject the Non-Additive Weak Non-Sadism Condition, thereby accepting an additive version of the Repugnant Conclusion. I shall consider each option in turn.

4.1 Acyclicity

It has been suggested, most prominently by Larry Temkin (1987, 1996, 2012) and Stuart Rachels (1998, 2001, 2004), that we might avoid the Repugnant Conclusion by denying the transitivity of the at-least-as-good-as relation. While both authors also deny Acyclicity, it seems to me that the two moves are not on a par: it is less plausible to deny Acyclicity than it is to deny transitivity. To see why, let us first consider the case for (in)transitivity, and next consider the case for Acyclicity.

Why might we antecedently want to accept transitivity, beyond the mere intuitive plausibility of that principle? One standard argument holds that we must accept transitivity because it is an analytic feature of comparatives: as a matter of logic, whenever A is F-er than B, and B is F-er than C, A must be F-er than C (Broome 2004). It might alternatively be claimed that transitivity is otherwise central to the concept of value, or that value is inherently quantitative (and therefore transitive).[28] I'm not sure whether these arguments succeed, but even if we assume that they do, sceptics of transitivity still have the nuclear option: they can say that talking about value is a mistake, and that we should instead theorize in terms of some other non-transitive normative relation.[29] If some such normative relation fulfils whatever role we wanted goodness to play in our moral theory, it is unclear that much is lost by abandoning value-talk in favour of theorizing in terms of this new relation.

As an example, say that one population is *impartially preferable* to another just in case there is all-things-considered reason to hope, from an impartial perspective, that the first population would come about rather than the second.[30] The impartial preferability relation would appear to be able to do everything we want a value relation to do, but as far as I can see, there is no obvious argument to the effect that the impartial preferability relation must be transitive.[31]

There are, however, strong pragmatic arguments for the irrationality of cyclic preferences. The most prominent of these are money pump arguments, some versions of which even apply to agents who take into account their expected future choices in their present decision-making.[32] If these arguments successfully show that

[28]See Klocksiem (2016).

[29]Rachels (2001: 218–219) suggests that even if transitivity is held to be central to value, one might nevertheless understand intransitivity in terms of some other normative relation.

[30]Some authors, such as Parfit (2011: 41–42), understand *value* in a similar "impartial-reason-implying sense".

[31]There are some pragmatic arguments against acyclic intransitivity; see for example Gustafsson (2010, 2022). The problem is that these arguments assume completeness, which on its face is less secure than transitivity itself. Gustafsson (2022) provides pragmatic arguments for completeness, but these arguments are less compelling than the pragmatic arguments for acyclicity.

[32]A recent version of this argument, provided by Gustafsson and Rabinowicz (2020), shows that any agent with cyclic preferences is susceptible to exploitation, provided only that she obeys a minimal principle of backwards induction. Moreover, the pump they provide can be iterated, showing that any such agent can

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

cyclic preferences are irrational, the betterness relation must be acyclic (given that choosing in accordance with the betterness relation is not irrational). Importantly, pragmatic arguments seem to apply just as well to any relation which might be offered up as a replacement for betterness by a transitivity sceptic taking the nuclear option. If a normative relation R is to play the role of traditional value relations in our conceptual theorizing, one feature it must have is to be morally decisive in cases in which only R-relevant factors are at stake. As a result, if R is cyclic, abiding by morality will sometimes require one to act on cyclic preferences. Therefore, given that successfully abiding by morality is not irrational, that rational agents are not susceptible to exploitation by money pump, and that agents with cyclic preferences are susceptible to exploitation by money pump, any normative relation R purporting to stand in for traditional value relations must satisfy Acyclicity.

Note also that the only way to deny that agents with cyclic preferences are susceptible to exploitation by money pump is to claim that an agent's preferences at some point in a decision tree can turn on more than just the options achievable for her going forward from that point.[33] But if that is true, then even if a theory says (for instance) that populations of excellent lives are better than very large populations of lives barely worth living, the same theory may nevertheless instruct an agent to bring about the population of lives barely worth living, if the agent is at some suitable point in a decision tree. Any non-exploitable theory must avoid actually issuing its usual judgement for at least some pair of populations involved in a betterness cycle, for at least some points in a decision tree. Denial of Acyclicity is therefore not a clean way of avoiding the Repugnant Conclusion without incurring other problematic commitments. Even if Acyclicity is false, we must still deny at least one of the premises of the Additive Impossibility Result in at least some sequential decision contexts, on pain of having to say that rational moral agents sometimes *should* get money pumped, even when they can see it coming.

4.2 General Non-Elitism and General Non-Extreme Priority

Recall that according to General Non-Elitism, rather than slightly benefiting one person who is comparatively well off, it would be better to instead benefit some sufficiently large number of people who are worse off. According to General Non-Extreme Priority, rather than slightly benefiting one person, who may initially be very badly off, it would be better to instead provide large benefits to a sufficiently large number of people whose lives are barely worth living.

For reasons mentioned in §3.3, we shall understand these claims as applying in every option set. Such claims to the effect that some population A is worse than B in every option set can be denied in a number of ways, some of which are more plausible than others. First, one might reject the claim outright, insisting instead that there are cases where A is better than B in every option set. Second, one could adopt

be forced to give up arbitrarily much of something they care about (Gustafsson and Rabinowicz 2020: 586). In the present context, the relevant quantity would be people's wellbeing.

[33]See McClennen (1985), or more recently, Ahmed (2017).

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

the weaker position that A and B might be incomparable in every option set.[34] Third, one could admit that it is determinately true that there is some case in which A is not worse than B (in every option set), but deny, for each pair of populations A and B falling under the principle, that it is determinately true that A is not worse than B (in every option set).[35] Fourth, one could accept that A is worse than B in the pairwise option set consisting of just A and B, while allowing that this may fail to be the case in some larger option sets.[36]

There is little to be said for the first option. It would be very implausible to claim that it would be better to benefit a single better-off person, rather than arbitrarily many worse-off people. It would be only slightly less implausible to claim that it would be better to benefit a single badly off person by a very small amount, rather than benefiting a large number of people with lives barely worth living by a great amount. Moreover, Thornley (2021) has shown that the two fixed-population principles can be weakened in the following way: the potential losses to one person in each principle can be replaced by arbitrarily small probabilities of a loss to one person. To my mind, this sinks the first option.[37]

The remaining three options represent different paths for mitigating the intuitive costs of denying General Non-Elitism or General Non-Extreme Priority, but it is unclear how much they really help. Consider, for instance, General Non-Elitism. It seems false that benefiting a single person by a small amount could be better than benefiting an arbitrarily large number of less well-off people by larger amounts. It also seems false that benefiting the better-off person could be not worse than benefiting the less well-off people, or that benefiting the better-off person could fail to be determinately worse than benefiting the less well-off people. It is hard to see how even the last, weakest judgement could fail in *any* option set. Again, these claims become even more secure if we shift to probabilistic versions of our fixed-population principles. The prospects for rejecting General Non-Elitism or General Non-Extreme Priority, even when mitigated by taking one or more of options two to four, thus look dim to me.

[34]I define two populations to be incomparable just in case they are unrelated to each other by $\leq$. It does not matter for the purposes of this paper whether this means that no positive evaluative relation at all holds between the populations, or whether some positive evaluative relation not definable in terms of $^\circ$, such as parity (Chang 2002) or imprecise equality (Parfit 2016), might nevertheless hold. Hájek and Rabinowicz (2022) have recently suggested we appeal to incomparability in order to defuse the paradoxes of population ethics.

[35]See Thomas (2022a, b).

[36]This option is advocated by Frick (2022) and, for a non-evaluative interpretation of $\succ$, by Boonin-Vail (1996).

[37]An anonymous reviewer points out that although each instance of the General Non-Elitism and General Non-Extreme Priority conditions looks compelling by itself, the Additive Impossibility Result proceeds via repeated application of these principles; and insofar as the Additive Impossibility Result is paradoxical, one response might be to doubt the general principles, even though their instances seem almost undeniable. It still seems to me, however, that the general principles themselves remain more compelling than the other premises of the Additive Impossibility Result. And, indeed, efforts to deny the various aggregation principles derived from these two conditions focus more often on Finite Fine-Grainedness than on the General Non-Elitism and General Non-Extreme Priority conditions; see for instance Thomas (2018), Carlson (2022) and Nebel (2022).

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

4.3 Accepting the Repugnant Conclusion

To reject the Additive Anti-Repugnance Condition is to accept that a population of excellent lives may fail to make a better addition than a population consisting of some very bad lives, together with arbitrarily many lives barely worth living. If this is accepted, it would seem natural to also accept a stronger, non-additive version of this claim: a population of excellent lives can be worse, by itself, than a combination of many bad lives with arbitrarily many lives which are barely worth living. That is, if we reject the Additive Anti-Repugnance Condition, we should accept the Very Repugnant Conclusion.

One might deny this last claim. Avoidance of the Very Repugnant Conclusion is *consistent* with the negation of the Additive Anti-Repugnance Condition, just as the Non-Additive Weak Non-Sadism Condition is consistent with the negation of Weak Non-Sadism. (To see that the first of these claims is true, note that Average Utilitarianism avoids the (Very) Repugnant Conclusion, but does not satisfy the Additive Anti-Repugnance Condition.)[38] One might therefore think that the Additive Impossibility Result tells us nothing new that is important: Weak Non-Sadism is arguably more plausible than the Additive Anti-Repugnance Condition, and we already know from the Sixth Impossibility Theorem that (assuming the other premises) we must choose between Weak Non-Sadism and avoidance of the Repugnant Conclusion. So why worry? If we deny the Additive Anti-Repugnance Condition, we are not much worse off than before with respect to avoiding the Very Repugnant Conclusion, because we have only rejected a principle which is less plausible than Weak Non-Sadism.

Put this way, I think the worry is overstated, because it seems to me that even if the Additive Anti-Repugnance Condition is intrinsically less plausible than Weak Non-Sadism, there is still an important sense in which rejecting this condition would be "repugnant". We are inclined to think that the Repugnant Conclusion is false because of its particular character: it *could not be*, we think, that a population of lives barely worth living is better than a population of excellent lives, just because the former contains a very large number of people. But if the Additive Anti-Repugnance Condition is false, this intuition is undermined: it *can* be that a population of lives barely worth living is better as an addition than a population of excellent lives, just because the former contains a very large number of people. In contrast, if we deny Weak Non-Sadism, this does not by itself undermine the intuition of repugnance. Rejecting the Weak Non-Sadism condition can be a way of robustly avoiding the Repugnant Conclusion; rejecting the Additive Anti-Repugnance Condition cannot.

Accepting the Repugnant Conclusion is a radical step. Most people who have thought seriously about the Repugnant Conclusion find it deeply counterintuitive, taking there to be decisive reason to reject any population axiology which has it, or a version of it, as a consequence.[39] Even so, it seems to me that unlike in the case of rejecting other premises of the Additive Impossibility Result, such as the Non-Additive Weak Non-Sadism Condition, it is at least explicable how the Repugnant Conclusion could be true. A large enough population of lives barely worth living can indeed contain more – indeed, *much* more – of whatever makes life worth living than a smaller population of excellent lives. This fact alone cannot justify accepting

[38]This was first noticed by Bill Anglin (1977), and more recently by Gustaf Arrhenius (n.d.).

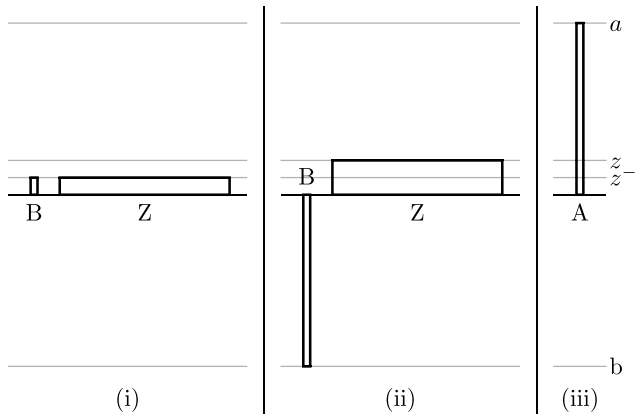
[39]That said, the traditional view that the Repugnant Conclusion is to be avoided at all costs no longer enjoys the near-unanimity it once had; see Zuber *et al.* (2021).

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

the Repugnant Conclusion, because as most moral philosophers rightly observe, people are not mere containers of value. But it can at least help to make it intelligible that the Repugnant Conclusion might be true. Moreover, there are many impossibility theorems in the population ethics literature, many of which include premises which are quite different in character to those appealed to in the Additive Impossibility Result, but almost all of which include a condition for avoiding the Repugnant Conclusion. These can be seen as providing multiple lines of evidence supporting the Repugnant Conclusion.

It may well be objected at this point that what we are dealing with here is a version of the *Very* Repugnant Conclusion not Parfit's original Repugnant Conclusion. It is widely held that even those who are able to hold their nose and accept the Repugnant Conclusion should nevertheless baulk at its Very Repugnant cousin. I think we should reject this conventional wisdom. Let me explain why.

Suppose that b be a terrible wellbeing level, that z is the level of a life barely worth living, and that z^- is the level of a slightly worse life, which is still worth living. Let B be a large group of people, and let Z be a much larger group of people. The Very Repugnant Conclusion has it that a population $B[b] + Z[z]$ can be better than a population of excellent lives, while the regular Repugnant Conclusion has it that a large enough population of lives barely worth living can be better than a population of excellent lives. The former is supposed to be harder to believe than the latter. But if this is so, it must be because there is something implausible about Axiological Aggregation, and therefore something implausible about either General Non-Elitism or General Non-Extreme Priority. For Axiological Aggregation implies that great losses of wellbeing for a large number of people can matter less than tiny gains for a huge number of people. It therefore implies that the population $B[b] + Z[z]$ could be sequentially better than the population $B[z^-]+Z[z^-]$. This of course guarantees that if the latter population of lives barely worth living is sequentially better than a smaller population of excellent lives A, then the former population must also be sequentially better than A. This argument is illustrated by Figure 9.



(i) ◁ (ii); hence if (ii) ◁ (iii) then (i) ◁ (iii).

Figure 9. The Very Repugnant Proposition.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

This argument can be made more precise. Specifically, it can be shown that, given Axiological Aggregation, the avoidance condition for the regular Repugnant Conclusion implies the avoidance condition for the Very Repugnant Conclusion. This is proved, as the "Very Repugnant Proposition", in the Appendix.

(i) ◁ (ii); hence if (ii) ◁ (iii) then (i) ◁ (iii).

The upshot is that insofar as the Very Repugnant Conclusion is less plausible than the regular Repugnant Conclusion, the fault must lie with at least one of the same-person principles of General Non-Elitism or General Non-Extreme Priority, or with Finite Fine-Grainedness. But, as I have already argued, I do not think that we should reject these conditions.

For these reasons, I think that the least implausible response to the Additive Impossibility Result is to reject the Additive Anti-Repugnance Condition, and thereby accept a version of the Very Repugnant Conclusion. But I expect that this opinion will not be widely shared. It is therefore worth mentioning one final option for those who simply cannot reject any particular premise of the Additive Impossibility Result, which is to flip the table and reject the entire basis of the argument.[40] One might take the Additive Impossibility Result, together with the many other impossibility theorems in the literature, to show that there is no such thing as a complete, justifiable theory of population ethics, and that we therefore have reason to adopt a moral error theory.[41] While I cannot offer a detailed argument against this sort of response here, it strikes me as an overreaction. It seems to me better to reject one premise than to reject all six.

5. Conclusion

I have presented the Additive Impossibility Result, which demonstrates that an additive version of the Repugnant Conclusion is extremely difficult to avoid. Unlike most other impossibility theorems, the Additive Impossibility Result does not include any version of the Mere Addition Principle, or any different-number principle which could feasibly be denied on the grounds that Mere Addition is false. Instead, it assumes the compelling Non-Additive Weak Non-Sadism Condition. The main practical upshot, I think, is this. In a choice between Weak Non-Sadism and avoidance of the Repugnant Conclusion, it may be unclear which should stay and which should go. Arguably, both are similarly compelling. In contrast, in a choice between the Non-Additive Weak Non-Sadism Condition and the Additive Anti-Repugnance Condition, it is clear that the former is by some margin the more plausible principle. Since these are the only two different-number principles involved in the Additive Impossibility Result, we may conclude that the Additive Anti-Repugnance Condition cannot reasonably be maintained by abandoning the Non-Additive Weak Non-Sadism Condition. Neither do denial of Acyclicity or Finite Fine-Grainedness offer a straightforward path to retaining the Additive Anti-Repugnance Condition. Instead, the most plausible response to the Additive Impossibility Result, for opponents of the Repugnant Conclusion, may be to reject the fixed-population assumptions: either General Non-Elitism or General Non-

[40]I am grateful to an anonymous referee for suggesting this response.

[41]See Arrhenius (2011: 23) and particularly Cowie (2023: 467–469) on this sort of response.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

Extreme Priority (or both). I have suggested that the intuitive costs of doing so might be mitigated by appealing to incompleteness, option set dependent betterness, indeterminacy, or some combination of the three. Yet even with these mitigating factors in play, I find these principles too compelling to deny.

Let me make one final point. While the relation $\succeq$ was interpreted throughout this paper as the moral at-least-as-good-as relation, its structure as a three-place relation incorporating option set dependence means that it can be given other interpretations. In particular, it can be interpreted as the at-least-as-much-reason-to-bring-about relation. The Additive Impossibility Result on this normative interpretation of $\succeq$ might be more troubling than the axiological interpretation of the theorem. On the normative version of my favoured response to the axiological Additive Impossibility Result, we would sometimes have more reason to bring about a future consisting of many tortured lives, and many more barely good lives, than we would have to bring about a future consisting of a smaller number of excellent lives. This is a conclusion I would be prepared to accept, but I don't expect I shall have much company.

Acknowledgements. I would like to thank Ralf Bader, Timothy Campbell, Roger Crisp, Hilary Greaves, Todd Karhu, Kacper Kowalczyk, Seán O'Neill McPartlin, Aidan Penn, Daniel Ramöller, Teruji Thomas, Biqing Wang and three anonymous referees for helpful comments on this paper.

References

Ahmed A. 2017. Exploiting cyclic preference. *Mind* **126**(504), 975–1022.

Anglin B. 1977. The repugnant conclusion. *Canadian Journal of Philosophy* **7**(4), 745–754.

Arrhenius G. 2000. An impossibility theorem for welfarist axiologies. *Economics and Philosophy* **16**(2), 247–266.

Arrhenius G. 2003. The very repugnant conclusion. In *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Åqvist*, ed. S. Krister and R. Sliwinski, 167–180. Uppsala: Uppsala University Press.

Arrhenius G. 2009. One more axiological impossibility theorem. In *Logic, Ethics, and All That Jazz: Essays in Honour of Jordan Howard Sobel*, ed. L.-G. Johansson, J. Österberg and R. Sliwinski, 23–37. Uppsala: Uppsala University Press.

Arrhenius G. 2011. The impossibility of a satisfactory population ethics. In *Descriptive and Normative Approaches to Human Behaviour*, ed. E.N. Dzhafarov and L. Perry, 1–26. Singapore: World Scientific Publishing Co.

Arrhenius G. 2016. Population ethics and different-number based imprecision. *Theoria* **82**, 166–181.

Arrhenius G. n.d. Population ethics: The challenge of future generations. Unpublished manuscript.

Bader R.M. 2022a. The asymmetry. In *Ethics and Existence: The Legacy of Derek Parfit*, ed. J. McMahan, T. Campbell, J. Goodrich and K. Ramakrishnan, 15–37. Oxford: Oxford University Press.

Bader R.M. 2022b. Person-affecting utilitarianism. In *The Oxford Handbook of Population Ethics*, ed. G. Arrhenius, K. Bykvist, T. Campbell and E. Finneron-Burns, 251–270. Oxford: Oxford University Press.

Blackorby C., W. Bossert and D. Donaldson 1996. Quasi-orderings and population ethics. *Social Choice and Welfare* **13**(2). 129–150.

Blackorby C., W. Bossert and D. Donaldson 2003. The axiomatic approach to population ethics. *Politics, Philosophy & Economics* **2**(3), 342–381.

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

Boonin-Vail D. 1996. Don't stop thinking about tomorrow: two paradoxes about duties to future generations. *Philosophy & Public Affairs* **25**(4), 267–307.

Broome J. 2004. *Weighing Lives*. Oxford: Oxford University Press.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

Carlson E. 1998. Mere addition and two trilemmas of population ethics. *Economics and Philosophy* **14**(2), 283–306.

Carlson E. 2022. On some impossibility theorems in population ethics. In *The Oxford Handbook of Population Ethics*, ed. G. Arrhenius, K. Bykvist, T. Campbell and E. Finneron-Burns, 205–225. New York: Oxford University Press.

Chang R. 2002. The possibility of parity. *Ethics* **112**(4), 659–688.

Chang R. 2016. Parity, imprecise comparability and the repugnant conclusion. *Theoria* **82**, 182–214.

Cowie C. 2023. Why moral paradoxes support error theory. *Journal of Philosophy* **120**(9), 457–483.

Cusbert J. 2017. Acting on essentially comparative goodness. *Thought: A Journal of Philosophy* **6**(2), 73–83.

Francis T. 2023. Intrapersonal arguments for the repugnant conclusion. *Ethics* **134**(1), 89–107.

Frick J. 2014. *Making People Happy, Not Making Happy People: A Defense of the Asymmetry Intuition in Population Ethics*. PhD thesis, Harvard University.

Frick J. 2022. Context-dependent betterness and the mere addition paradox In *Ethics and Existence: The Legacy of Derek Parfit*, ed. J. McMahan, T. Campbell, J. Goodrich and K. Ramakrishnan, 232–263. Oxford: Oxford University Press.

Gustafsson J.E. 2010. A money-pump for acyclic intransitive preferences. *Dialectica* **64**(2), 251–257.

Gustafsson J.E. 2020. Population axiology and the possibility of a fourth category of absolute value. *Economics and Philosophy* **36**(1), 81–110.

Gustafsson J.E. 2022. *Money-Pump Arguments*. Cambridge: Cambridge University Press.

Gustafsson J.E. and W. Rabinowicz 2020. A simpler, more compelling money pump with foresight. *Journal of Philosophy* **117**(10), 578–589.

Hájek A. and W. Rabinowicz 2022. Degrees of commensurability and the repugnant conclusion. *Noûs* **56**(4), 897–919.

Klocksiem J. 2016. How to accept the transitivity of better than. *Philosophical Studies* **173**, 1309–1334.

McClennen E.F. 1985. Prisoner's dilemma and resolute choice. In *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb's Problem*, ed. R. Campbell and L. Sowden, 94–104. Vancouver: University of British Columbia Press.

Nebel J.M. 2019. An intrapersonal addition paradox. *Ethics* **129**(2), 309–343.

Nebel J.M. 2022. Totalism without repugnance. In *Ethics and Existence: The Legacy of Derek Parfit*, ed. J. McMahan, T. Campbell, J. Goodrich and K. Ramakrishnan, 200–231. Oxford: Oxford University Press.

Ng Y.-K. 1989. What should we do about future generations? Impossibility of Parfit's theory X. *Economics and Philosophy* **5**(2), 235–253.

Parfit D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

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

Parfit D. 2016. Can we avoid the repugnant conclusion?. *Theoria* **82**(2), 110–127.

Qizilbash M. 2007a. The mere addition paradox, parity and vagueness. *Philosophy and Phenomenological Research* **75**(1), 129–151.

Qizilbash M. 2007b. The parity view and intuitions of neutrality. *Economics and Philosophy* **23**(1), 107–114.

Qizilbash M. 2018. On parity and the intuition of neutrality. *Economics and Philosophy* **34**(1), 87–108.

Rabinowicz W. 2009. Broome and the intuition of neutrality. *Philosophical Issues* **19**, 389–411.

Rachels S. 1998. Counterexamples to the transitivity of *Better Than*. *Australasian Journal of Philosophy* **76**(1), 71–83.

Rachels S. 2001. A set of solutions to Parfit's problems. *Noûs* **35**(2), 214–238.

Rachels S. 2004. Repugnance or intransitivity: a repugnant but forced choice. In *The Repugnant Conclusion: Essays on Population Ethics*, ed. J. Ryberg and T. Tännsjö, 163–186. London: Kluwer Academic.

Roberts M.A. 2011. The asymmetry: a solution. *Theoria* **77**(4), 333–367.

Spears D. and M. Budolfson 2021, Repugnant conclusions. *Social Choice and Welfare* **57**(3), 567–588.

Temkin L.S. 1987. Intransitivity and the mere addition paradox. *Philosophy & Public Affairs* **16**(2), 138–187.

Temkin L.S. 1996. A continuum argument for intransitivity. *Philosophy & Public Affairs* **25**(3), 175–210.

Temkin L.S. 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*, Oxford: Oxford University Press.

Thomas T. 2018. Some possibilities in population axiology. *Mind* **127**(507), 807–832.

Thomas T. 2022a. Are spectrum arguments defused by vagueness? *Australasian Journal of Philosophy* **100**(4), 743–757.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

Thomas T. 2022b. On evaluative imprecision. In *Ethics and Existence: The Legacy of Derek Parfit*, ed. J. McMahan, T. Campbell, J. Goodrich and K. Ramakrishnan, 478–497. Oxford: Oxford University Press.
Thornley E. 2021. The impossibility of a satisfactory population prospect axiology (independently of finite fine-grainedness). *Philosophical Studies* **178**(11), 3671–3695.
Zuber S. et al. 2021. What should we agree on about the repugnant conclusion? *Utilitas* **33**(4), 379–383.

Tomi Francis is a Senior Researcher in philosophy at the University of Fribourg. His work focuses on obligations to future generations, value theory, interpersonal aggregation, and conflicts between moral ideals.

Appendix

The Transitivity Lemma

Transitivity Lemma Let A, B and C be any populations. Suppose $A \lhd B$ and $B \lhd C$. Then $A \lhd C$.

Proof of the Transitivity Lemma. Since $A \lhd B$, there exists some sequence $A_1, A_2, \ldots, A_n$, with $A_1 = A$ and $A_n = B$, such that, for each $i < n$, $A_i \prec A_{i+1}$ in every option set containing the two. Similarly, since $B \lhd C$, there exists some sequence $B_1, B_2, \ldots, B_m$, with $B_1 = B$ and $B_m = C$, such that, for each $i < m$, $B_i \prec B_{i+1}$ in every option set containing the two. The concatenation of these two sequences[42] is a sequence $C_1, C_2, \ldots, C_k$, with $C_1 = A$ and $C_k = C$, such that, for each $i < k$, $C_i \prec C_{i+1}$ in every option set containing the two. Thus we have $A \lhd C$.

The Inequality Aversion Lemma

The Inequality Aversion Lemma The General Non-Elitism Condition and Finite Fine-Grainedness imply

Inequality-Averse Addition. For any wellbeing levels $a > e > c$, and any number n, there is a number m such that, if A and C are groups containing n and at least m lives respectively, and I is any unaffected background population,

$$I + A[a] + C[c] \lhd I + A[e] + C[e]$$

Proof of the Inequality Aversion Lemma. We shall first prove that Inequality-Averse Addition holds under the additional restriction that e is close to a. Call this "Restricted Inequality-Averse Addition". Let $a > e > c$ be arbitrary wellbeing levels with a close to e. Let n be an arbitrary number. By General Non-Elitism, there exists some number k such that for any single-person group A, and any group of at least k people C, which is disjoint from A, and any disjoint unaffected background population I,

$$I + A[a] + C[c] \lhd I + A[e] + C[e]$$

Claim: kn is a witness to Inequality-Averse Addition under this additional restriction. To see this, let A be an arbitrary group of size n. Split this into single person-groups $A_1, \ldots, A_n$. Let C be an arbitrary group of at least kn people. Split C into k-person groups $C_1, \ldots, C_n$, with C_n containing any additional people over kn. Finally, let I be some arbitrary unaffected background population. We want to show that

$$I + A_1[a] + \ldots + A_n[a] + C_1[c] + \ldots + C_n[c] \lhd I + A_1[e] + \ldots + A_n[e] + C_1[e] + \ldots + C_n[e]$$

To see this, let D_j be the result of taking the left-hand side of (*) and replacing all $A_i[a]$ and $C_i[c]$ with $A_i[e]$ and $C_i[e]$, for all $i \leq j$. The left-hand side of (*) then equates to D_0, while the right-hand side equates to D_n. Furthermore, General Non-Elitism implies that $D_i \lhd D_{i+1}$ for each $i < n$. The Transitivity Lemma then implies that $D_0 \lhd D_n$, as required.

We now prove the general case. We shall show by induction on k that for any $a > e > c$ and number of lives n, if there exists a chain of consecutively close wellbeing levels of length $k \geq 2$ between a and e, then

[42]Excluding the first member $B_1 = B$ of the second sequence from this concatenation.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

Inequality-Averse Addition holds for a, e, c and n. The base case of $k = 2$ follows from Restricted Inequality-Averse Addition. Suppose for our induction step that for any $a' > e' > c'$ and n', Inequality-Averse Addition holds whenever there is a consecutively close chain of length j between a and any arbitrary e'. Let $a > e > c$ and n be arbitrary, and suppose that there is a consecutively close chain of length $j + 1$ between a and e, that is

$$a_1 > a_2 > \ldots > a_j = e' > a_{j+1} = e$$

By the induction hypothesis applied to $a > e' > c$ and n, there is some number m' of good lives such that, if A is a group of size n and C a disjoint group of size at least m', and I' any disjoint unaffected background population, then

$$I' + A[a] + C[c] \lhd I' + A[e'] + C[e'] \qquad (**)$$

We now apply Restricted Inequality-Averse Addition for wellbeing levels $e' > e > c$, and number of lives $n + m'$. We obtain that there is some number m such that, if A' is a group of size $n + m'$, and C' is a group of size at least m, and I is any disjoint unaffected background population, then

$$I + A'[e'] + C'[c] \lhd I + A'[e] + C'[e] \qquad (***)$$

In particular, if A is any group of size n and C is any disjoint group of size m', so that $A + C$ is a group of size $n + m'$, we have

$$I + A[e'] + C[e'] + C'[c] \lhd I + A[e] + C[e] + C'[e]$$

Thus, applying the Transitivity Lemma to (**) and (***), with $I' = I + C'[c]$, we have

$$I + A[a] + C[c] + C'[c] \lhd I + A[e] + C[e] + C'[e]$$

As required. This completes our induction.

Finally, by Finite Fine-Grainedness, for any $a > e > c$, there exists a finite consecutively close chain from a to e. Thus, by our induction, Inequality-Averse Addition holds for $a > e > c$ and number of lives n. As these were chosen arbitrarily, Inequality-Averse Addition holds in general.

The Sufficient Trade-Offs Lemma

The Sufficient Trade-Offs Lemma The General Non-Extreme Priority condition and Finite Fine-Grainedness imply

Sufficient Trade-Offs. For any barely good z, any wellbeing level $b < z$, and any number n, there is a good wellbeing level $a' > z$ and a number m such that for any $a \geq a'$, any group A of size at least m, any group C of size n, and any unaffected background population I,

$$I + A[z] + C[z] \lhd I + A[a] + C[b]$$

Proof of the Sufficient Trade-Offs Lemma. As in the proof of the previous lemma, we shall first prove a restricted version of the Sufficient-Trade-Offs principle, the "Restricted Trade-Offs Principle", which applies only to wellbeing levels $b < z$ which are close to z. Let z, then, be an arbitrary barely good wellbeing level, $b < z$ a close wellbeing level, and n an arbitrary number of lives. We want to show that, for some number m and wellbeing level $a' > z$, if A is a group of size at least m, C is a group of size n, $a \geq a'$, and if I is any unaffected background population, then

$$I + A[z] + C[z] \lhd I + A[a] + C[b]$$

Let number of lives k and wellbeing level a' be witnesses to General Non-Extreme Priority for z.[43] We then have, for any single-person group A' of size k, any $a \geq a'$, any single-person group C, and any unaffected background population I,

$$I + A'[z] + C'[z] \lhd I + A'[a] + C'[b]$$

[43]General Non-Extreme Priority universally quantifies over two wellbeing levels in its first string of quantifiers; set both to z.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

Claim: nk is a witness to Restricted Trade-Offs. To see this, split up any group A, consisting of at least kn people, into n k-person groups $A_1, \ldots A_n$, with A_n containing any excess people over kn. Split up any n-group C into single-person groups $C_1, \ldots, C_n$. Rewriting (*), we want to show that

$$I + A_1[z] + \ldots + A_n[z] + C_1[z] + \ldots + C_n[z] \lhd I + A_1[a] + \ldots + A_n[a] + C_1[b] + \ldots + C_n[b]$$

As in the proof of Restricted Inequality-Averse Addition, define, for $0 \leq i \leq n$, D_i to be the left-hand side of (**), with $A_j[z]$ and $C_j[z]$ replaced by $A_j[a]$ and $C_j[b]$ respectively for all $j \leq i$. Thus, D_0 is the left-hand side of (**), while D_n is the right-hand side of (**). Furthermore, General Non-Extreme Priority implies that $D_i \lhd D_{i+1}$ for all $i < n$. The Transitivity Lemma then implies that $D_0 \lhd D_n$, as required.

We now prove the general case. Let z be an arbitrary barely good wellbeing level, and n any number of lives. We shall prove by induction on k that, if some wellbeing level $b < z$[44] is reachable from z by a consecutively close chain of length k, then Sufficient Trade-Offs is true for b, z and n. The base case, in which $k = 1$, follows immediately from the General Non-Extreme Priority condition. Suppose then that our induction hypothesis is true for i, and consider some wellbeing level b which is reachable from z by a consecutively close chain of length $i + 1$. That is, we have a chain

$$z = z_1 > z_2 > \ldots > z_i = b' > z_{i+1} = b$$

From Restricted Trade-Offs applied to b', b and n, we have that there exists some number m and wellbeing level $a' > b'$ such that, for any group C_1 of size at least m, and any group C_2 of size n, and any $a \geq a'$, and any unaffected background population I',

$$I' + C_1[b'] + C_2[b'] \lhd I' + C_1[a] + C_2[b] \qquad (***)$$

By our induction hypothesis applied to z, b' and $m + n$, for some number k and wellbeing level a'', if A is of size at least k and C is of size $m + n$, and if $a \geq a''$, and if I is any unaffected background population, we have

$$I + A[z] + C[z] \lhd I + A[a] + C[b'] \qquad (****)$$

By the Directedness Property, there exists some $a^+ \geq a', a''$. Claim: wellbeing level a^+ and number $k + m$ satisfy Sufficient Trade-Offs for b, z and n. To see this, let C be an arbitrary group of size $m + n$, and A an arbitrary group of size k. Note that we can break down C into groups C_1 and C_2, containing m and n people respectively. Note also that (***) and (****) hold with any $a^{++} \geq a^+$ in place of a.[45] Finally, setting I' to $I + A[a^{++}]$, and applying the Transitivity Lemma on (***) and (****), we have

$$I + A[z] + C[z] \lhd I + A[a^{++}] + C_1[a^{++}] + C_2[b]$$

As required. This completes the induction. Finally, note that the Finite Fine-Grainedness Assumption implies that any wellbeing level $b < z$ is reachable from z by a consecutively close chain of finite length; therefore, Sufficient Trade-Offs holds for any b.

The Aggregation Lemma

Aggregation Lemma The General Non-Elitism Condition, General Non-Extreme Priority and Finite Fine-Grainedness imply

Axiological Aggregation. For any barely good z, any $z^+ > z$ which is not maximal, any $b < z$, and any number of lives n, there is a number of lives m such that if B and Z are groups of n and m people respectively, and I is any unaffected background population, then

$$I + B[z] + Z[z] \lhd I + B[b] + Z[z^+]$$

Proof of the Aggregation Lemma. We may assume Sufficient Trade-Offs and Inequality-Averse Addition. Let $z, z^+ > z, b < z$, and n be arbitrary. By Sufficient Trade-Offs applied to wellbeing levels $z > b$, and number of lives n, there exists a good wellbeing level $a' > z$ and a number m such that for any $a'' \geq a'$, and

[44] This restriction is harmless since a bad wellbeing level must be worse than a good one.

[45] This follows by transitivity, since we then have that $a^{++} \geq a', a''$.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

any group A of size at least m, and group C of size n, and any unaffected background population I,

$$I + A[z] + C[z] \lhd I + A[a''] + C[b] \quad (*)$$

Since z^+ is not maximal, there exists $z^{++} > z^+$. By the Directedness Property, there exists a such that $a \geq a'$, and $a \geq z^{++}$. From Sufficient Trade-Offs we then have, for any unaffected background population I', and any groups C and A of sizes n and at least m respectively,

$$I' + A[z] + C[z] \lhd I' + A[a] + C[b] \quad (**)$$

We now apply Inequality-Averse Addition for wellbeing levels $a > z^+ > z$, and number of lives m, and thus obtain some number k such that, if A and Z are groups containing m and at least k lives respectively, for any unaffected background population I'',

$$I'' + Z[z] + A[a] \lhd I'' + Z[z^+] + A[z^+] \quad (***)$$

Given an arbitrary unaffected background population I, and an arbitrary group C of size n, we now set $I' = I + Z[z]$ and $I'' = I + C[b]$. From (**), (***) and the Transitivity Lemma, it follows that for any groups A and Z of sizes at least m and at least k respectively,

$$I + Z[z] + A[z] + C[z] \lhd I + Z[z^+] + A[z^+] + C[b]$$

Thus, since $A + Z$ is an arbitrary group of size at least $m + k$, $m + k$ is a witness to Axiological Aggregation for z^+, z, b and number of lives n.[46]

A Non-Additive Version of Arrhenius's Sixth Impossibility Theorem

This theorem is identical to the Additive Impossibility Result, with three exceptions:

(i) The Additive Anti-Repugnance Condition is replaced by Arrhenius's "Weak Quality Condition".
(ii) The Non-Additive Weak Non-Sadism Condition is replaced by Weak Non-Sadism.
(iii) The General Non-Extreme Priority condition is slightly strengthened.

The strengthened version of Non-Extreme Priority required is

More General Non-Extreme Priority For any wellbeing level w, and any non-maximal a, there is a good wellbeing level $a' > a$, and a number n, such that for any wellbeing levels $a^+ \geq a'$ and $w^+ > w$, where w^+ is close to w, any group A of size at least n, and single-person group C disjoint from A, and any unaffected background population I disjoint from A and C,

$$I + A[z] + C[w^+] \lhd I + A[a] + C[w]$$

This principle differs from General Non-Extreme Priority only in that it is not required that the potentially large benefits accrue to individuals who have barely good lives. Rather, the large benefits may accrue to individuals who are already well-off. While it is logically possible for General Non-Extreme Priority to be true and More General Non-Extreme Priority to be false, it is difficult to imagine how this could be the case.

By means exactly analogous to those involved in the proof of the Aggregation Lemma of §3.6, one can prove

Aggregation Lemma 2 The General Non-Elitism Condition, More General Non-Extreme Priority and Finite Fine-Grainedness imply

Unrestricted Axiological Aggregation. For any non-maximal a, any non-maximal $a^+ > a$, any $b < a$, and any number of lives n, there is a number of lives m such that if B and Z are groups of n and m people respectively, and I is any unaffected background population, then

$$I + B[a] + Z[a] \lhd I + B[b] + Z[a^+]$$

[46]That is, $m + k$ satisfies the existential claim made by the instance of Unrestricted Axiological Aggregation taking z^+, z, b and n in the places specified by the initial string of universal quantifiers.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

Our (Very) Repugnant Conclusion avoidance condition shall be Arrhenius's

Weak Quality Condition[47] There are some wellbeing levels $a > b$, with a being non-maximal, and number of lives n, such that for any barely good wellbeing level z, any groups A and B, both of size at least n and any group Z,

$$B[b] + Z[z] \lhd A[a]$$

We then have

The Non-Additive Impossibility Result There is no population axiology for which all of the following conditions are satisfied:

(1) Finite Fine-Grainedness.
(2) Acyclicity.
(3) The Weak Quality Condition.
(4) Weak Non-Sadism.[48]
(5) General Non-Elitism.
(6) More General Non-Extreme Priority.

Proof. Let $z^+ > z > z^-$ be barely good wellbeing levels. Let wellbeing level b' and number of lives n' witness Weak Non-Sadism. Let wellbeing levels $a > b''$ (with a non-maximal), and number of lives n'', witness the Weak Quality Condition. The Directedness Property implies that there exists $b \leq b', b''$. Let n be the maximum of n' and n''. b and n thus witness Weak Non-Sadism and, together with a, witness the Weak Quality Condition. Let $a^+ > a$. Let j witness Unrestricted Axiological Aggregation 3 for $a^+ > a$, b, and number of lives n. Let k' witness Inequality-Averse Addition for wellbeing levels $a^+ > z > z^-$, and number of lives j. Let k'' witness Unrestricted Axiological Aggregation for wellbeing levels $z^+ > z > b$, and number of lives n. Finally, let k be the maximum of k' and k''.

Let A be a group consisting of j lives. Let B be a group consisting of n lives. Let Z be a group consisting of k lives.

We then have:

$$A[a^+] + B[b] \lhd A[a^+] + Z[z^-] + B[z^-] \tag{1}$$

$$A[a^+] + Z[z^-] + B[z^-] \lhd A[z] + Z[z] + B[z] \tag{2}$$

$$A[z] + Z[z] + B[z] \lhd A[z^+] + Z[z^+] + B[b] \tag{3}$$

$$A\,[z^+] + Z[z^+] + B[b] \lhd A[a] + B[a] \tag{4}$$

$$A[a] + B[a] \lhd A[a^+] + B[b] \tag{5}$$

The argument is illustrated by Figure 10. (1) follows from Weak Non-Sadism with unaffected background population $A[a^+]$. (2) follows from Inequality-Averse Addition. (3) follows from Unrestricted Axiological Aggregation. (4) follows from the Weak Quality Condition. (5) follows from Unrestricted Axiological Aggregation. Applying the Transitivity Lemma, we get a violation of Acyclicity.

The Very Repugnant Proposition

This argument assumes our Axiological Aggregation condition. It shows that the Weak Quality Condition as defined in the Non-Additive Impossibility Result, which is an avoidance condition for the Very Repugnant Conclusion, implies the apparently stronger Quality Condition, which is an avoidance condition for the

[47]The condition that the wellbeing levels a and b are non-maximal is not present in Arrhenius's statement of the Weak Quality Condition. Additionally, the version of the Weak Quality Condition given here uses the sequential worseness rather than the worseness relation. These differences do not significantly affect the plausibility of the condition.

[48]As in the case of the Weak Quality Condition, we assume a version of Weak Non-Sadism which makes a claim of sequential worseness rather than worseness.

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press

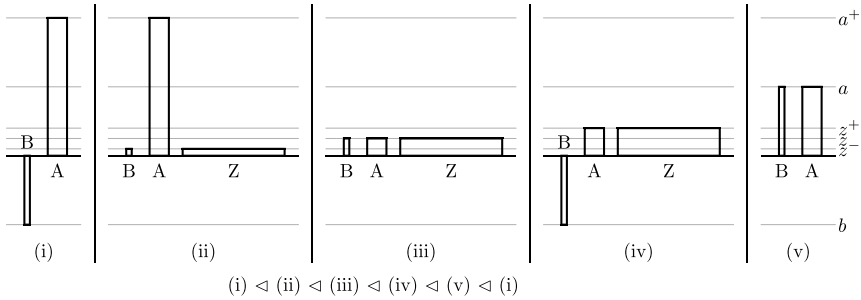


Figure 10. Proof of the Non-Additive Impossibility Result.

regular version of the Repugnant Conclusion. Arrhenius's Quality Condition, adapted for our sequential worseness notation, is defined as follows:

> **Quality Condition** There is a wellbeing level a, a number of lives n and a number of lives m such that, for any barely good wellbeing level z, for any group A of size at least n and any group Z of size at least m,
>
> $$Z[z] \triangleleft A[a]$$

> **The Very Repugnant Proposition** The Weak Quality condition and Axiological Aggregation imply the Quality Condition.

Proof. Let $z > z^-$ be barely good wellbeing levels. Let wellbeing levels $a > b$ and number of people n witness the Weak Quality Condition. That is, for any barely good wellbeing level z', any groups A' and B' of size at least n and any group Z', we have

$$B'[b] + Z'[z] \triangleleft A'[a]$$

Let A and B be two groups of size at least n and let Z be a group containing m people, where m is an arbitrarily large number. We may assume, in particular, that m is large enough to witness Axiological Aggregation for wellbeing levels $z > z^- > b$ and number of lives n. That is, if B' and Z' are groups of n and m people respectively, and I is any unaffected background population, then

$$I + B'[z^-] + Z'[z^-] \triangleleft I + B'[b] + Z'[z]$$

We then have:

$$B[z^-] + Z[z^-] \triangleleft B[b] + Z[z] \tag{1}$$

$$B[b] + Z[z] \triangleleft A[a] \tag{2}$$

$$B[z^-] + Z[z^-] \triangleleft A[a] \tag{3}$$

(1) follows from Axiological Aggregation, setting the unaffected background population to be empty. (2) follows from the Weak Quality Condition. (3) follows from (1) and (2), applying the Transitivity Lemma. (3) is then a witness to the Quality Condition, since group Z (and thus, group $B + Z$) was chosen so that it can be arbitrarily large.

Given the Aggregation Lemma, we have the following corollary:

> **The Very Repugnant Proposition (2)** The Weak Quality Condition, General Non-Elitism, General Non-Extreme Priority and Finite Fine-Grainedness imply the Quality Condition.

Cite this article: Francis T (2026). Repugnance without Mere Addition. *Economics and Philosophy* **42**, 645–673. https://doi.org/10.1017/S0266267125100631

https://doi.org/10.1017/S0266267125100631 Published online by Cambridge University Press