16
Population Ethics and Conflict-of-Value Imprecision

Gustaf Arrhenius

16.1 Introduction

Derek Parfit's infamous "Repugnant Conclusion" can be stated as follows:

> *The Repugnant Conclusion*: For any population consisting of people with very high positive welfare, there is a better population in which everyone has very low positive welfare, other things being equal.[1]

In Figure 16.1, the width of each block represents the number of people whereas the height represents their lifetime welfare. Dashes indicate that the block in question should be much wider than shown, that is, the population size is much larger than shown.

These populations could consist of all the past, present, and future lives (a possible world), or all the present and future lives, or all the lives during some shorter time span in the future such as the next generation, or all the lives that are causally affected by, or consequences of, a certain action or series of actions, and so forth.[2]

[1] Here's how Parfit (1984), p. 388, formulates the conclusion: "For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living." Hence, my formulation is more general than his. The *ceteris paribus* clause in the formulation is meant to imply that the compared populations are roughly equal in all other putatively axiologically relevant aspects apart from individual welfare levels. Although it is through Parfit's writings that this implication of Total Utilitarianism has become widely discussed, it was already noted by Henry Sidgwick (1907), p. 415, in the first edition 1874. For other early sources of *the Repugnant Conclusion*, see McTaggart (1927), pp. 452–3; Narveson (1967); and Broad (1979), pp. 249–50.

[2] More exactly, a population is a finite set of lives in a possible world. A, B, C, ... A_1, A_2, ..., A_n, $A \cup B$, and so on, denote populations of finite size. We shall adopt the convention that populations represented by different letters, or the same letter but different indexes, are pairwise disjoint. For example, $A \cap B = A_1 \cap A_2 = \emptyset$. We shall assume that for any natural number n and any welfare level X, there is a possible population of n people with welfare X (for a discussion of this *No-Limit Assumption*, see Arrhenius (2000b), ch. 3, (forthcoming)).

Gustaf Arrhenius, *Population Ethics and Conflict-of-Value Imprecision* In: *Ethics and Existence: The Legacy of Derek Parfit*. Edited by: Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan, Oxford University Press.
 DOI: 10.1093/oso/9780192894250.003.0017

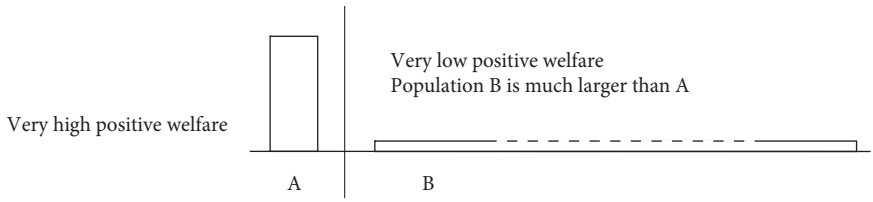

Figure 16.1

All the lives in Figure 16.1 have positive welfare, or, as we also could put it, all the people have lives worth living. The A-people have very high welfare whereas the B-people have very low positive welfare.[3] The reason for this could be that in the B-lives there are, to paraphrase Parfit, only enough ecstasies to just outweigh the agonies, or that the good things in those lives are of uniformly poor quality, e.g., eating potatoes and listening to Muzak.[4] However, since there are many more people in B, the total sum of welfare in B is greater than in A. Hence, a theory like Total Utilitarianism, according to which we should maximize the welfare in the world, ranks B as better than A—an instance of *the Repugnant Conclusion*.[5]

Notice that problems like *the Repugnant Conclusion* are not just problems for total utilitarians or those committed to welfarism, the view that welfare is the only value that matters from the moral point of view, since we have assumed that other axiologically relevant aspects are roughly equal. Hence, other values and considerations are not decisive for the value comparison of populations A and B. Thus, *the Repugnant Conclusion* is a problem for all moral theories according to which welfare matters at least when all other things are equal, which arguably is a minimal adequacy condition for any moral theory.

The Repugnant Conclusion highlights a problem in an area that has become known as *population ethics*. It involves foundational questions regarding axiology and our duties to future generations. The main problem in population ethics has been to find an adequate theory about the value of outcomes where the

[3] We shall say that a life has *neutral welfare* if and only if it is equally as good for the person living it as a neutral welfare component, and that a life has *positive* (*negative*) welfare if and only if it has higher (lower) welfare than a life with neutral welfare. A welfare component is neutral relative to a certain life x if and only if x with this component has the same welfare as x without this component. A hedonist, for example, would typically say that an experience which is neither pleasurable nor painful is neutral in value for a person and as such doesn't increase or decrease the person's welfare. The above definition can of course be combined with other welfarist axiologies, such as desire satisfaction and objective list theories. Moreover, there are a number of alternative definitions of a neutral life in the literature, many of which would also work fine in the present context. For a discussion, see Parfit (1984), pp. 357–8 and appendix G; Broome (1999), (2004); Arrhenius (2000b), (forthcoming), chs. 2 and 9); and Bykvist (2007), p. 101. Notice also that we actually don't need an analysis of a neutral welfare in the present context but rather just a criterion, and the criterion can vary with different theories of welfare.

[4] See Parfit (1984), p. 388, and (1986), p. 148. For a discussion of different interpretations of *the Repugnant Conclusion*, see Parfit (1984), (2014), (2016); and Arrhenius (2000b), (forthcoming).

[5] Throughout this chapter, "better" means "better, all things considered" if not otherwise indicated.

number of people, the quality of their lives, and their identities may vary. Since any reasonable moral theory has to take these aspects of possible outcomes into account when determining the normative status of actions, the study of population ethics is of general import for moral theory. Through his pioneering and seminal contributions, Parfit can rightly be said to be the founding father of this important field.[6]

As the name indicates, Parfit finds *the Repugnant Conclusion* unacceptable and most philosophers seem to agree. However, it has been surprisingly difficult to find a theory that avoids *the Repugnant Conclusion* without implying other very counterintuitive conclusions.[7] Actually, it is impossible to avoid *the Repugnant Conclusion* (or even worse conclusions) without violating some intuitively very convincing conditions. We know this for sure through a number of so-called impossibility theorems.[8] The proofs of these theorems show that there is no theory that fulfils a number of intuitively compelling adequacy conditions—conditions which everyone seems to agree that a reasonable moral theory must fulfil. Examples of such conditions are that one future is better than another if everyone is better off in the former as compared to the latter, or that it is better to create happy rather than unhappy people. The question as to how *the Repugnant Conclusion* should be dealt with has become one of the cardinal challenges of modern ethics and the inquiry into what it shows about the nature of ethics has opened up many new avenues for research. Population ethics has proved a very fruitful area of research, having implications for all areas of moral and political philosophy. This is all thanks to Parfit's ground-breaking work in this area.

Parfit has suggested a novel way of avoiding *the Repugnant Conclusion* by introducing what he calls "imprecision" in value comparisons.[9] He suggests that in a range of important cases, outcomes are only imprecisely comparable. In such cases, transitive relations such as "equally as good as" are not applicable. Instead, we have to make use of imprecise relations that are non-transitive. This imprecision is not due to any cognitive or epistemic limitations but a fact about the value comparisons of certain types of outcomes.

In his Rolf Schock Prize Lecture, he suggested that "[w]hen two possible worlds would contain different numbers of people, this fact makes these worlds less precisely comparable."[10] From this "Different-Number-Based Imprecision", as he called it, follows that many of the comparisons of different future populations will involve imprecise comparisons and transitivity of the involved relations might fail. Parfit suggests that this feature will open up a way of avoiding the Repugnant

[6] Another pioneer is Jan Narveson. See Narveson (1967), (1973), (1978).

[7] For a summary, see Arrhenius (2000b), (2013b), (forthcoming); Arrhenius, Ryberg, & Tännsjö (2014).

[8] See, e.g., Parfit (1984), (1986); and Arrhenius (2000b), (2000a), (2003), (2004), (2009b), (2011), (forthcoming).

[9] Parfit (2014), (2016).

[10] Parfit (2014).

Conclusion without implying other very counterintuitive conclusions, or at least less counterintuitive than *the Repugnant Conclusion*, and thus solve one of the major challenges in ethics.

Parfit and I had many exchanges about this proposal in which I, among other things, tried to convince him to instead go for what I called "Conflict-of-Value Imprecision" (to be explained below). Our presentations at the *Rolf Schock Prize Symposium in Logic and Philosophy in honour of Derek Parfit* were scheduled to appear in the same special issue of the journal *Theoria*. Mine, focusing on the idea of Different-Number-Based Imprecision, appeared unaltered but at the last minute Parfit changed his view and, surprisingly, went for Conflict-of-Value Imprecision instead of Different-Number-Based Imprecision.[11] So in the charming absent-minded way so characteristic of him, he completely wrongfooted me. As he graciously wrote to me later, he had had to change his talk (of which his article was a transcript) because of the criticism in my paper. I'm happy but also sad now. Parfit has left us and we no longer can have those delightful and insightful exchanges that I miss so much. But I'm happy to be able at last to reply to his revised proposal and consider whether it will help us with the paradoxes in population ethics, as he had hoped.

16.2 Imprecision and Conflict-of-Value Imprecision

Here's an example of what Parfit means by imprecision in value comparisons:

> Suppose that I ask you whether Einstein or Bach was a greater genius, or achieved more. You may first assume that this question couldn't have an answer, since it makes no sense to compare the genius, or achievements, of scientists and composers. But I might then point out that Bach was clearly a greater genius than many bad scientists, and Einstein was a greater genius than many bad composers. When you realize that there *can* be truths of this kind, you would not suddenly come to believe that as geniuses, or in their achievements, Einstein and Bach might be precisely equally as great. As you would see, the truth could be only that one of these people was imprecisely greater than the other, or more plausibly, that they were imprecisely equally as great.[12]

"Equally as good as" is transitive: If A is equally as good as B, and B is equally as good as C, then A is equally as good as C. "Imprecisely equally as good as", however, is non-transitive: Even if A is imprecisely equally as good as B, and B is imprecisely equally as good as C, A might not be imprecisely equally as good as C. For example, that Einstein and Mozart are imprecisely equally as good and Mozart

[11] Arrhenius (2016); Parfit (2016). [12] Parfit (2014).

and Bohr are imprecisely equally as good is compatible with Einstein being greater than Bohr. According to Parfit, "[s]uch imprecision is not the result of our lack of knowledge, but is part of what we would know if we knew the full facts."[13]

That there are non-transitive value relations of this kind seems likely. A well-known case is "not better than". As Parfit puts it:

> If your life could go in different ways, it might be true that your being a writer would not be better than your being a doctor, which would not be better than your being a slightly less successful writer. But your being a writer would be better than your being a slightly less successful writer. Not better than would not here be a transitive relation.[14]

But how should this help us with the paradoxes in population ethics? Parfit adds another important assumption which I shall call "Conflict-of-Value Imprecision": "We might claim that ... given the conflict between ... values, [w]orlds are only imprecisely comparable, and would be imprecisely equally good."[15] So the idea is that imprecision may arise from weighing different values against each other, such as perfectionist values versus total welfare or equality. The values Parfit has is mind are the quality of people's lives and the quantity of welfare:

> we can ... distinguish between the quality of people's lives and the quantity of well-being per person. These might diverge. The best things in your life might be of a higher quality than the best things in mine, and your life might go worse than mine only because you would have many fewer of these best things.[16]

What does Parfit have in mind when he talks about "quality of life" and the "the best things in life"? He doesn't say so much about it in Parfit (2016) but in a comment on one of the cases he discusses, he clarifies it a bit:

> Some people in Alpha would have a much higher quality of life than anyone in Y, since this quality would be at level 200 rather than at level 2. This higher quality of life, we should assume, would not be merely a difference in the amount of well-being per person. At level 200, the best things in life would be very good, and lives at level 2, in World Y, would not include any of these good things. *There would be no art, or science, no deep loves or friendships, no other achievements, such as that of bringing up our children well, and no morally good people*. World Y would be much worse than Alpha in what we can call *qualitative or perfectionist terms*. In

[13] Parfit (2014). His idea is similar to, but not the same as, Ruth Chang's (2002), (2005) proposal that there is a forth value-relation: "on a par". On a par is also a non-transitive relation.
[14] Parfit (2014). [15] Parfit (2016), p. 126. [16] Parfit (2016), p. 126.

one version of this case, lives at levels 1 and 2 would be like the lives of never-developing one-year-old and two-year-old children.[17]

Much more needs to be said about Parfit's idea of the quality of welfare components and how they contribute to the welfare of a life and the value of a population. However, the above will suffice for our discussion here I think. Parfit's central idea, as I take it, is that such qualitative difference in welfare components in people's lives contributes to the value of worlds or populations in addition to how they contribute to people's welfare, and might make populations imprecisely comparable.

This source of imprecision is very much along the lines of what I suggested to Parfit in our discussion: It is much more likely with imprecision arising from weighing different values against each other than from populations differing in size. Importantly, with Conflict-of-Value Imprecision we could get imprecision also in same number cases, not only in different number cases as with Different-Number-Based Imprecision. Hence, this imprecision could potentially undermine the same number conditions in the impossibility theorem such as Non-Elitism and Non-Extreme Priority (more on this below).

How could Conflict-of-Value Imprecision help us with the impossibility results in population ethics? Let's first look at some derivations of *the Repugnant Conclusion* and see whether imprecision can help us blocking some step in them.

16.3 The Quantity Sequence

There are other axiologies apart from Total Utilitarianism that imply *the Repugnant Conclusion*. These can be characterized by a set of conditions. Consider the following condition:

Quantity: For any pair of positive welfare levels $\mathbb{A}$ and $\mathbb{B}$, such that $\mathbb{B}$ is slightly lower than $\mathbb{A}$, and for any number of lives n, there is a greater number of lives m, such that a population of m lives at level $\mathbb{B}$ is better than a population of n lives at level $\mathbb{A}$, other things being equal.[18]

Quantity has some intuitive plausibility and should appeal to those thinkers that find some truth in the saying "the more good, the better". However, it implies

[17] Parfit (2016), p. 123 (my emphasis). On p. 118 he gives the example of "the earliest sentient animals who had lives that were just worth living, because these animals had enough slight pleasures like those of cows munching grass or lizards basking in the sun".

[18] A welfare level is an equivalence class on the set of all possible lives with respect to the relation "has at least as high welfare as". For an exact statement of this principle, see Arrhenius (2000b), (forthcoming), where this condition is formulated in terms of "at least as good as".

the Repugnant Conclusion together with the following reasonable assumption regarding welfare levels:

> *Finite Fine-grainedness*: There exists a finite sequence of slight welfare differences between any two welfare levels.

The idea here is that one can get from one welfare level to another in a finite number of steps of intuitively slight welfare difference. Examples of such welfare differences could be some minor pain or pleasure or a shortening of life by a minute or two.[19] These differences don't have to be of the same size or type. Let's say that a life of type *a* has higher welfare than a life of type *b*, and suppose that you are successively making *a* slightly worse, perhaps by shortening it by a minute or two or by adding some minor pain. *Finite Fine-grainedness* implies that there is a finite (but possibly great) number of such slight worsenings from *a* to another type of life *c* such that a life of this type will have the same welfare or lower welfare than a life of type *b*. It is quite hard to deny the intuitive force of this assumption.[20]

Consider the sequence in Figure 16.2 of populations for an informal demonstration that these two conditions together imply *the Repugnant Conclusion*:[21]

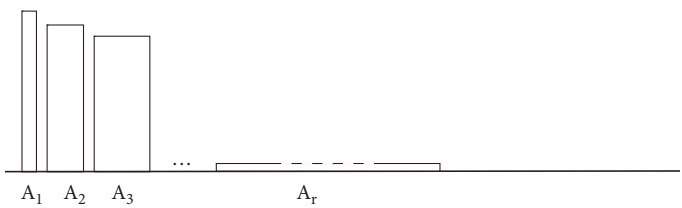

Figure 16.2 The Quantity Sequence

Assume that A_1 in Figure 16.2 is a population with very high welfare and that A_r is a population with very low positive welfare (again, the width of the blocks represents the number of lives in the population, the height represents their lifetime welfare; dashes indicates that the block in question should intuitively be

[19] For a precise definition of "slight welfare difference", see Arrhenius (2000b), (2011), (forthcoming).

[20] Notice that Finite Fine-grainedness doesn't imply that all sequences of slight welfare differences between two welfare levels are finite, just that there exists at least one such sequence. It is compatible with the welfare ordering being continuous as well as discreet. It just rules out that there are, so to speak, big "jumps" or "holes" in the order of welfare levels. For a discussion of Finite Fine-grainedness and possible theories of welfare that violate this condition, see Arrhenius (2005), (forthcoming); Arrhenius & Rabinowicz (2015). For an interesting effort to challenge Finite Fine-grainedness (in light of the impossibility theorems in population ethics), see Thomas (2018) and Carlson (forthcoming). Actually, a complete discussion of Parfit's new ideas on imprecision would have to involve a more detailed discussion of Finite Fine-grainedness but due to space restrictions, this will have to wait to another time.

[21] For an exact proof, see Arrhenius (2000b), (forthcoming). Parfit (2016) informally derives *the Repugnant Conclusion* with this kind of sequence argument.

much wider than shown). According to *Quantity*, there is a population A_2 with slightly lower welfare than A_1 and which is better than A_1; a population A_3 with slightly lower welfare than A_2 and which is better than A_2; and so forth. We can assume that the welfare levels in this sequence of populations satisfy Finite Fine-grainedness. Hence, we will finally reach population A_r with very low positive welfare. By transitivity, A_r is better than A_1. Since A_1 is an arbitrary population with very high welfare, this shows that for any population with very high welfare, there is a population with very low positive welfare which is better, that is, *the Repugnant Conclusion*.

Here's Parfit suggestion for how an appeal to imprecision can block this derivation:

> [This] assume that any slight loss in the quality of people's lives could be outweighed by a sufficient gain in the number of people who would exist and have lives that would be slightly less worth living. As we can more briefly say, any slight loss of quality could be outweighed by a sufficient gain in quantity. If we assumed precision, it would be hard to reject these arguments.—But we should deny that such truths would be precise . . . It would not be better if there existed many more people whose quality of life would all be lower, since two such worlds would at most be imprecisely equally good. Though the larger of these worlds would not be worse, this relation is not transitive. So we could claim that it would be worse if, in other, larger worlds, everyone's quality of life would be much lower.[22]

So the idea is that A_2 is not better than A_1 but only imprecisely equally as good "since two such worlds would at most be imprecisely equally good". Likewise for A_3 and A_2, and so on. Since "imprecisely equally as good as" is non-transitive, *the Repugnant Conclusion* doesn't follow. Actually, it doesn't even follow that A_1 is imprecisely equally as good as A_r which would be a slightly weaker version of *the Repugnant Conclusion*. This result is compatible with A_1 being better than A_r, which is what most people seem to believe. This appeal to imprecision implies, of course, a rejection of Quantity. However, it is compatible with a weaker version formulated in terms of "imprecisely equally as good as".

Is this argument convincing? Well, given that we can claim that there is a loss of the best things in life when we move down the sequence, this might work. Not obvious, however, since it is not clear why there couldn't be the same amount of the best things in life in A_1 and A_r but much more pain and suffering in A_r (more on this below). Moreover, there are other derivations of *the Repugnant Conclusion* that are trickier to deal with.

[22] Parfit (2016), p. 120.

16.4 The Sequential Dominance Addition Paradox

The next derivation is based on two principles. Here's the first one:

> *Dominance Addition*: An addition of lives with positive welfare and an increase in the welfare of everyone in the original population makes a population better, other things being equal.[23]

The idea is that you don't make a population worse by adding lives worth living and increasing the welfare of the individuals in the original population. It is a logically weaker and intuitively more compelling version of the more well-known *Mere Addition Principle*: An addition of people with positive welfare does not make a population worse, other things being equal.[24] Yet, although it might seem a compelling principle at first glance, it is controversial. Several authors have rejected it, some for quite good reasons.[25] One might, for example, object to it on egalitarian grounds since a mere addition can introduce great inequality in an otherwise perfectly equal population.[26] Likewise for Dominance Addition albeit then the disvalue of the introduced inequality also has to be weighed against the positive value of the increased welfare of the lives in the original population, not only against the possible positive value of more lives with positive welfare. However, we shall set this aside for now since we are here interested in how imprecision might help us with the paradoxes in population ethics.

The next condition is a weak egalitarian condition:

> *Inequality Aversion*: For any triplet of welfare levels $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{C}$, $\mathbb{A}$ higher than $\mathbb{B}$, and $\mathbb{B}$ higher than $\mathbb{C}$, and for any population A with welfare $\mathbb{A}$, there is some larger population C with welfare $\mathbb{C}$ such that a perfectly equal population B of the same size as A∪C and with welfare $\mathbb{B}$ is better than A∪C, other things being equal.[27]

Another way of stating Inequality Aversion is to say that for any welfare level of the best off and worst off, and for any number of best off lives, there is some

[23] See Arrhenius (2000b), (forthcoming).

[24] Cf. Hudson (1987); Ng (1989); Sider (1991); and Parfit (2014), p. 420ff, Cf. fn. below.

[25] Ng (1989), p. 244; Blackorby, Bossert, & Donaldson (1995), p. 1305; Blackorby, Bossert, & Donaldson (1997), pp. 210–11; and Fehige (1998). Ng (1989), p. 238, ascribes to Parfit the view that a population axiology should satisfy the Mere Addition Principle, and one might get that impression from Parfit (2014), p. 420ff. In personal communication, however, Parfit has expressed doubts about the Mere Addition Principle in cases where the added people are much worse off than the rest of the population. See also Kavka (1982); Feldman (1997) ch. 10; and Carlson (1998), pp. 288–9.

[26] See Arrhenius (2009a), (2013a), (forthcoming).

[27] For an exact statement of this principle, see Arrhenius (2000b), (forthcoming), where this condition is formulated in terms of "at least as good as". I've here formulated it in terms of "better than" to simplify the exposition.

(possibly much) greater number of worst off lives such that it would be better to have an equal distribution of welfare on any level higher than the worst off, other things being equal.

It is a very weak egalitarian condition since it can be satisfied by a theory which demands that the total welfare must be greater for a population with perfect equality to be better than an unequal population of the same size. Moreover, it is also compatible with principles that give much greater weight to the welfare of the best off as compared to the welfare of the worst off. For example, a theory which requires that to compensate for one person falling from twenty to ten units of welfare, a hundred people have to be moved from zero to ten units, is compatible with Inequality Aversion. In that sense, its name is a bit misleading since it is compatible with quite non-egalitarian theories. Roughly, Inequality Aversion only rules out theories that imply that we should always or sometimes give some kind of "lexical priority" the best off.[28] A simple example of such a theory is "Maximax": Maximize the welfare of the best off.

Now consider the populations in Figure 16.3. All the people in population A enjoy very high welfare. In A+, we have added a second group of lives with positive welfare a bit lower than the lives in A, and increased the welfare of the original group in A. In B, which is of the same size as A+, we have equalized the welfare at a level higher than the +-lives but lower than the A-lives. We can assume that A+ and B fulfil the antecedent of Inequality Aversion.[29]

The Dominance Addition Principle implies that A+ is better than A. Inequality Aversion implies that B is better than A+. Likewise for populations B, B+, and C, and so forth until we finally reach population Z with very low positive welfare. By transitivity, Z is better than A, that is, *the Repugnant Conclusion*.

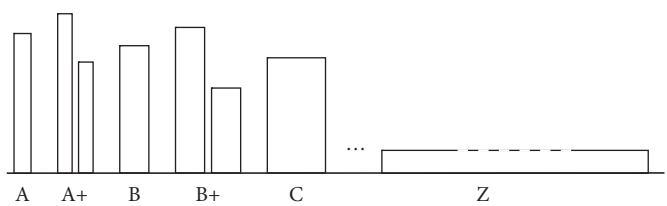

Figure 16.3 The Sequential Dominance Addition Paradox

[28] There are some more subtle theories that violate Inequality Aversion, such as theories that invoke some form of superiority in value. See Arrhenius (2005) and Arrhenius & Rabinowicz (2005), (2015), for a discussion. As we shall discuss below, Inequality Aversion can be derived from an even more intuitively compelling condition, Non-Elitism.

[29] If welfare is measurable on at least an interval scale, we could also assume that the total and average welfare in B is higher than in A+.

Here's Parfit's take on this case:

> Though World B would be better than [A+] in utilitarian and egalitarian terms, B would be worse in qualitative terms, *since the best things in people's lives would be worse in B*. We might claim that (K) *given the conflict between these values*, Worlds B and [A+] are only *imprecisely comparable*, and would be imprecisely equally good. This claim would not be in itself plausible.—(K) seems implausible because, in a change from [A+] to B, there would be only a slight qualitative loss. The best lives would fall only from level 101 to 99. It may be hard to believe that this slight qualitative loss could make [A+] not better than B, but only imprecisely equally good. But it would be much harder to believe that, compared with the existence of many people whose quality of life would be very high, it would be better if there existed instead some vast number of people whose lives were barely worth living.[30]

So the idea is that B is not better than A+ but, because of Conflict-of-Value Imprecision, B is instead imprecisely equally as good as A+. Given this, transitivity fails and we cannot derive that Z is better than A. Hence, in this case imprecision makes it possible for us to hold onto Dominance Addition but still avoid *the Repugnant Conclusion*. We can still hold on to the belief that A is better than Z. This appeal to imprecision implies, of course, a rejection of Inequality Aversion. However, it is compatible with a weaker version formulated in terms of "imprecisely equally as good as". We would also have to accept (K) which is quite counterintuitive but Parfit's idea is that this is less implausible than accepting *the Repugnant Conclusion*.[31]

Here's a first problem with this line of argument, however. Parfit claims that "the best things in people's lives would be worse in B". But the best things in life could be better for the extra people in a move to B (which explains their rise in welfare from 95 to 99). Moreover "in a change from [A+] to B, there would be only a slight qualitative loss" since "[t]he best lives would fall only from level 101 to 99". So in this case we have to weigh a small decrease in the quality of the best thing in life for the best off, against both a bigger increase in the quality of best things in life for the worst off, and a total and average increase of the best things in life. Hence, there doesn't seem to be a conflict of different values in this case.

Here's a second problem. Pace Parfit, it could be the same quality and amount of the best things in life in A+ and B but the bad things (pains, suffering) are more equally distributed in B. Hence, the same quality and amount of the best things in

[30] Parfit (2016), emphasis added.

[31] Of course, even if we grant Parfit that (K) is less implausible than *the Repugnant Conclusion*, that doesn't suffice to show that of all the counterintuitive conclusions we could accept, we should accept (K). That is, without further argument, it isn't clear that the view entailing (K) is the least counterintuitive way of avoiding *the Repugnant Conclusion*. Thanks to Jimmy Goodrich for drawing my attention to this point.

each life in A+ and B but in A+ there is an unequal distribution of the bad things whereas in B, these are equally distributed. And it is this that explain the different welfare pattern in A+ and B. Again, there would be no conflict of value.

A rejoinder would be to claim that this is possible when comparing A+ and B but not further down the sequence. To get to a life barely worth living, the quality and amount of the best things in life have to go down at some point, one might argue.

This raises an essential question for the appeal to Conflict-of-Value imprecision in population ethics that Parfit unfortunately never discusses: Could the best things in life be the same in a life with very high positive welfare and a life with very low positive welfare?

I think this is true since we are discussing lifetime welfare and one could just add more bad things to a life to move it from very high welfare to very low positive welfare (or further down) albeit with the same best things as a life with very high positive welfare. For example, to a life which is splendid for the first sixty years, involving the best things in life, we could add horrible suffering for the next twenty years. It seems intuitive that such a life would have very low positive lifetime welfare although it involves the best things in life.

Actually, Parfit agrees that such lives are possible and seems to have realized that this possibility limits the reach of an appeal to imprecision as a solution to the paradoxes when he writes:

> in Roller-Coaster Z, everyone would live as long as everyone in World A, and all of the good things in these people's lives would be just as good, but these people's lives would be barely worth living because their lives would also contain much that was very bad. This version of Z also raises questions that I shall not discuss here.[32]

One possibility is that Parfit's idea was that Conflict-of-Value Imprecision could only help us with one special version of *the Repugnant Conclusion*, namely one that involves "Drab Z, [in which] there would be nothing in people's lives that would be bad, but there would also be very little that was good. The only good features . . . might be muzak and potatoes" or some versions of this kind of Z-population, such as the ones that involve "the earliest sentient animals who had lives that were just worth living, because these animals had enough slight pleasures like those of cows munching grass or lizards basking in the sun."[33] He might then argue that other versions of *the Repugnant Conclusion*, albeit counterintuitive, are "significantly less repugnant" and that we could accept these versions of *the Repugnant Conclusion*. He never claims this about Roller-Coaster Z but he does

[32] Parfit (2016), p. 118. [33] Parfit (2016), p. 118.

make that claim about "Short-Lived Z, [in which] our imagined people would live for only as long as some flowers bloom".[34]

There is, however, a problem with this solution to the Drab Z-version of *the Repugnant Conclusion*.[35] If we accept the Roller-Coaster Z-version of *the Repugnant Conclusion*, we cannot avoid the Drab Z-version without violating an adequacy condition which is as uncontroversial as it gets in population ethics. Consider a Roller-Coaster Z with very low positive welfare. We can then take a Drab Z version in which everyone also has very low positive welfare but everyone is better off than in the Roller-Coaster Z. It is very hard to deny that the former population is better than the latter and it follows from the following uncontroversial condition:

> *Egalitarian Dominance*: If population A is a perfectly equal population of the same size as population B, and every person in A has higher welfare than every person in B, then A is better than B, other things being equal.

Hence, if an appeal to Conflict-of-Value Imprecision cannot block a derivation of the Roller-Coaster Z-version of *the Repugnant Conclusion*, then it cannot block the Drab-Z-version.

One could try to deny that there exists a Drab Z in which everyone has higher welfare than in Roller-Coaster Z but that would be very counterintuitive for the same reasons that we discussed above with respect to the question whether the best things in life could be the same in a life with very high positive welfare and a life with very low positive welfare. Just consider the possibility of a Roller-Coaster Z- in which everyone has slightly negative welfare since the bad things just outweighs the very good things in life. Such a Z is surely worse than Drab Z. Then consider a version of Roller-Coaster Z- where the lives are slightly improved by some extra happy days so that all lives are just barely worth living. Surely such a population could be worse than a version of Drab Z.[36]

16.5 Non-Elitism

In the Sequential Dominance Addition Paradox we made use of Inequality Aversion and Parfit's way out was to deny this condition by an appeal to imprecision. The application of this condition may involve comparisons of a few great losses against a greater number of small gains—another intuitive source for imprecision. However, that wasn't how we applied Inequality Aversion in the Sequential Dominance Addition Paradox. We just had roughly equally big gains

[34] Parfit (2016), p. 118.
[35] I'm grateful to Tim Campbell for pointing this out.
[36] Cf. fn. 20.

and losses throughout the sequence. This shows that the paradox can be derived with a weaker condition that never involves comparisons of a few greater losses against a greater number of small gains. Hence, the Sequential Dominance Addition Paradox doesn't need to involve comparisons that could work as a source of imprecision. Here's the condition:

> *Non-Elitism*: For any triplet of welfare levels $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{C}$, $\mathbb{A}$ slightly higher than $\mathbb{B}$, and $\mathbb{B}$ higher than $\mathbb{C}$, and for any one-life population A with welfare $\mathbb{A}$, there is a population C with welfare $\mathbb{C}$, and a population B of the same size as A∪C and with welfare $\mathbb{B}$, such that, for any population X consisting of lives with welfare ranging from $\mathbb{C}$ to $\mathbb{A}$, B∪X is better than A∪C∪X, other things being equal.[37]

The intuition which this fairly densely formulated condition tries to capture is simply and roughly that there is some (possibly great) number of worst off people such that a slight decrease in welfare (from $\mathbb{A}$ to $\mathbb{B}$) for *one* of the best off persons can be compensated for by an at least as great increase in welfare (from $\mathbb{C}$ to $\mathbb{B}$) for all those worst off people to the effect that the involved people enjoy the same level of welfare ($\mathbb{B}$).

Figure 16.4 provides an illustration. It shows two populations of the same size. Population A consists of a number of best off lives, $a_1, a_2, \ldots, a_n$, and a number of worst off *groups* of lives, $\alpha_1, \alpha_2, \ldots, \alpha_m$. In population B, one of the best off lives (a_1) has been replaced with a life (b_1) enjoying welfare just lower than the welfare of the best off. Moreover, one of the worst off groups of lives (α_1) has been replaced by a same sized group of lives (β_1) with the same welfare as life b_1. Population B is then better than A according to Non-Elitism.

By repeated application, Non-Elitism will yield that B is better than A+ in the Sequential Dominance Addition Paradox.[38] Hence, since the application of

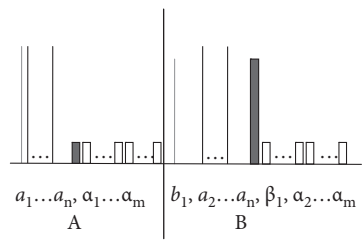

Figure 16.4

[37] For an exact statement of this principle, see Arrhenius (2000b), (2011), (forthcoming), where this condition is formulated in terms of "at least as good as". I've here formulated it in terms of "better than" to simplify the exposition.

[38] This result presumes Finite Fine-grainedness. Actually, Inequality Aversion can be derived from Non-Elitism, see Arrhenius (2000b), (2003), (2011), (forthcoming).

Non-Elitism doesn't involve any comparisons of a few great losses against a greater number of small gains, an appeal to imprecision based on such comparisons won't help here. In Non-Elitism, the gains and losses for each involved individual is either of roughly the same size, or the gains for the worst off individuals are greater than the slight loss for the best off individual.

Another related point is that it would be counterintuitive if the zone of imprecision wasn't limited but "greedy".[39] Even if (and that is indeed a big "if" in this case) we could get imprecision when we weigh one best off person slight loss against one worst off person's gain of roughly the same size or greater, we should get out of the zone of imprecision at some point if we increase the number of people that gain. Non-Elitism is compatible with this idea since it talks about some, possibly great number of worst off people such that a slight decrease in welfare for *one* of the best off persons can be compensated for by an at least as great increase in welfare for the, possibly many more, worst off people.[40]

References

Arrhenius, G. (2000a). An Impossibility Theorem for Welfarist Axiologies. *Economics and Philosophy*, *16*(02), 247–66.

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

Arrhenius, G. (2003). The Very Repugnant Conclusion. In *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Åqvist* (Vol. 51, pp. 167–80). Uppsala: Department of Philosophy, Uppsala University: Uppsala Philosophical Studies.

Arrhenius, G. (2004). The Paradoxes of Future Generations and Normative Theory. In Jesper, Ryberg & T. Tännsjö (Eds.), *The Repugnant Conclusion: Essays on Population Ethics* (pp. 201–18). Dordrecht: Kluwer Academic Publishers.

Arrhenius, G. (2005). Superiority in Value. *Philosophical Studies*, *123*(1/2), 97–114. https://doi.org/10.1007/s11098-004-5223-0.

Arrhenius, G. (2009a). Egalitarianism and Population Change. In A. Gosseries & L. Meyer (Eds.), *Intergenerational Justice* (1st edition, pp. 325–49). Oxford: Oxford University Press.

[39] Broome (2004), ch. 12, uses the term "greedy" to describe incomparability that spreads way beyond its intuitive limitations. See also Handfield & Rabinowicz (2018); Herlitz (2020); Arrhenius (forthcoming), ch. 9.

[40] I would like to thank Andrea Asker, Krister Bykvist, Jimmy Goodrich, Anders Herlitz, Jeff McMahan, Julia Mosquera, Joe Roussos and especially Tim Campbell for helpful discussions. Thanks also to the audiences at the Moral Philosophy Seminar, Oxford, 16 May, 2016; Franco-Swedish Program in Philosophy and Economics Final Conference, Uppsala, 8 June, 2015; and the workshop Incommensurability: Vagueness, Parity and other Non-Conventional Comparative Relations, IFFS, Stockholm, 6–7 December 2019, for useful questions and comments. Financial support from Riksbankens Jubileumsfond (grant M17-0372:1) is gratefully acknowledged.

Arrhenius, G. (2009b). One More Axiological Impossibility Theorem. In L.-G. Johansson, J. Österberg, & R. Sliwinski (Eds.), *Logic, Ethics and All That Jazz: Essays in Honour of Jordan Howard Sobel* (Vol. 57, pp. 23–37). Uppsala: Department of Philosophy, Uppsala University: Uppsala Philosophical Studies.

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

Arrhenius, G. (2013a). Egalitarian Concerns and Population Change. In O. Frithjof Norheim, N. Eyal, S. A. Hurst, & D. Wikler (Eds.), *Inequalities in Health: Concepts, Measures, and Ethics* (pp. 74–91). Oxford: Oxford University Press.

Arrhenius, G. (2013b). Repugnant Conclusion. In *The International Encyclopedia of Ethics* (pp. 4560–63). Oxford: Blackwell Publishing Ltd.

Arrhenius, G. (2016). Population Ethics and Different-Number-Based Imprecision. *Theoria*, *82*(2), 166–81. https://doi.org/10.1111/theo.12094.

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

Arrhenius, G., & Rabinowicz, W. (2005). Millian Superiorities. *Utilitas*, *17*(2), 127–46. https://doi.org/10.1017/S0953820805001494.

Arrhenius, G., & Rabinowicz, W. (2015). Value Superiority. In I. Hirose & J. Olson (Eds.), *The Oxford Handbook of Value Theory* (pp. 225–48). New York: Oxford University Press.

Arrhenius, G., Ryberg, J., & Tännsjö, T. (2014). The Repugnant Conclusion. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014, pp. 1–30). Retrieved from http://plato.stanford.edu.ezp.sub.su.se/archives/spr2014/entries/repugnant-conclusion/.

Blackorby, C., Bossert, W., & Donaldson, D. (1995). Intertemporal Population Ethics: Critical-Level Utilitarian Principles. *Econometrica*, *63*(6), 1303–20. https://doi.org/10.2307/2171771.

Blackorby, C., Bossert, W., & Donaldson, D. (1997). Critical-Level Utilitarianism and the Population-Ethics Dilemma. *Economics and Philosophy*, *13*(02), 197–230. https://doi.org/10.1017/S026626710000448X.

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

Broome, J. (1999). *Ethics out of Economics*. Cambridge: Cambridge University Press.

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

Bykvist, K. (2007). The Good, the Bad, and the Ethically Neutral. *Economics and Philosophy*, *23*(01), 97–105. https://doi.org/10.1017/S0266267107001253.

Carlson, E. (1998). Mere Addition and Two Trilemmas of Population Ethics. *Economics and Philosophy*, *14*(02), 283–306.

Carlson, E. (forthcoming). *On Some Impossibility Theorems in Population Ethics*. Mimeo, Department of Philosophy, Uppsala University.

Chang, R. (2002). The Possibility of Parity. *Ethics*, *112*(4), 659–88.

Chang, R. (2005). Parity, Interval Value, and Choice*. *Ethics*, *115*(2), 331–50.

Fehige, C. (1998). A Pareto Principle for Possible People. In C. Fehige & U. Wessels (Eds.), *Preferences* (pp. 508–43). Berlin: W. de Gruyter.

Feldman, F. (1997). *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*. Cambridge: Cambridge University Press.

Handfield, T., & Rabinowicz, W. (2018). Incommensurability and Vagueness in Spectrum Arguments: Options for Saving Transitivity of Betterness. *Philosophical Studies*, *175*(9), 2373–87. https://doi.org/10.1007/s11098-017-0963-9.

Herlitz, A. (2020). Spectrum Arguments, Parity and Persistency. *Theoria*, *86*(4), 463–81. https://doi.org/10.1111/theo.12249.

Hudson, J. L. (1987). The Diminishing Marginal Value of Happy People. *Philosophical Studies*, *51*(1), 123–37. https://doi.org/10.1007/BF00353967.

Kavka, G. S. (1982). The Paradox of Future Individuals. *Philosophy & Public Affairs*, *11* (2), 93–112.

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

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

Narveson, J. (1973). Moral Problems of Population. *The Monist*, *57*(1), 62–86.

Narveson, J. (1978). Future People and Us. In R. I. Sikora & B. Barry (Eds.), *Obligations to Future Generations* (pp. 38–60). Cambridge: White Horse Press.

Ng, Y.-K. (1989). What Should We Do about Future Generations? *Economics and Philosophy*, *5*(02), 235–53.

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

Parfit, D. (1986). Overpopulation and the Quality of Life. In P. Singer (Ed.), *Applied Ethics* (1st edition, pp. 145–64). Oxford: New York: Oxford University Press.

Parfit, D. (2014). *How We Can Avoid the Repugnant Conclusion*. Mimeo, University of Oxford, Faculty of Philosophy.

Parfit, D. (2016). Can We Avoid the Repugnant Conclusion? *Theoria*, *82*(2), 110–27. https://doi.org/10.1111/theo.12097.

Sider, T. R. (1991). Might Theory X Be a Theory of Diminishing Marginal Value? *Analysis*, *51*(4), 265–71.

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

Thomas, T. (2018). Some Possibilities in Population Axiology. *Mind*, *127*(507), 807–32. https://doi.org/10.1093/mind/fzx047.