THEORIA

ORIGINAL ARTICLE OPEN ACCESS

The Continuum Fallacy in Moral Philosophy

John Broome

University of Oxford, Oxford, UK

Correspondence: John Broome (john.broome@philosophy.ox.ac.uk)

Received: 1 November 2024 | **Revised:** 22 May 2025 | **Accepted:** 4 September 2025

Funding: The author received no specific funding for this work.

Keywords: asymptote | continuum argument | derek parfit | imprecision | incommensurability | repugnant conclusion | reverse repugnant conclusion | spectrum argument

ABSTRACT

'Spectrum arguments' or 'continuum arguments' in moral philosophy are sometimes invalid because they commit a particular fallacy I call the 'Continuum Fallacy'. An important example is an argument in population ethics described by Derek Parfit, which purports to derive a conclusion that he and others find repugnant on the basis of a weak and plausible premise. Parfit treats this as a paradox, and takes up the challenge of resolving it, looking for a way to avoid the Repugnant Conclusion. The solution he offers depends on the existence of imprecision within the relation of betterness among populations of people. Other philosophers have taken up the same challenge, following Parfit's lead, and offered solutions also based on imprecision or incommensurability. I show that actually the Repugnant Conclusion is not implied by Parfit's appealing premise. There is therefore no paradox and no real challenge. Moreover, the explanation of why this is so has nothing to do with imprecision, incompleteness, incommensurability, indeterminacy or vagueness in betterness. It is consistent with a sharp, complete betterness ordering.

1 | Introduction

One style of argument appears frequently in the literature of moral philosophy. It is generally called a 'spectrum argument' or a 'continuum argument'. Some arguments of this type are invalid because they commit a particular fallacy, which I shall call the 'Continuum Fallacy'.

This fallacy appears in various guises, but its core is always the same. Take a set of quantities that is totally ordered and constitutes a continuum from smaller quantities to larger ones. The quantities might be anything, but in the literature of moral philosophy, they are often levels of a person's wellbeing or the intensity of a pain. Now take two particular quantities, one higher than the other. Take an infinite diminishing sequence of quantities, starting from the higher of these two. Each quantity in the sequence is less than the previous one. The fallacy is to suppose that, necessarily, this sequence eventually gets down to the lower of the two quantities or below it.

The argument in this paper requires some mathematical precision, and for that purpose, I shall use some notation.

Continuum Fallacy. Let Q be a set of quantities that is totally ordered by $>$ and forms a continuum. From Q, take two quantities q_h and q_l where $q_h > q_l$, and an infinite sequence $q_1, q_2, \ldots$ where $q_1 = q_h$ and for all i, $q_i > q_{i+1}$. Then for some k, $q_l \geq q_k$.

This proposition is supposed to be a necessary truth. Actually, it is a fallacy, which is to say not a necessary truth.

Stated baldly like this, the fallacy is obvious. The sequence could converge to a limit that is either q_l or above it. But in the thick of an argument, some authors do not notice this possibility. Also, some authors may be used to thinking of quantities in discrete units, with the implication that between any two given quantities, there is at most a finite number of other quantities. If quantities are discrete, the proposition is necessarily true and not a fallacy at all. But it is a fallacy if the set of quantities is dense, which means that between any two given quantities, there is always another quantity. A continuum is dense.

For whatever reason, the fallacy crops up in moral philosophy. For instance, it occurs explicitly in Temkin (1996), where Temkin uses it to argue that the relation of betterness is intransitive. Temkin assumes 'there is a continuum of unpleasant or "negative" experiences ranging in intensity from extreme forms of torture to the mild discomfort of a hangnail'. He sets up a sequence of unpleasant experiences. The sequence starts with an extreme form of torture, and each member of the sequence is less intense than the previous one. Then he simply assumes that the sequence eventually reaches the mild discomfort of a hangnail. This is exactly the continuum fallacy. This example of it was identified by Binmore and Voorhoeve (2003).

The fallacy occurs again in Dorsey (2009), where Dorsey uses it to argue for a particular sort of discontinuity in value. Dorsey's continuum of bad things ranges from the loss of a person's life down to a headache. His argument is less explicit than Temkin's, but it has just the same structure. I identified this example of the continuum fallacy in Broome (2010). Later, I described the fallacy differently in Broome (2019).

It continues to crop up. A more recent example appears in Parfit (2016). I shall concentrate on this example because it is still influential in discussions of population ethics. Parfit sets out what he calls the 'Continuum Argument', which purports to demonstrate that a weak, plausible premise implies a conclusion that he and many other people find repugnant. He treats this as a paradox and takes up the challenge of resolving it, looking for a way to avoid this Repugnant Conclusion. The solution he offers depends on the existence of imprecision within the relation of betterness among populations of people. Parfit's argument has led other philosophers to continue knocking their heads against the Repugnant Conclusion. Like Parfit, they find its derivation paradoxical. Like Parfit, they solve the paradox in ways that, like Parfit's, are based on imprecision, incommensurability, indeterminacy, parity or vagueness in betterness.[1]

But actually, the Continuum Argument is invalid. It commits the Continuum Fallacy. There is therefore no paradox, and the argument gives no valid grounds for thinking betterness is anything other than a sharp, complete ordering. (I do not suggest it actually is a sharp, complete ordering.) I think this argument has diverted population ethics into an unproductive direction.

The Repugnant Conclusion can be validly derived from other, stronger premises. For instance, it can be derived from some versions of utilitarianism. Section 4 of this paper discusses one stronger premise that is closely related to Parfit's own, but apart from that, this paper does not examine alternative arguments for the Repugnant Conclusion.

Section 2 sets up the formal framework that underlies my argument. Section 3 presents Parfit's plausible premise and shows that it does not imply the Repugnant Conclusion. Section 4 presents a related, stronger premise that does imply the Repugnant Conclusion. It explains that, although this stronger premise can seem attractive at first, it is not really defensible. Sections 5 and 6 explain how vagueness and incommensurability in value, respectively, have been thought to contribute to refuting the Repugnant Conclusion, and why actually they do not contribute. Section 7 describes the Reverse Repugnant Conclusion, which is commonly taken to be at least as repugnant as the Repugnant Conclusion. A problem with some ways of avoiding the Repugnant Conclusion is that they imply the Reverse Repugnant Conclusion instead. Section 7 shows that my refutation of the Repugnant Conclusion does not suffer from this problem; it can refute the Reverse Repugnant Conclusion simultaneously. Section 8 summarises my conclusions.

2 | The Framework

I shall conduct my argument within a limited domain of possible populations. Each member of the domain is a population (n, w) consisting of n people each having the same lifetime wellbeing w. Within each population in the domain, everyone is equal. The question at issue is which of these populations is better than which.

I shall assume that the wellbeings of populations included in the domain are strictly ordered. This is not true of wellbeings in general because often it is indeterminate which of two people has more wellbeing than the other; neither of them is determinately better off than the other. But my domain is constructed around a particular continuous chain of wellbeings that are strictly ordered, extending from the level of a very bad life to the level of a very good one.

I assume there is such a chain. This is very plausible because a person's wellbeing can always be determinately improved in a continuous manner. For example, the length of some painful episode in the person's life can be continuously diminished. Each diminution makes life determinately better. So it is very plausible that a continuous, increasing chain of wellbeings can be found.

The argument in the previous paragraph assumes time is continuous. I understand that physics does not rule out the possibility that time is discrete rather than continuous on the quantum scale. If that is so, wellbeings may be discrete on a very small scale, too. The Appendix to this paper explains that this sort of discreteness would make no difference to the paper's conclusions.

The members of each population in the domain have a wellbeing that belongs to this chain. Other wellbeings are omitted from the domain, just as unequal populations are omitted. My purpose in adopting such a sparse domain is partly to keep things simple and partly to make it transparent that no imprecision of any sort is involved in my demonstration that the Continuum Argument is invalid.

Wellbeings in the domain are strictly ordered, but I assume no scale of wellbeing beyond the ordinal. I define no zero and no unit of wellbeing.

The domain is illustrated in Figure 1 in a two-dimensional graph with numbers of people n on the horizontal axis and wellbeing w on the vertical axis. I pick as a reference point a population

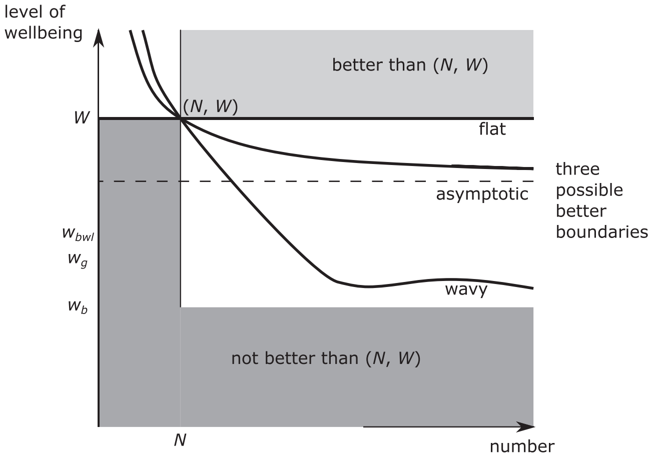

FIGURE 1 | Some possible better boundaries.

(N, W). This is intended to be what Parfit (2016, 110) describes as a population of 'many people who would all have some very high quality of life.' I call it 'the very good population'.

Parfit's argument is supposed to rest on minimal premises; that is why the Repugnant Conclusion is supposed to be hard to escape. I shall therefore conduct my argument within a minimalist framework. But I do need to make three framework assumptions about the ordering of populations. I shall not question them. They are not unquestionable, but I would not expect them to be questioned in this debate. I use '$>$' for greater than and '$\succ$' for better than.

Framework Assumption 1. Increasing wellbeing is good: $(n, w) \succ (n, w')$ if $w > w'$.

Framework Assumption 2. Above some level of wellbeing, increasing numbers makes the population better or equally good: there is a w' such that, for all $w > w'$, and for all n and n' with $n > n'$, $(n, w) \succeq (n', w)$.

Let w_g be the lower bound of all such w'. Call the set of wellbeings above w_g the 'good range'.

Framework Assumption 3. Below some level of wellbeing, increasing numbers makes the population worse or equally bad: there is a w' such that, for all $w < w'$, and for all n and n' with $n > n'$, $(n', w) \succeq (n, w)$.

Let w_b be the upper bound of all such w'. Call the set of wellbeings below w_b the 'bad range'.

A technical note. The bad range and the good range of wellbeings may overlap, so w_g is below w_b. The overlap will consist of wellbeings at which increasing numbers leave the population equally good. The possibility of an overlap plays no part in this paper, and it will soon be ruled out by the assumption below called 'Substitution'.

Some consequences of the three framework assumptions are shown in Figure 1. These assumptions imply that populations within the lightly shaded area are better than the very good population (N, W) and those in the more heavily shaded area are not better than (N, W) (They are actually worse).

Next I pick out in the diagram the set of populations that are better than (N, W) by drawing its boundary, which I shall call the 'better boundary' of (N, W). Because of the three framework assumptions, this better boundary has to include all the lightly shaded area and exclude all the more heavily shaded area. This means it has to have a lower bound. The assumptions also ensure that, so long as the boundary is in the good range, it does not slope upwards. Figure 1 shows three possible better boundaries.

Populations on the better boundary itself may or may not be better than (N, W); I make no assumption about this. I do assume the law of excluded middle, so populations below the boundary are not better than (N, W). However, I do not assume that populations are completely ordered, so I do not assume that all populations below the boundary are either worse than (N, W) or equally as good as (N, W). They might be incommensurable with it.

A technical note. If the better zone and the worse zone overlap, and if (N, W) is within the overlap, the only possible better boundary is the horizontal line through (N, W).

3 | Substitution and the Repugnant Conclusion

I now add the premise that Parfit finds plausible. In his words, it is:

> Compared with the existence of many people who would all have lives that were equally worth living, there are some much larger numbers of people whose existence would be better, though these people would all have lives that would be slightly less worth living.
>
> (Parfit 2016, 116)

More formally:

Substitution. For any (n, w) where $w > w_g$, there is a population (n', w') such that $(n', w') \succ (n, w)$ and $w' < w$.

This is an almost literal translation of Parfit's words into more formal language, translating 'numbers of people' as 'populations'. Just two expressions of Parfit's are not translated: 'much larger' and 'slightly'. In the minimalist framework I have set up, these expressions have no formal meaning. In any case, they are redundant. If we added enough structure to the framework to give them meaning, Substitution together with the framework assumptions would immediately imply:

For any (n, w) where $w > w_g$, there is a population (n', w') such that $(n', w') \succ (n, w)$ and $w' < w$, and n' is much larger than n, and w' is slightly less than w.

So those words contribute nothing to Parfit's premise.

Substitution says that a sufficiently large increase in numbers can make up for a sufficiently small decrease in wellbeing. In the diagram, it means that within the better zone, the better boundary eventually has to decline. It can have flat sections, but any flat section has to be followed by a decline. This assumption knocks out the possibility of the flat boundary in Figure 1. The other two boundaries remain possible.

Let w_{bwl} be the wellbeing of a life that Parfit would consider barely worth living. The conclusion Parfit considers repugnant is:

Repugnant Conclusion. There is a number n such that $(n, w_{bwl}) \succ (N, W)$.

In Parfit's words:

> Compared with the existence of many people who would all have some very high quality of life, there is some much larger number of people whose existence would be better, even though these people would all have lives that were barely worth living.
>
> (110)

Parfit's Continuum Argument is supposed to derive the Repugnant Conclusion from Substitution (117).

Parfit clearly thinks the Continuum Argument is valid. In discussing it, he says 'An argument fails if this argument has some premise whose rejection is less implausible than this argument's conclusion' (120), and on this basis, he rejects the premise Substitution. But an argument also fails if it is invalid, and he clearly excludes this possibility. Because he thinks it plausible to reject Substitution only if the betterness relation is imprecise, he concludes this relation is indeed imprecise (120).

But the Continuum Argument cannot be valid, because Substitution together with the three framework assumptions does not imply the Repugnant Conclusion. The better boundary may reach an asymptote above the level of w_{bwl}, as the asymptotic boundary does in Figure 1. In that case, no population with wellbeing w_{bwl} is better than (N, W).

Where does the Continuum Argument go wrong? It starts by defining a sequence of populations, A, B, C and so on (117). In my notation, A is the very good population, (N, W). B is a population (n_1, w_1) that is better than A but that has a lower level of wellbeing $w_1 < W$. Substitution guarantees that such a population exists. C is a population (n_2, w_2) that is better than B but that has a lower level of wellbeing $w_2 < w_1$. And so on, at each step, Substitution guarantees that the next step exists. So this is a sequence of populations, each better than the one before and each with a wellbeing less than the one before. Then Parfit says:

> This series continues down to world Z, in which there would exist some vast number of people who would have lives that were barely worth living. Since better than is a transitive relation, this argument implies that Z would be better than A. That is the Repugnant Conclusion.
>
> (117)

Z is a population (n, w_{bwl}) for some number n. So Parfit assumes that, necessarily, the sequence of diminishing wellbeings W, w_1, w_2, … eventually gets down to w_{bwl}. This is precisely the Continuum Fallacy. When the better boundary has an asymptote above w_{bwl}, the sequence converges to a limit at or above that asymptote.

The asymptote that blocks the argument belongs to the better boundary, specifically, of the very good population (N, W). The better boundaries of different reference populations may have different asymptotes. Consequently, none of these asymptotes need to mark any particular threshold in the ordering of wellbeings. The better boundaries of some less good reference populations may well pass below w_{bwl}. The invalidity of the Continuum Argument does not depend on the existence of a particular threshold level of wellbeing.

There is nothing quirky about an asymptotic better boundary. In particular, it implies no discontinuity in the betterness ordering. An ordering is defined to be continuous if, whenever an element x is above an element y in the ordering, there is a neighbourhood of x and a neighbourhood of y such that any element in the neighbourhood of x is above any element in the neighbourhood of y. Many people consider a discontinuous betterness ordering implausible. In our context, it would imply that an improvement in a population's wellbeing, however small, can cause a jump in the population's position in the ordering—pushing it suddenly above a lot of other populations. Many people find that implausible.

Discontinuity is an implication of a true lexical ordering. Because of this and because many people consider discontinuity implausible, when John Rawls introduced the idea of a lexical ordering he treated it as only an approximation and said it 'cannot be strictly correct' (Rawls 1972, 45). But an asymptotic better boundary does not imply a true lexical ordering, nor does it imply any discontinuity.

I am using 'discontinuity' in the mathematical sense. The asymptotic boundary does imply a discontinuity in a different sense that was introduced by Griffin (1986, 85–6). In Griffin's words, there is a discontinuity when 'enough of *A* outranks any amount of *B*'. There is nothing implausible about that. Let *A* be a life at the very good level *W*, and *B* a life at the barely worth living level w_{bwl}. We can take the very good population (N, W) to be 'enough of *A*'. There is discontinuity in Griffin's sense if (N, W) outranks any number of people at level w_{bwl}. That is to say, for all n, $(N, W) \succ (n, w_{bwl})$. This is nothing more than the negation of the Repugnant Conclusion, slightly strengthened. It is not implausible to deny the Repugnant Conclusion, and nor is it implausible to make this slightly stronger claim. The slightly stronger claim implies discontinuity in Griffin's sense, so this sort of discontinuity is not implausible.

Also, even when the better boundary has an asymptote, the betterness ordering may be complete and precise, containing no incommensurability or imprecision. The framework assumptions allow for the possibility of incommensurability or imprecision, but this possibility plays no part in explaining why the Continuum Argument is invalid.

Elsewhere in the same paper, Parfit (2016, 112) makes a case against what he calls 'diminishing value views', according to which, as more and more people are added to the world's population, the value of each further person diminishes. It might be thought that an asymptote in the better boundary implies that lives have diminishing value in this way.

Actually, the asymptote does not have this implication. I described the asymptotic boundary without any reference to a numerical scale of value. We could add a numerical scale of value that represents value, but it would still not be implied that people have diminishing value. We could choose a scale in such a way as to keep the value of each extra person constant.[2]

However, it is true that, if we were to travel rightwards along an asymptotic better boundary, we would find it eventually becoming flatter. Its downward slope eventually diminishes. This slope is called the 'marginal rate of substitution'. If we were to represent betterness by a numerical scale of value, the marginal rate of substitution would be the ratio of the value of an extra person to the value of increasing wellbeing. This ratio must eventually diminish along the asymptotic better boundary.

Is the diminishing marginal rate of substitution an objection to the asymptote? I do not see why it should be. Moreover, it is inevitable. The marginal rate of substitution is not constant unless the better boundary is a straight line. The framework assumptions ensure that the better boundary cannot be a straight line unless it is flat. So if there is no diminishing marginal rate of substitution, every better boundary, for every reference point, is flat. This implies that the value of an extra person is always zero, or at least that it is always lexically dominated by an increase in wellbeing.

If all better boundaries are flat, an implication is that a diminution in wellbeing can *never* be compensated for by an increase in numbers. This would not be a ridiculous conclusion: it is an implication of average utilitarianism, for example. But Parfit would not want to draw this conclusion from his rejection of diminishing value. He thinks that adding people to the world is sometimes good. If it is, then the marginal rate of substitution of each better boundary must eventually diminish. This is a consequence of the framework assumptions alone.

4 | Sequential Substitution

Recognising the failure of Parfit's Continuum Argument, other authors[3] have replaced Substitution with this stronger premise:

Sequential Substitution. There is a finite sequence of populations $(n_0, w_0), (n_1, w_1), \dots (n_k, w_k)$, where $(n_0, w_0) = (N, W)$ and $w_k = w_{bwl}$, such that, for any r with $0 \leq r \leq k-1$, $(n_{r+1}, w_{r+1}) \succ (n_r, w_r)$.

This definition says that at each step in the sequence, wellbeing diminishes, but its diminution is more than compensated for by an increase in numbers, so each step is an improvement. The assumption of Sequential Substitution is that there is a finite descending sequence of wellbeings starting at W and finishing at w_{bwl}, such that decreasing wellbeing from one member of the sequence to the next can always be more than compensated for by an increase in numbers.

The Repugnant Conclusion implies Sequential Substitution because it just *is* Sequential Substitution with $k=2$, which means the sequence has just two members. Sequential Substitution implies the Repugnant Conclusion because betterness is transitive. So Sequential Substitution and the Repugnant Conclusion are equivalent.

Parfit seems to think Sequential Substitution is implied by Substitution. It might be more accurate to say that he does not clearly recognise the difference between the two. Other philosophers (e.g., Hájek and Rabinowicz 2022, 901) recognise the difference but nevertheless think Sequential Substitution is independently appealing. They think it can revive the aura of paradox that Parfit casts over the Repugnant Conclusion.

The appeal of Sequential Substitution—in so far as it has one—is a combination of two thoughts. First is the thought that a finite sequence of slight reductions in wellbeing, starting at the level of the very good population, can eventually come down to the level of a life barely worth living. Second is the thought that a slight reduction in wellbeing can always be more than compensated for by a sufficiently big increase in numbers. The first thought gives us a sequence of wellbeings and the second thought allows us to create a sequence of populations satisfying Sequential Substitution by choosing at each stage a number of people that more than compensates for the reduction of wellbeing at that stage.

I have been assuming no more than an ordinal scale of wellbeing, which is not enough to give these thoughts a clear sense. It gives no meaning to 'slight'. But those who are attracted to Sequential Substitution would evidently be happy to tighten up the measurement of wellbeing enough to endow this word with a meaning. For example, they might assume a cardinal scale of wellbeing and take a 'slight reduction in wellbeing' to be a reduction less than some particular distance on this scale. This would make the first thought true. I shall accept for the sake of argument that there are indeed finite sequences of wellbeings, each slightly less than the one before, that lead from W down to w_{bwl}. This is to accept the first thought.

The second thought as I expressed it is equivocal. It might mean that *there is* a slight reduction in wellbeing that can be more than compensated for by an increase in numbers. This is just the assumption of Substitution, which I have been taking for granted. It gives the second thought some appeal, but it does not imply Sequential Substitution. My intended meaning is that *any* slight reduction in wellbeing can be more than compensated for by an increase in numbers. Together with the first thought, this does imply Sequential Substitution.

However, once we have a meaning for 'slight' that makes the first thought true, the second thought with my intended meaning is not necessarily true. In particular, once we have picked a sequence of slight reductions leading from W to w_{bwl}, it may not be true of each step in this sequence.

It will definitely not be true of each step if the better boundary of the very good population (N, W) has an asymptote above w_{bwl}. As a hypothesis for reductio, suppose we have a sequence that satisfies Sequential Substitution. One step of the sequence—say from (n_r, w_r) to (n_{r+1}, w_{r+1})—must cross the asymptote, going from a level slightly above the asymptote to either the asymptote itself or a level slightly below it. By hypothesis, (n_r, w_r) is better than (N, W), so it is on or above the better boundary of (N, W). (n_{r+1}, w_{r+1}) is either on the asymptote or below it. Figure 2 shows that, however large the number n_{r+1} may be, (n_{r+1}, w_{r+1}) cannot be better than (n_r, w_r). This is a step that cannot be more than compensated for by an increase in numbers. The hypothesis is false and the second thought is false.

The upshot is that, once we recognise that the better boundary of (N, W) can have an asymptote above w_{bwl}, the appeal of Sequential Substitution should dissolve.

A parallel argument appears in Section 2 of Thomas (2018). In that section, Thomas examines the first thought, which he calls 'Small Steps', and the second thought, which he calls the 'Quantity Condition' following Arrhenius (Arrhenius Forthcoming). Recognising that the two together imply Sequential Substitution and the Repugnant Conclusion, he recommends dropping the first thought rather than the second. In effect, this means choosing a meaning for 'slight' that validates the second thought and falsifies the first. I made the opposite choice above, and in Section 4 of the same paper, Thomas himself gives support to the first thought and implicitly rejects the second. Either way, the consequence is that he rejects Sequential Substitution.

Still, despite the possibility of an asymptote, some philosophers feel an intuitive attraction towards Sequential Substitution. An example is Thornley (2022). Thornley accepts a version of Thomas's Small Steps, but he is reluctant to give up 'the second thought', whose negation he calls 'Weak Noninferiority Across Slight Differences'.[4] So he finds it hard to avoid a commitment to Sequential Substitution.

Philosophers who are intuitively attracted to Sequential Substitution but who also find the Repugnant Conclusion repugnant are in a bind, since the two are equivalent. They have an easy way to escape: they should recognise that their intuition in favor of Sequential Substitution cannot be trusted.

First, they should remember that even the weaker claim Substitution is by no means universally supported by intuition. I believe that most non-philosophers would deny it because they are instead gripped by what I call 'the intuition of neutrality'. This is the intuition that adding a person to the population of

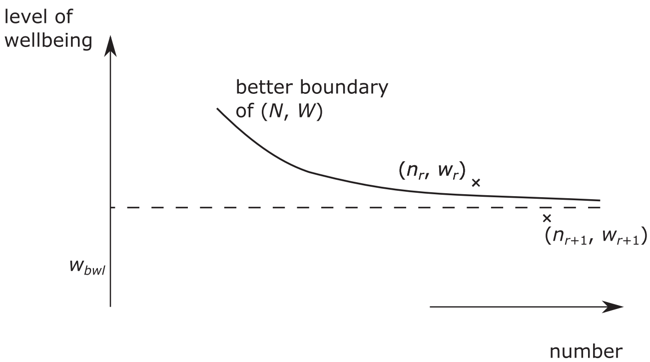

FIGURE 2 | Where a sequence crosses an asymptote.

the world is neutral in value: neither good nor bad. It implies that no reduction in the level of wellbeing could ever be made up for by an increase in the number of people, contrary to what Substitution claims. One example of the popularity of the intuition of neutrality shows itself when economists set a value on human life. They do this often when, for instance, they make a cost–benefit analysis of a project that will improve safety on the roads. They count as a benefit the lives of the people who are saved. But it never occurs to them to count as a benefit the new lives that will result when some of the people who are saved later have babies. Why not? It can only be because they do not think it is indeed a benefit to add these new people to the world's population.

Moreover, many philosophers share this popular intuition. As the philosopher Narveson (1973, 73) put it, 'We are in favour of making people happy but neutral about making happy people'. Many philosophers accept what is called 'the person-affecting restriction', which is the view that no change can be good if it is not good for someone (Roberts 2023). Since adding a person to the population is not in itself good for anyone, adding people to the population cannot make up for a reduction in the level of wellbeing.

These contrary intuitions show us that mere intuition is not enough to support Substitution; it requires argument too. Consequently, intuition is also not enough to support the stronger proposition Sequential Substitution.

Second, philosophers attracted to Sequential Substitution should notice that what I called 'the second thought' differs from Substitution in a mathematically subtle way—basically by a change in the order of quantifiers. Substitution claims that, for any population, there is a degree of reduction in its wellbeing that can be made up for by a sufficiently large increase in numbers. The second thought requires that there is a degree of reduction (called 'slight') in wellbeing such that, for any population, a reduction in its wellbeing of that degree can be made up for by a sufficiently large increase in numbers. Intuition cannot be expected to respond reliably to such a subtle difference, so whatever intuitive support Sequential Substitution may seem to have, it may truly support nothing more than Substitution.

To see why the order of quantifiers matters, think of this version of one of Zeno's paradoxes.[5] When an arrow is approaching its target, it is still a small distance away from the target. Wherever it has got to in its flight, it always has a small distance yet to cover. So it will never reach the target. What is wrong with this reasoning?

If it were true that there is a small distance such that, wherever the arrow has got to, it still has at least that small distance yet to cover, then the reasoning would be correct: the arrow would indeed never reach the target. But the true situation has the quantifiers in the opposite order. Wherever the arrow has got to, there is a small distance such that the arrow has at least that small distance yet to cover. The small distance can depend on where the arrow is. And actually, as the arrow approaches the target, the small distance converges towards zero. Correspondingly, the time remaining in the arrow's flight, which has to cover that small distance, also converges to zero. This means the arrow reaches the target.

In Substitution—for any population, there is a degree of reduction in its wellbeing that can be made up for by a sufficiently large increase in numbers—the existential quantifier 'there is' follows the universal 'for any'. So, the degree of reduction in wellbeing can depend on the population. This makes it possible for the wellbeings in a sequence of populations, where each is better than the one before, to converge to a particular level above w_{bwl}. It makes an asymptote possible. But in the second thought—there is a degree of reduction in wellbeing such that, for any population, a reduction in its wellbeing of that degree can be made up for by a sufficiently large increase in numbers—the degree of reduction in wellbeing does not depend on the population. Convergence is therefore impossible and there can be no asymptote. The consequence of mistaking the order of quantifiers is therefore severe, but we cannot rely on intuition to get the order right.

5 | Vagueness

Thomas (2022) accepts the appeal of Sequential Substitution but aims to 'defuse' it by arguing that betterness is vague. In this, he follows the inspiration of Parfit's argument against the Repugnant Conclusion, which is based on 'imprecision' in betterness.

Thomas explains that Sequential Substitution can be reformulated as a sorites argument. A sorites argument is paradoxical: it has plausible premises that lead to an implausible conclusion. Its structure is as follows. It is applied to some particular predicate. The argument identifies a sequence of objects such that it is plausible that (1) the first member of the sequence satisfies the predicate, (2) the last member of the sequence does not, yet (3) if any member of the sequence satisfies the predicate, so does the following member. These three plausible premises are inconsistent with each other.

For example, take a sequence of collections of grains of sand. It starts with a collection of a million grains. Then it diminishes by one grain at each step and finishes with just one grain. The first collection is plausibly a heap. Plausibly, if one member of the sequence is a heap, so is the next. But plausibly, one grain of sand is not a heap.

Thomas constructs a sorites argument with a predicate he names 'super'. Transposed into my terms, this is a predicate of wellbeings defined as follows. A wellbeing w is super if and only if, for some n, the population (n, w) is better than the very good population (N, W). The sorites sequence is the sequence of wellbeings $w_1, w_2, \ldots w_k$ specified in the statement of Sequential Substitution in Section 4. Sequential Substitution implies that w_1 is super and that, if w_r is super, so is w_{r+1}. The conclusion is that w_k is super. But this is just the Repugnant Conclusion. So plausible premises lead to an implausible conclusion.

Thomas takes the existence of this sorites argument as evidence that 'super' is vague. Since 'super' is defined on the basis of 'better than', he concludes that 'better than' is vague (751). He takes the existence of the sorites argument as evidence because vagueness is the standard account of the sorites paradox. In further support, he cites some philosophers who take the existence of a sorites argument to be *decisive* evidence of vagueness.

But actually, alternative explanations of the sorites paradox are available in some cases. Here is an example. You are presented with a collection of coins, all about the same weight. They are arranged in a sequence such that each coin is indistinguishable in weight from its successor in the sequence, even using the best available scales. Nevertheless, let us assume that each member of the sequence is actually heavier than its successor and that the first member is distinguishably heavier than the last. Take the predicate 'above the average weight'. The first member of the sequence satisfies this predicate and the last member does not. Nevertheless, since each member is indistinguishable in weight from its successor, it is plausible that if it is above the average weight, so is its successor. So we have a sorites argument. Yet the predicate 'above the average weight' is sharp, not vague.

The explanation of this sorites paradox is that at some unidentifiable place in the sequence there is a sharp cut between objects that satisfy the predicate and those that do not. No vagueness is involved. And I have explained this is just the case with Sequential Substitution. There, the sequence is cut by a sharp asymptote. Wellbeings above the asymptote are super; wellbeings on or below the asymptote are not. Vagueness is not the right explanation in this case either.

6 | Incommensurability

Hájek and Rabinowicz (2022), like Thomas, accept the initial appeal of Sequential Substitution. But they set out to refute this premise in order to block the derivation of the Repugnant Conclusion. They do so in a different way from Thomas's or mine. Parfit argues against the Repugnant Conclusion on the basis of imprecision in betterness. Like Thomas, Hájek and Wlodek Rabinowicz follow Parfit's lead but treat imprecision as incommensurability, rather than vagueness. I shall argue that incommensurability does not contribute to refuting either Sequential Substitution or the Repugnant Conclusion. It is the possibility of an asymptote in a better boundary that refutes both.

These authors take a finite sequence of populations as described in Sequential Substitution—that is, a sequence $(n_0, w_0), (n_1, w_1), \ldots (n_k, w_k)$, where $(n_0, w_0) = (N, W)$ and $w_k = w_{bwl}$—and suppose that each member of this sequence appears to be better than its predecessor. Hájek and Rabinowicz argue on the basis of incommensurability that this appearance may be mistaken. When two populations are incommensurable with each other, they argue that one may be almost better than the other—and so appear to be better—without actually being better. One or more members of the sequence may be almost better than their predecessors but not actually better, despite appearances. In that case, the Repugnant Conclusion does not follow.

Even if Hájek and Rabinowicz are right about this, it does not refute Sequential Substitution. Sequential Substitution is an existential claim rather than a claim about a particular sequence. Even when some particular sequence of populations does not satisfy its specifications, another sequence may.

Take a sequence of the sort that Hájek and Rabinowicz describe. At least one of its members is not better than its predecessor but only almost better. We can improve this member, for instance, by increasing its level of wellbeing. Surely, we could improve it enough to make it actually better than its predecessor. And then surely we could improve subsequent members of the sequence enough to ensure that each is better than its predecessor. If necessary, we could add some further members to the end of the sequence to ensure the sequence eventually gets down to w_{bwl}. All this seems possible on the face of it. In this way, the original sequence, which does not satisfy the specifications of Sequential Substitution, can be repaired to make a sequence that does satisfy them. Sequential Substitution would then be true, and the Repugnant Conclusion would follow.

Hájek and Rabinowicz's paper does not block this difficulty. But in a later paper, Rabinowicz (2022) describes the difficulty and goes some way towards blocking it. He introduces the idea of 'persistent incommensurability' (443). Take some sequence $(n_0, w_0), (n_1, w_1), \ldots (n_k, w_k)$ that appears to instantiate Sequential Substitution. Each member of this sequence appears to be better than its predecessor. But suppose that actually a particular member (n_{r+1}, w_{r+1}) is not better than its predecessor (n_r, w_r) but only almost better. It is therefore incommensurable with (n_r, w_r). Rabinowicz defines it as persistently incommensurable with (n_r, w_r) if it remains incommensurable however much its numbers are increased. That is to say: if (n, w_{r+1}) is incommensurable with (n_r, w_r) for any value of n greater than n_{r+1}.

If the incommensurability between (n_{r+1}, w_{r+1}) and (n_r, w_r) is persistent, it blocks one way of repairing the sequence. The given sequence does not instantiate Sequential Substitution because (n_{r+1}, w_{r+1}) is not better than (n_r, w_r). We might hope to make it better by increasing the numbers in (n_{r+1}, w_{r+1}), and also appropriately increasing the numbers in all subsequent members of the sequence. But this will not work if the incommensurability is persistent.

I suggested repairing the sequence differently: by increasing the level of wellbeing in (n_{r+1}, w_{r+1}) rather than the numbers. Persistent incommensurability does not directly block this method of repair. Yet it would block it indirectly if it could be shown that, when there is persistent incommensurability in the given sequence, there will be persistent incommensurability somewhere in any sequence that runs from the very high level of wellbeing down to the level of a life barely worth living.

This is in fact true, as will appear in the next few paragraphs. But it will also appear that this route to blocking Sequential Substitution is irrelevant for refuting the Repugnant Conclusion. The feature of an ordering that explains why incommensurability, when it exists, is persistent itself directly refutes Sequential Substitution and the Repugnant Conclusion, even if there is no incommensurability.

What could explain persistent incommensurability in the first place? It is possible only where the better boundary converges to an asymptote. Figure 2 shows why. Because (n_{r+1}, w_{r+1}) lies on or below the asymptote of the better boundary of (N, W), increasing the number of people n with wellbeing w_{r+1} cannot bring us to a population (n, w_{r+1}) that is better than (n_r, w_r). So if (n_{r+1}, w_{r+1}) is incommensurable with (n_r, w_r), then (n, w_{r+1}) will always be incommensurable with (n_r, w_r). But if there were no asymptote, increasing n would eventually make (n, w_{r+1}) cross the better boundary and eventually become better than (n_r, w_r). Only the asymptote explains how persistent incommensurability is possible.

We already know that the existence of an asymptote above the level of a life barely worth living explains why the Repugnant Conclusion is false. Remember this is an asymptote in the boundary of populations that are better than the very good population (N, W). It is the shape and position of this boundary that determines whether or not the Repugnant Conclusion is true.

Outside the boundary are populations that are worse than those inside, and there may also be some that are incommensurable with some inside. If there is incommensurability, it will be persistent around an asymptote. But none of this contributes to explaining why the Repugnant Conclusion is false. That depends only on the asymptote of the better boundary; incommensurability has nothing to do with it.

One of Hájek and Rabinowicz's (2022, 906) aims is to ease the discomfort you might feel in giving up Sequential Substitution and recognising that sometimes it may not be possible to compensate for a particular slight reduction in wellbeing by an increase in numbers. Likewise, Thomas (2018, 829) claims that vagueness in the betterness relation can 'soften the blow' of giving up Sequential Substitution. But you should feel no blow and no discomfort. As I explained in Section 4, once you recognise that the better boundary may have an asymptote above the level of a life barely worth living, the appeal of Sequential Substitution should dissolve.

7 | The Reverse Repugnant Conclusion

Some philosophers think that escaping from the Repugnant Conclusion is jumping out of the frying pan into the fire. The fire in this case can be called the 'Reverse Repugnant Conclusion' (e.g., Mulgan 2002). To define it, let (N, V) be a very bad population, where N is a large number of people and V is the wellbeing of a life of great suffering. And let w_{awl} be the wellbeing of a life that Parfit would consider almost, but not quite, worth living. The Reverse Repugnant Conclusion is:

Reverse Repugnant Conclusion. There is a number n such that $(n, w_{awl}) \succ (N, V)$.

There is a population of people, all of whose lives are almost worse living, that is worse than the population of great suffering. This appears to be at least as repugnant as the Repugnant Conclusion. The concern is that, though it may be possible to avoid the Repugnant Conclusion, it is not possible to avoid the Repugnant Conclusion and the Reverse Repugnant Conclusion simultaneously.

Let us adopt:

Reverse Substitution. For any (n, w) where $w \succ w_b$, there is a population (n', w') such that $(n', w') \succ (n, w)$ and $w' > w$.

This means that, starting from a population in the bad zone, there is always a small improvement in wellbeing that can be cancelled out by a sufficiently large increase in numbers. It is as acceptable as Substitution itself.

Just as there is no valid argument from Substitution to the Repugnant Conclusion, there is no valid argument from Reverse Substitution to the Reverse Repugnant Conclusion. As we did for the better boundary for the very good population, we can plot a worse boundary for the very bad population (N, V). All populations that are worse than (N, V) lie below this boundary. Consistently with Reverse Substitution, we may draw a worse boundary that converges to an asymptote below the level of w_{awl}. The Reverse Repugnant Conclusion therefore does not follow. Figure 3 illustrates.

The Reverse Repugnant Conclusion could be derived from Reverse Sequential Substitution, defined as the dual of Sequential Substitution. But Section 4 explained that any attraction Sequential Substitution might have should be dissolved by recognising that a better boundary can have an asymptote above w_{bwl}. Similarly, any attraction Reverse Sequential Substitution might have should be dissolved by recognising that a worse boundary can have an asymptote below w_{awl}. It offers us no real grounds for the Reverse Repugnant Conclusion.

The worse boundary of the very bad population does not have to converge to the same asymptote as the better boundary of the very good population. There is room for both w_{bwl} and w_{awl} to lie between the two asymptotes. The Repugnant Conclusion and the Reverse Repugnant Conclusion may therefore both be false together. This can happen even if $w_g = w_b$, which means there is no space between the good range of populations and the bad range. I have chosen to show this case in Figure 3.

The borderline wellbeing $w_g = w_b$ may be in the good range, or the bad range, or neither. In the latter case, it may be a neutral level of wellbeing, at which increasing numbers make the population neither better nor worse. In Figure 3 I have shown the two levels w_{bwl} and w_{awl}—barely worth living and almost worth living—on opposite sides of the borderline wellbeing $w_g = w_b$. But

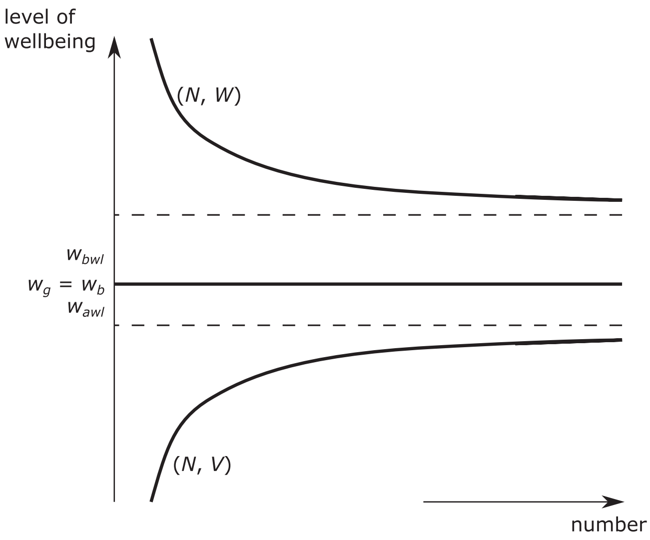

FIGURE 3 | A better boundary and a worse boundary.

I do not affirm that they necessarily span the borderline in this way; I leave that open. If a person has a life worth living, that is a matter of her personal good. If it means anything, it presumably means it is better for the person that she lives this life rather than not living at all. But if a wellbeing is in the good range, that is a matter of general good. It means it is good to increase the number of people living such a life. I leave it open whether the borderline of a life worth living coincides with the borderline of the good range of wellbeings.

A final reminder: everything I have said is consistent with the betterness relation's being complete and precise. Incompleteness, incommensurability or imprecision do not contribute to explaining how the Repugnant Conclusion and the Reverse Repugnant Conclusion may be simultaneously avoided.

8 | Conclusion

I have argued that the Continuum Argument is invalid because it depends on the Continuum Fallacy. It is impossible to derive the Repugnant Conclusion from the simple, appealing premise of Substitution, even given some uncontroversial framework assumptions. The argument is invalid even if the betterness relation is complete and precise.

Other arguments for the Repugnant Conclusion are based on stronger premises. For example, an argument might be based on utilitarianism. If we assume utilitarianism, it may even be that the Repugnant Conclusion and the Reverse Repugnant Conclusion cannot be avoided simultaneously except on the basis of the incompleteness, vagueness or imprecision of the betterness ordering. (See Broome 2004, 213–14).

Appendix on discrete wellbeing.

I argued in Section 2 that wellbeing can vary continuously, on the grounds that the length of a good or bad episode in a life can vary continuously. But what if that is incorrect and wellbeing is not continuous? Gustaf Arrhenius and Erik Carlson have pointed out to me that physics has not ruled out the possibility that time is discrete rather than continuous. If that is so, wellbeing might also come in discrete units. What difference would that make to the conclusions of this paper?

It seems obvious that it could make no difference to ethical conclusions. If wellbeing were discrete, it would be discrete on the minutest, quantum scale. That could not possibly make any difference to ethics. Such an esoteric piece of science, far beyond anything apparent in our ordinary living, and indeed beyond any currently possible experiments, could make no difference to how we ought to live.

Yet, if wellbeing were discrete, Parfit's Continuum Argument would be valid (though incorrectly named). The Repugnant Conclusion would be validly derivable from Substitution, so Parfit's paradox would be reinstated. This makes it seem that the discreteness of wellbeing would make a serious difference to our ethical conclusions, contrary to what I have just said. How can this apparent contradiction be resolved?

I explained in Section 3 that, if wellbeing is continuous, a better boundary can have a horizontal asymptote that blocks the Continuum Argument. An asymptotic boundary is shown in Figure 1. If wellbeing is discrete, a better boundary cannot strictly have an asymptote, but it can approximate one. As we move rightwards along the asymptote shown in Figure 1, the better boundary declines, converging towards the asymptotic level of wellbeing but never reaching it. If wellbeing is discrete, the better boundary can follow just the same course except that it cannot decline forever. Eventually, it must reach some particular level of wellbeing and go no lower. It has to become flat. Since the discreteness is very small, this flatness will be reached very far to the right—far beyond any number that the world's population could conceivably attain. Like the genuine asymptote of the continuous case, this approximate asymptote will block descent in the Continuum Argument to the level of a life barely worth living. Our substantive ethical conclusions will be exactly the same.

I explained in Section 3 that Substitution implies that a better boundary has eventually to decline. So if the better boundary ever becomes flat, Substitution is false. With discrete wellbeing, the Repugnant Conclusion could be validly derived from Substitution, but its premise, Substitution, may be false. Its falsity will show itself only far out to the right of the diagram, far beyond any numbers of practical concern. Substitution can remain true for all practical purposes, and the small-scale discreteness of wellbeing would leave the ethical arguments unchanged.

Acknowledgements

I have greatly benefited from discussions with Erik Carlson, Alan Hájek, Jake Nebel and Wlodek Rabinowicz.

Endnotes

[1] Examples are: Teruji Thomas (2022), Hájek and Rabinowicz (2022), Chang (2022), Rabinowicz (2022) and Thornley (2022).

[2] How to do this is described in Broome (2019).

[3] For instance, Carlson (2005) and Pummer (2018) examine Sequential Substitution as discussed by Nebel (2018). Nebel eventually rejects it, but not for the reason I shall give. Thomas (2018) objects to Sequential Substitution on grounds similar to mine.

[4] See particularly Thornley (2022, 406).

[5] Binmore and Voorhoeve (2003) first made the connection between Sequential Substitution and Zeno's paradoxes.

References

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

Binmore, K., and A. Voorhoeve. 2003. "Defending Transitivity Against Zeno's Paradox." *Philosophy and Public Affairs* 31: 272–279.

Broome, J. 2004. *Weighing Lives*. Blackwell.

Broome, J. 2010. "No Argument Against the Continuity of Value: Reply to Dorsey." *Utilitas* 22: 494–496.

Broome, J. 2019. "Lessons From Economics." In *The Oxford Handbook of Ethics and Economics*, edited by M. D. White, 585–608. Oxford University Press.

Carlson, E. 2005. "Intransitivity Without Zeno's Paradox." In *Recent Work on Intrinsic Value*, edited by M. J. Zimmerman and T. Rønnow-Rasmussen, 273–277. Springer.

Chang, R. 2022. "How to Avoid the Repugnant Conclusion." In *Ethics and Existence: The Legacy of Derek Parfit*, edited by J. McMahan, T. Campbell, J. Goodrich, and K. Ramakrishnan, 389–429. Oxford University Press.

Dorsey, D. 2009. "Headaches, Lives and Value." *Utilitas* 21: 36–58.

Griffin, J. 1986. *Well-Being: Its Meaning, Measurement and Moral Importance*. Oxford University Press.

Hájek, A., and W. Rabinowicz. 2022. "Degrees of Commensurability and the Repugnant Conclusion." *Noûs* 56: 897–919.

Mulgan, T. 2002. "The Reverse Repugnant Conclusion." *Utilitas* 14: 360–364.

Narveson, J. 1973. "Moral Problems of Population." *Monist* 57: 62–86.

Nebel, J. 2018. "The Good, the Bad and the Transitivity of Better Than." *Noûs* 52: 874–899.

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

Pummer, T. 2018. "Spectrum Arguments and Hypersensitivity." *Philosophical Studies* 175: 1729–1744.

Rabinowicz, W. 2022. "Can Parfit's Appeal to Incommensurabilities in Value Block the Continuum Argument for the Repugnant Conclusion?" In *Ethics and Existence: The Legacy of Derek Parfit*, edited by J. McMahan, T. Campbell, J. Goodrich, and K. Ramakrishnan, 430–460. Oxford University Press.

Rawls, J. 1972. *A Theory of Justice*. Harvard University Press.

Roberts, M. 2023. *The Existence Puzzles: An Introduction to Population Ethics*. Oxford University Press.

Temkin, L. 1996. "A Continuum Argument for Intransitivity." *Philosophy & Public Affairs* 25: 175–210.

Thomas, T. 2018. "Some Possibilities in Population Axiology." *Mind* 127: 807–832.

Thomas, T. 2022. "Are Spectrum Arguments Defused by Vagueness?" *Australasian Journal of Philosophy* 100: 743–757.

Thornley, E. 2022. "A Dilemma for Lexical and Archimedean Views in Population Axiology." *Economics and Philosophy* 38: 395–415.