

# Population Ethics

## Table of Contents

1. Introduction
2. The Total View
  - Objecting to the Total View: The Repugnant Conclusion
3. The Average View
4. Variable Value Theories
5. Critical Level and Critical Range Theories
6. Person-Affecting Views and the Procreative Asymmetry
7. Practical Implications of Population Ethical Theories
8. Conclusion
9. Resources and Further Reading

## Introduction

Utilitarians agree that if the number of people that were ever to exist is held constant, we should promote the *sum total of well-being* in that fixed population.<sup>1</sup> But in reality, the population is not fixed. We have the option of bringing more people into existence, such as by having children. If these additional people would have good lives, is that a way of making the world better? This question falls in the domain of *population ethics*, which deals with the moral problems that arise when our actions affect who and how many people are born and at what quality of life.

Population ethics is not just an academic exercise. It is relevant to many important practical questions, such as how many children we ought to have, if any; how much we should invest in climate change mitigation; and how much we should worry about near-term risks of human extinction.

This article will survey five major approaches to population ethics:

1. The *total view* that evaluates populations according to the total amount of well-being that they contain.
2. The *average view* that instead focuses on the *average* well-being level in the population.

3. *Variable value theories* that take both factors into account, approximating the total view for smaller populations and the average view for larger populations.
4. *Critical level* (and *critical range*) *theories* that tweak the total view to only count positive well-being above a critical baseline level (or range).
5. *Person-affecting views* that deny we have (non-instrumental) reason to add happy lives to the world.<sup>2</sup>

## The Total View

According to the *total view* of population ethics:

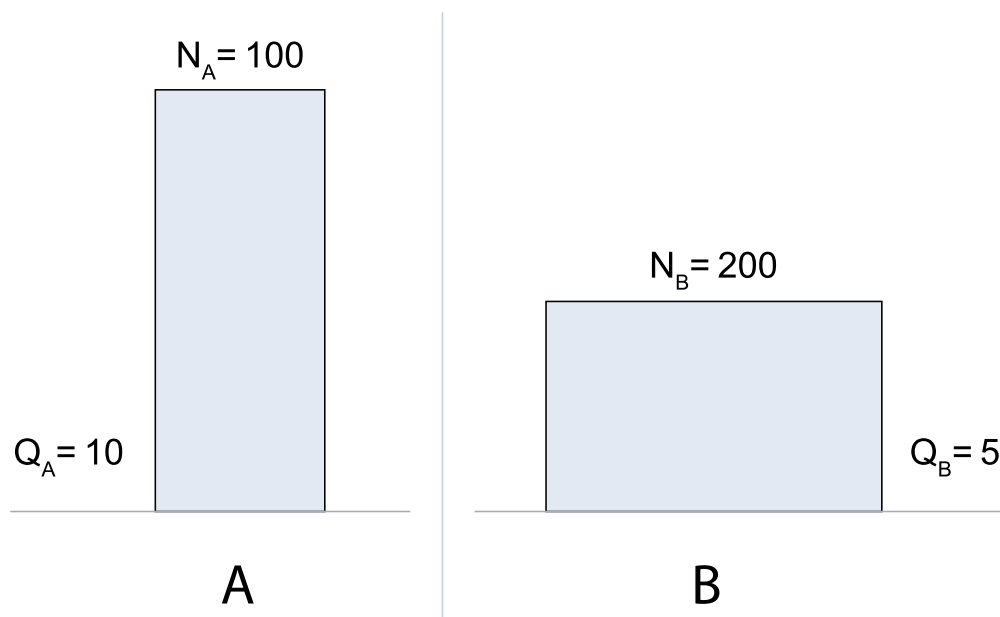
**One outcome is better than another if and only if it contains greater total well-being.**<sup>3</sup>

Importantly, one population may have greater total well-being than another in virtue of having more people. One way to calculate this total is to multiply the number of individuals ( $N$ ) by their average quality of life ( $Q$ ).<sup>4</sup> We can, therefore, represent the total view by the following value function:

$$\text{Value}_{\text{total view}} = N * Q$$

Consider a hypothetical world A with 100 inhabitants ( $N_A$ ) at an average well-being level 10 ( $Q_A$ ) and another hypothetical world B with 200 inhabitants ( $N_B$ ) at well-being level 5 ( $Q_B$ ). On the total view, worlds A and B are equally good because they both have 1,000 units of well-being ( $N_A * Q_A = N_B * Q_B = 1,000$ ).

When comparing hypothetical worlds in population ethics, these worlds are often illustrated graphically. The width of the following graphs represents the number of people, and the height represents their average well-being level. Consequently, the graphs' area—width times height—represents the total welfare in the hypothetical worlds. Illustrated graphically, worlds A and B are equally valuable, according to the total view, since their graphs have the same area.



The total view implies that we can improve the world in two ways: either we can improve the quality of life of existing people, or we can increase the number of people living positive lives.<sup>5</sup> The total view makes no distinction between whether the additional well-being would come to people who already exist or to entirely new people. For example, the total view regards having a child who lives a happy and fulfilled life as something that makes the world better, all else equal, since it adds to the total well-being.<sup>6</sup> Importantly, the claim is not that having the child might make the world better in virtue of enriching the lives of others; instead, it's that having the child is good in itself. How good it is to bring a new person into existence depends on how much better or worse that person's life is than a "neutral life". According to this view, happy people are good and having more of a good thing, other things being equal, makes an outcome better.

In practice, there might be trade-offs between making existing people happier and creating additional happy people. For example, you might think that, on a planet with limited resources, adding more people to an already large population may at some point diminish the quality of life of everyone else severely enough that total well-being decreases. It's an open empirical question whether our world has reached the population size at which adding a person increases or decreases overall well-being.<sup>7</sup>

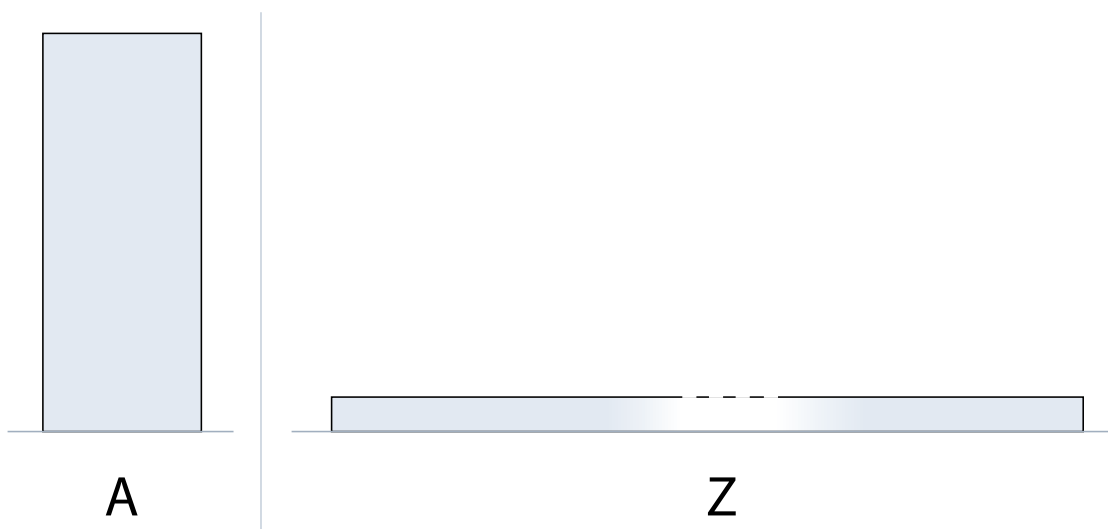
The total view's foremost practical implication is giving great importance to ensuring the long-term flourishing of civilization. This entails taking existential risk reduction very seriously as a moral priority.<sup>8</sup> Existential risks—such as all-out nuclear war, or extreme climate change, or an engineered global pandemic—threaten the survival of humankind. If humanity survives, civilization may flourish over billions of years, and an enormous number of future people may get to enjoy highly fulfilling lives. The total well-being across all future generations may be

astronomically large, and an existential catastrophe would irreversibly deprive humanity of this potentially grand future. From the perspective of the total view and many other moral views, the stakes involved with existential risks are so immense that the mitigation of these risks becomes one of the principal moral issues facing humanity.

## Objecting to the Total View: The Repugnant Conclusion

The most prominent objection to the total view is the *repugnant conclusion*, originally raised by Derek Parfit.<sup>9</sup> In its simplest form, the repugnant conclusion is that:<sup>10</sup>

**For any world A, there is a better world Z in which no one has a life that is more than barely worth living.**



All lives in world Z are positive, but they are only barely worth living. If the population in Z is sufficiently large, Z's total well-being—represented by the graph's area—is greater than A's. Consequently, the total view implies that world Z is better than world A: the repugnant conclusion.

On the total view, a sufficiently large increase in the quantity of a population can compensate for any loss in the average quality of lives in this population, as long as average well-being remains positive. Most people find some trade-offs between quantity and quality intuitive; for instance, almost everyone would agree that our world of about 8 billion people is better than a world with only one person with a slightly higher average well-being level. However, many people find the repugnant conclusion counterintuitive and think that the total view takes quantity-quality trade-offs too far. Given that no one in world Z has a life more than barely worth living, it's tempting to think that Z must be worse than A, regardless of Z's population size.

Importantly, the total view need not imply that we should maximize population size *in practice*. It's an open empirical question how best to promote total well-being in real-life circumstances: blindly increasing population without also ensuring a high quality of lives is not guaranteed (or even likely) to be the best approach. But even if the total view avoids repugnant implications in practice, it remains important to assess whether its broader implications (for various hypothetical scenarios) are ones that we are willing to accept.

As such, proponents of the total view might respond to the challenge presented by the repugnant conclusion by debunking the intuition, biting the bullet, or attacking the alternatives. We will consider these in turn.

### **Debunking the Intuition**

Our intuitions about these cases may be unreliable or based on subtle misunderstandings.<sup>11</sup>

First, the repugnant conclusion involves cases with extremely large numbers of individuals with low but positive welfare. Many philosophers argue that this is a situation where we should expect our intuitions to be unreliable: human brains struggle to intuitively grasp both very large numbers and how adding up lots of small values results in a very large value.<sup>12</sup> So we may fail to understand on an intuitive level how the vast number of lives in world Z could ever compound to something more valuable than world A.

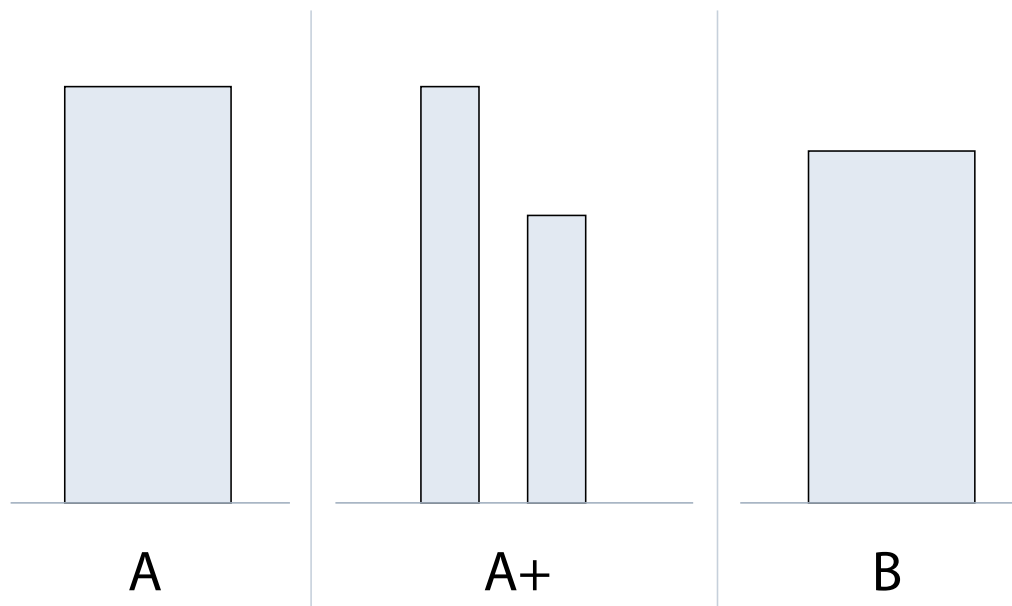
Second, we may not adequately appreciate that lives “barely worth living” are good rather than bad.<sup>13</sup> A life “barely worth living” *is* worth living, and a person with such a life has reason to be glad they exist. Misleading representations of lives “barely worth living” in the academic literature may contribute to this misunderstanding. While it's controversial how to determine which lives are worth living, it has been argued that influential examples in the literature—such as lives containing no goods besides “muzak and potatoes”<sup>14</sup>—are actually not worth living.<sup>15</sup>

Third, we may mistakenly imagine ourselves as part of the populations being compared in the repugnant conclusion. Consequently, an egoistic bias may push us to favor populations with a high quality of life.<sup>16</sup>

### **Tolerating the Intuition**

Proponents of the total view may “bite the bullet” and simply accept that world Z is better than world A. They may point out that, while initially counterintuitive, this conclusion is based on the compelling goal of creating a world with as much total well-being as possible. On the total view, reductions in the average well-being level of a population can be more than compensated for by adding sufficiently many lives that are worth living. Since lives worth living are (one would naturally think) an inherently good thing, it seems rhetorically overblown to call this implication “repugnant”.

Indeed, it turns out to be remarkably difficult to avoid the repugnant conclusion. Strong arguments, such as Parfit's "mere addition paradox",<sup>17</sup> entail the repugnant conclusion without assuming the total view to begin with. Consider the choice between the following three worlds, A, A+, and B.



In world A, everyone's well-being is very high. World A+ contains (i) one population group that is identical to the A population in terms of population size and average well-being, and (ii) a second group of the same size but with slightly lower well-being. Most people would agree that A+ is not worse than A (and may even be better) since, intuitively, merely adding people with lives worth living (without thereby harming anyone else) cannot make an outcome worse. Now consider world B, which has the same overall population size as A+. The average well-being in B is slightly higher than in A+ but lower than in A. When we compare worlds A+ and B, it seems that B must be better. Not only are the average *and* total welfare in B higher than in A+, but well-being is also more equally distributed. Yet, if B is better than A+, and A+ is not worse than A, it follows that B must be better than A.<sup>18</sup> Repeating this process—comparing world B with B+ and C etc.—leads to world Z with an enormous number of people with lives barely worth living. Following the above reasoning, world Z must be better than world A: the repugnant conclusion.

### Attacking the Alternatives

A final response is to note that counterintuitive implications are by no means unique to the total view. Several impossibility theorems prove that it's *logically impossible* for any population ethical theory to satisfy every intuitively desirable principle and axiom.<sup>19</sup> One such axiom is to avoid the repugnant conclusion. However, some philosophers argue that avoiding the repugnant conclusion is not worth the theoretical costs of giving up other axioms or fundamental principles. Accepting

the repugnant conclusion provides an easy response to the impossibility theorems since the total view is consistent with all the other axioms of these theorems.<sup>20</sup> In light of this, a growing number of ethicists have come to accept and defend the repugnant conclusion and the total view.<sup>21</sup>

To evaluate this comparative claim, we must consider the merits (and demerits) of the competing views. Philosophers have proposed several alternatives to the total view that seek to avoid (at least the original version of) the repugnant conclusion. These theories include the *average view*, *variable value theories*, *critical level and range theories*, and *person-affecting views*.

## The Average View

According to the *average view* of population ethics:

**One outcome is better than another if and only if it contains greater average well-being.**

Since the average view aims only to improve the *average well-being level*, it disregards—in contrast to the total view—the number of individuals that exist.<sup>22</sup> Consequently, the average view is represented by a simple value function, with average well-being level  $Q$ :

**Value<sub>average view</sub> =  $Q$**

The average view avoids the repugnant conclusion because it entails that reductions in the average well-being level can never be compensated for merely by adding more people to the population.

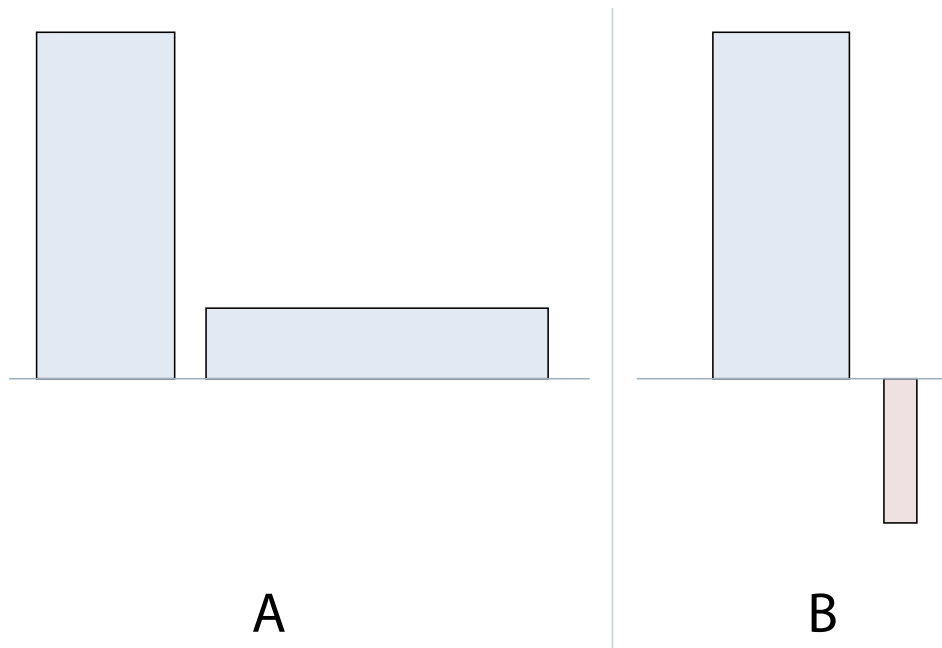
However, the average view has very little support among moral philosophers since it suffers from severe problems.

First, consider a world inhabited by a single person enduring excruciating suffering. The average view entails that we could improve this world by creating a million new people whose lives were also filled with excruciating suffering, if the suffering of the new people was ever-so-slightly less bad than the suffering of the original person.<sup>23</sup>

Second, the average view entails the *sadistic conclusion*:<sup>24</sup>

**It can sometimes be better to create lives with negative well-being than to create lives with positive well-being from the same starting point, all else equal.**

Adding a small number of tortured, miserable people to a population diminishes the average well-being less than adding a sufficiently large number of people whose lives are pretty good, yet below the existing average. To see this, consider the following graph where world A has lower average well-being than world B. Counterintuitively, the average view thus entails that we should prefer world B over world A.



Third, the average view prefers arbitrarily small populations over very large populations, as long as the average well-being is higher. For example, a world with a single, extremely happy individual would be favored to a world with ten billion people, all of whom are extremely happy but just ever-so-slightly less happy than that single person.

## Variable Value Theories

To compromise between the total and average views, some philosophers have proposed *variable value theories*. According to these theories, the marginal value of creating additional people diminishes—the larger the existing population, the smaller the value of adding another person.

The simplest variable value theories may be represented by the following value function, with number of individuals  $N$  and average quality of life  $Q$ :<sup>25</sup>

$$\text{Value}_{\text{variable value}} = Q * f(N)$$

This diminishing marginal value allows variable value theories to avoid some of the weaknesses of the total and average views. For instance, Hilary Greaves writes that:<sup>26</sup>

Arguably, [the average view] is intuitively less plausible for small populations: if there are otherwise only ten persons who will ever live, for instance, it (perhaps) seems more worthwhile to add an additional person with a given positive well-being level than if there are already 100 billion persons.

At first glance, it's an attractive feature of variable value theories that they approximate the total view at small population sizes and the average view at large population sizes since they may

thereby avoid the repugnant conclusion.<sup>27</sup>

However, variable value theories face problems of their own. First, in approximating the average view at large population sizes, they risk susceptibility to the same objections. So, to avoid approving of adding (above-average) negative lives to the world, variable value theorists must invoke an asymmetry according to which only the value of positive lives diminishes but not the disvalue of negative lives. Adding negative lives to a world always makes the world non-instrumentally worse, on such a view, even if it happens to improve the average. However, such an asymmetry leads to an analogue of what Parfit calls the *absurd conclusion*:<sup>28</sup> that a population considered to be good, with many happy and few miserable lives, can be turned into a population considered to be bad merely by proportionally increasing the number of both positive and negative lives.<sup>29</sup> To escape this objection, variable value theorists must allow additional good lives to sometimes *compensate* for additional bad lives, without introducing further unintended consequences that undermine their view. This is no easy task.<sup>30</sup>

## Critical Level and Critical Range Theories

According to critical level theories:

**Adding an individual makes an outcome better to the extent that their well-being exceeds some critical level.**

Consequently, an outcome can be made worse not only by bringing into existence an individual with negative well-being, but also if their well-being is positive yet below the critical level.<sup>31</sup>

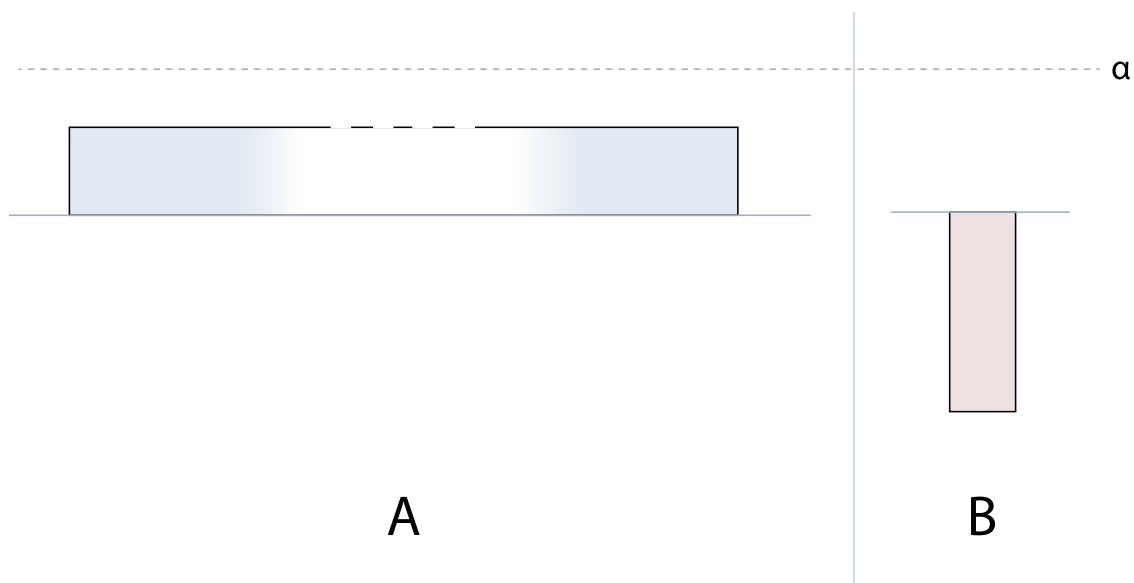
The total value of an outcome, according to critical level theories, can be represented by the following value function, with number of individuals  $N$ , average quality of life  $Q$ , and critical level  $\alpha$ :

$$\text{Value}_{\text{critical level}} = N * (Q - \alpha)$$

By looking at this value function, we may observe that the total view of population ethics is simply a critical level theory with a critical level of zero ( $\alpha = 0$ ). Critical level theories, including the total view, agree that the value of adding an individual to the world depends only on that individual's well-being level and—in contrast to the average view and variable value theories—not on the number of existing persons or their well-being levels.

Critical level theories avoid the repugnant conclusion if their critical level is greater than the well-being level that makes a life “barely worth living”. This follows from the definition of critical level theories since only adding people whose welfare exceeds the critical level can compensate for reductions in the average quality of life.

However, a positive critical level entails an especially appalling version of the sadistic conclusion, which gets worse the higher the chosen critical level.<sup>32</sup> This is because it implies that it's preferable for a world to contain individuals with negative well-being—lives not worth living, such as lives of constant torture—than to contain vastly more individuals with lives that are worth living yet below the critical level. Graphically illustrated, critical level theories counterintuitively imply that world A (with a sufficiently large population) is *worse* than world B.



To avoid this unacceptable result, we may move to a *critical range* theory,<sup>33</sup> according to which:

**Adding an individual makes an outcome better to the extent that their well-being exceeds the upper end of a *critical range*, and makes an outcome worse to the extent that their well-being falls below the lower limit of the critical range.**

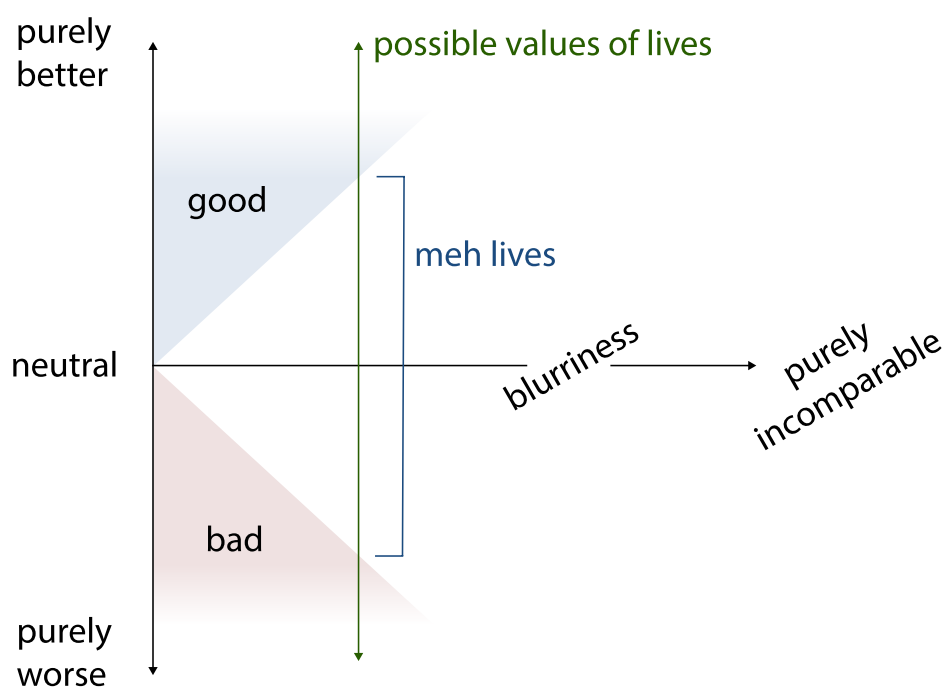
By setting the lower limit of the critical range to include all neutral lives that are neither good nor bad for the person living them, critical range theories can avoid the sadistic conclusion. And by setting the upper limit of the range at the point where lives become clearly worthwhile, these theories avoid the repugnant conclusion.

What about lives that fall within the critical range? Life within this range may strike us as *meh*:<sup>34</sup> neither good nor bad, but also not precisely equal to *zero* in value, either. After all, some meh lives (those toward the upper end of the range) are better than others (those toward the lower end), so it cannot be that adding any life in this range results in an equally valuable outcome. Instead, the outcome's value must be *incomparable* or *on a par* with that of the prior state: neither better, nor worse, nor precisely equal in value.<sup>35</sup> Note that it may be better to add an upper-range meh life to

the world than to add a lower-range meh life, even though adding *either* life is merely “meh”, or results in an outcome that is incomparable with the world in which neither life is added.<sup>36</sup>

To further develop this view (along lines suggested by Johan Gustafsson), we may think of the value of a life as having two dimensions.<sup>37</sup> In addition to the familiar negative-vs-positive dimension, there is a second dimension of what we might call *value blur*. When there is zero blur, the resulting values are perfectly precise and comparable: any positive life, however barely so, then constitutes an intrinsic improvement to the world. But as we increase blur, the resulting value becomes increasingly “meh”, or *incomparable*. If life’s value had infinite blur, then *all* lives would be meh. (We will consider such a view in the next section.) Alternatively, if we think that life’s value admits of just moderate blur, then sufficient positive (or negative) value may overcome this blurriness to qualify the life in question as one that would be in itself good (or bad) to add to the world.

The key implication of this critical range theory (with moderate value blur) is that an intrinsically good life must contain *significantly* more welfare than an intrinsically bad life, because between these two levels there is a moderate range of lives that are meh, as illustrated below:<sup>38</sup>



The resulting view, while theoretically complex, seems less susceptible to severe objections than the other views we’ve surveyed. In particular, it can simultaneously avoid both the repugnant conclusion and the sadistic conclusion. But it cannot accommodate the strong “anti-repugnance” intuition that the idyllic world A is strictly *better* than the repugnant world Z.<sup>39</sup> Critical range theories instead regard the two worlds as incomparable, due to the immense value blur introduced by all those meh lives in world Z.

## Person-Affecting Views and the Procreative Asymmetry

All of the theories discussed above—the total view, the average view, variable value theories, and critical level and range theories—are *impersonal theories*.<sup>40</sup>

Impersonal theories	Value function (value of a particular state of the world)
Total view	$N * Q$
Average view	$Q$
Variable value theory	$f(N) * Q$
Critical level theory <sup>41</sup>	$N * (Q - \alpha)$

**N** = number of persons; **Q** = average well-being level; **α** = critical level; **f()** = strictly increasing and strictly concave function with a horizontal asymptote

Impersonal theories imply that creating an additional person with a (sufficiently) good life makes the world better, other things being equal.<sup>42</sup> However, some people reject this implication. They say that there are no moral reasons to bring additional people into existence—at least no reasons based on the well-being these people would enjoy if created. Jan Narveson put this idea in slogan form: “We are in favor of making people happy, but neutral about making happy people”.<sup>43</sup> *Person-affecting views* of population ethics attempt to capture this intuition of neutrality, and are especially common among non-consequentialists.<sup>44</sup>

To this end, standard person-affecting views accept the *person-affecting restriction*:

**An outcome cannot be better (or worse) than another unless it is better (or worse) for someone.**  
45

At first glance, this principle sounds eminently plausible. When considering only fixed-population cases, it amounts to an expression of *welfarism*: the view that well-being is the only value. But problems arise when comparing outcomes in which different people exist, especially if combined with the principle of *existence incomparability*.<sup>46</sup>

**If a person exists in one outcome but not the other, it is not possible to compare their well-being across these outcomes.**

According to this principle, since the value of existence and non-existence are assumed to be incomparable, existing cannot be better (or worse) for you. When combined with the person-affecting restriction, we get the result that your existence likewise cannot make the outcome better (or worse). But we commonly think that lives of unrelenting suffering would be worse than not

existing at all. And an outcome containing additional suffering lives would surely be, in that respect, a worse outcome.<sup>47</sup>

To accommodate these intuitions, person-affecting theorists must allow that *intrinsically bad states* (like undeserved suffering) can count as *non-comparative harms*. Even if, strictly speaking, we cannot compare existence to non-existence, we can certainly hold that a life of unrelenting suffering is *bad for you*. The person-affecting restriction can then be tweaked to specify that one outcome can be worse than another if *either* it's worse for someone *or* it's bad for someone (while the alternative is not).

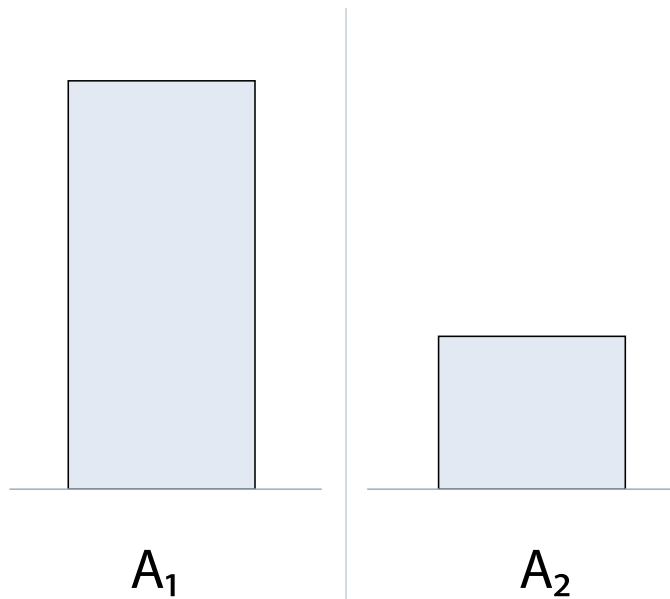
Generalizing this reasoning would lead us to similarly hold that intrinsically *good* states (or positive welfare) can count as a non-comparative benefit of existence.<sup>48</sup> But this result would rob the person-affecting view of its distinctiveness. It could easily end up coinciding with the total view, for example, and endorsing the repugnant conclusion on behalf of the multitudes in world Z who would each receive a tiny non-comparative benefit from getting to exist.

Many person-affecting theorists wish to avoid this result and instead endorse the *procreative asymmetry*, according to which:

**It is bad to create people with negative well-being, but not good to create people with positive well-being, all else equal.**

While many find this principle intuitive,<sup>49</sup> it's notoriously difficult to provide a principled basis for it.<sup>50</sup> The procreative asymmetry also has several deeply problematic implications, stemming from its failure to consider positive lives to be a good thing.

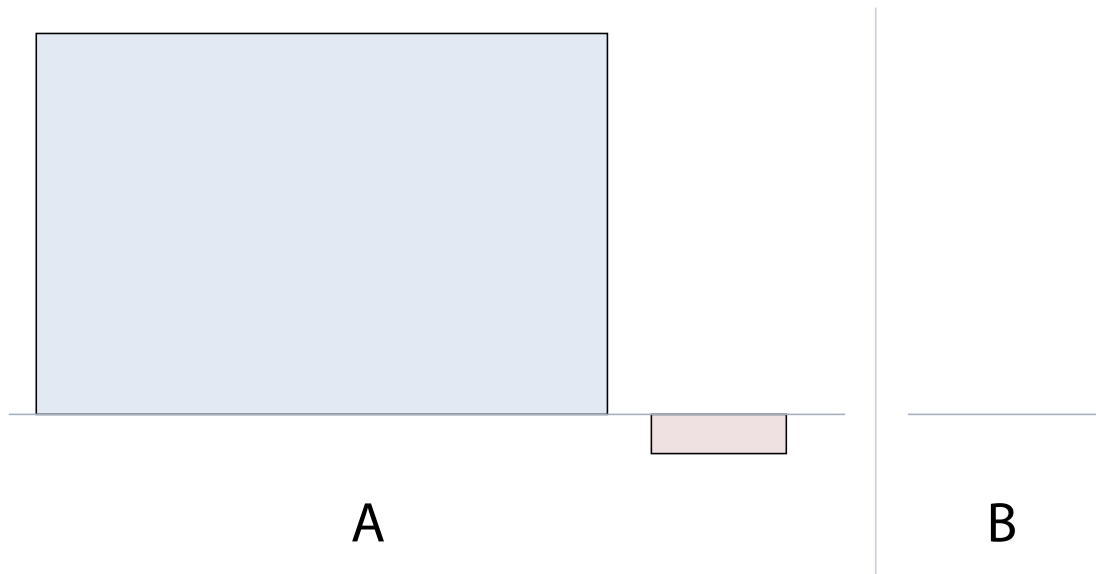
The simplest such view holds that positive lives make *no difference in value* to the outcome. But this falsely implies that creating lives with low positive welfare is just as good as creating an equal number of lives at a high welfare level. For example, consider the choice between creating either of two worlds inhabited by different sets of future people. In world A<sub>1</sub>, everyone has a wonderful life. In world A<sub>2</sub>, all people have lives that are much worse than in A<sub>1</sub> yet still positive.



Clearly, we should prefer world  $A_1$  over  $A_2$ . However, the simplest version of the procreative asymmetry implies that both worlds are equally good—because they are not good at all.

In cases involving comparisons to empty worlds, the simple procreative asymmetry sometimes yields verdicts that seem even more misguided. Consider the following choice between world A and world B:<sup>51</sup>

**In world A, all but a few people have excellent lives. But some people suffer from an extremely rare disease that makes life not worth living. In world B, no people exist.**



Most people would prefer world A over an empty world B. But the simple procreative asymmetry would seem, perversely, to favor the empty world B since it counts the many good lives in world A for nothing while the few bad lives dominate the decision. On this view, there are no worthwhile trade-offs between good and bad lives. It would be better, supposedly, to have no lives at all.

To help address these problems, we may consider a more complex person-affecting view—one analogous to the critical range theory, discussed above, but with infinite value blur, yielding the result that *all* (positive) lives are “meh”.<sup>52</sup> On such a view, it’s better to create a flourishing life than a mediocre one (so  $A_1$  is indeed better than  $A_2$ , at least if they contain the exact same number of people). However, *either* choice is merely on a par with creating neither.

But this brings us to a deeper problem with the procreative asymmetry, which is that it has trouble accounting for the idea that *we should be positively glad that the world (with all its worthwhile lives) exists*.<sup>53</sup> Granted, the immense incomparability introduced by all the putatively “meh” lives in A at least blocks the perverse conclusion that we must outright prefer the empty world B. Even so, holding the two worlds to be incomparable or “on a par” also seems wrong.

We should recognize that A is better. But to do that, we must reject the strict procreative asymmetry and hold that there is an upper limit to the “critical range” of lives that are merely meh. And this is independently plausible. After all, when thinking about what makes some possible universe *good*, the most obvious answer is that it contains a predominance of awesome, flourishing lives. How could that *not* be better than a barren rock? Any view that denies this verdict is arguably too nihilistic and divorced from humanistic values to be worth taking seriously.

We may also raise more theoretical objections to such a view. Person-affecting views typically fall afoul of one or more of the following problems:<sup>54</sup>

1. having moral rankings change when “irrelevant alternatives” are introduced (such as preferring A to B when they are the only choices, but then B over A when an inferior option C is also available);
2. having cyclical orderings of value (saying that A is better than B, B is better than C, and C is better than A); and
3. saying that all outcomes that differ even slightly in how many people exist are incomparable.

For all these reasons, utilitarians are largely united in rejecting person-affecting views, even as they continue to debate which impersonal theory provides the best way forward.<sup>55</sup>

## Practical Implications of Population Ethical Theories

Population ethics is a field of great importance for real-world decision-making. In particular, which population ethical view we adopt shapes the importance we should assign to preserving and improving the long-term future of humanity. If human civilization persists, and perhaps one day spreads to the stars, there could be an inconceivably large number of future people with good lives. Their existence and well-being depend in part on the choices we make today: especially how many resources we, the current generation, invest in [preventing existential risks](#) that threaten the continued survival and long-term flourishing of humankind.<sup>56</sup>

The total well-being enjoyed by all future people is potentially enormous. Consequently, on the total view, the disvalue of losing our future is immense, and the mitigation of existential risks becomes correspondingly important. The same conclusion also holds for critical level (or range) theories, assuming that the average well-being of future generations exceeds the critical level (or range).<sup>57</sup>

Even on the average view, there is reason to expect the long-term future to make a big difference to the overall value of the world. Human welfare has improved dramatically in recent centuries due to technological, social, and moral progress. Consequently, our generation is driving up the average of well-being to date.<sup>58</sup> Further scientific and medical breakthroughs will likely continue to improve the average quality of life in the future. Therefore, even on the average view, it should be a priority to avoid existential risks by virtue of the large future gains in average well-being. The same is true for variable value theories, as they tend to approximate the average view for large populations.<sup>59</sup>

Proponents of standard person-affecting views (with a strict procreative asymmetry) are sceptical of the claim that reducing existential risk is of enormous importance. They would still think that reducing these risks has some value because this reduces the risk of death for those alive today. But they would not regard the absence of future generations as an intrinsic moral loss. However, while

these strict asymmetric views may not be concerned about the prospect of human extinction, they may seriously worry about the possibility of a dystopian future containing many miserable lives. Averting such a negative future would be critically important on these views.<sup>60</sup> Others might be drawn to a weaker (and correspondingly more plausible) version of the asymmetry, according to which we do have some reason to create flourishing lives, but *stronger* reason to help existing people or to avoid lives of negative well-being. On these moderate views, ensuring that the future goes well would still be very important since so many lives are at stake.

Finally, there is an argument from moral uncertainty: given the difficult terrain of population ethics, we may not be entirely confident of any particular view. Therefore, we should figure out what degree of belief we ought to have in each theory and then take the action that is the best compromise between those theories. As we've seen, many plausible theories agree that improving the long-term future is of great moral importance. Therefore, unless one can be extremely confident in standard person-affecting views, it would seem prudent to pay heed to this conclusion and take significant steps to safeguard our future.<sup>61</sup>

## Conclusion

Our actions affect the quality, quantity, and identity of future lives. Population ethics deals with the thorny moral issues arising from such effects on future generations.

According to the total view of population ethics, an outcome's goodness depends only on the total well-being, which may be increased by either improving existing people's lives or creating more happy people. In contrast, the average view considers only average well-being, and so only regards above-average lives as contributing (positive) value to the world. Variable value theories seek to better reflect commonly held intuitions about population ethics by approximating the total view for small populations and the average view for large populations. Critical level (or range) theories hold that adding an individual makes an outcome better to the extent that their well-being exceeds some critical level (or range). Finally, person-affecting views deny that additional lives ever make the outcome (non-instrumentally) better.

All of these views are subject to serious objections. The total view entails the repugnant conclusion, according to which for any world A (however idyllic), there is a better world Z in which no one has a life that is more than barely worth living. The average view, variable value theories, and critical level theories all entail versions of the sadistic conclusion; that it can sometimes be better to create (few) lives with negative well-being than to create (more) lives with positive well-being. Person-affecting views rely on an unsupported asymmetry and struggle to explain the value of existence (even in the most idyllic cases). Critical range theories may do better, but even they cannot support the anti-repugnance intuition that an idyllic world A is strictly better than the repugnant world Z. The ubiquity of these problems is no coincidence: impossibility theorems prove that no population ethical theory can satisfy all the intuitive principles or axioms that we might have hoped for.

The most important practical implications of population ethics concern how much value we should assign to preserving and improving the long-term future of humanity, and hence how important it is to [reduce existential risks](#).

The next chapter discusses the most important implications of utilitarianism for how we should think about leading an ethical life.

Next Chapter: Utilitarianism and Practical Ethics

---

## Como Citar esta Página

```
Chappell, R.Y., Meissner, D., and MacAskill, W. (2023). Population Ethics. In R.Y. Chappell, D. Meissner e W. MacAskill (eds.), An Introduction to Utilitarianism, <https://www.utilitarismo.net/population-ethics>, acessado em 22/05/2025.
```

---

## Resources and Further Reading

### General Discussions of Population Ethics

- Gustaf Arrhenius (2000). [Future Generations: A Challenge for Moral Theory](#). PhD Dissertation, Uppsala University.
- Richard Y. Chappell (2021). Section 7: Population Ethics, [Parfit's Ethics](#). Cambridge: Cambridge University Press.
- Hilary Greaves (2017). [Population Axiology](#). *Philosophy Compass*. 12(11).
- Derek Parfit (1984). Part Four: Future Generations, [Reasons and Persons](#). Oxford: Clarendon Press.

### The Total View and Repugnant Conclusion

- Gustaf Arrhenius, Jesper Ryberg, & Torbjörn Tännsjö (2017). [The Repugnant Conclusion](#). *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.).
- Gustaf Arrhenius (2003). The Very Repugnant Conclusion. In Krister Segerberg & Ryszard Sliwinski (eds.), *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Åqvist*. Uppsala, pp. 29–44.

- Johan E. Gustafsson (2022). [Our Intuitive Grasp of the Repugnant Conclusion](#). In Gustaf Arrhenius, Krister Bykvist, Tim Campbell, and Elizabeth Finneron-Burns (eds.), *The Oxford Handbook of Population Ethics*. Oxford University Press.
- Spears, D. & Budolfson, M. (2021). [Repugnant conclusions](#). *Social Choice and Welfare*. 28.
- Michael Huemer (2008). [In Defence of Repugnance](#). *Mind*. 117(468): 899–933.
- Torbjörn Tännsjö (2002). [Why We Ought to Accept the Repugnant Conclusion](#). *Utilitas*. 14(3): 339–359.
- Stéphane Zuber et al. (2021). [What Should We Agree on about the Repugnant Conclusion?](#) *Utilitas*. 33(4): 379–383.

## Variable Value Theories

- Theodore Sider (1991). [Might Theory X be a theory of diminishing marginal value?](#) *Analysis*. 51(4): 265–271.
- Thomas Hurka (1983). [Value and Population Size](#). *Ethics*, 93(3): 496–507.

## Critical Level and Critical Range Theories

- Charles Blackorby, Walter Bossert, & David J. Donaldson (1995). [Intertemporal Population Ethics: Critical-Level Utilitarian Principles](#). *Econometrica*, 63(6): 1303–1320.
- Charles Blackorby, Walter Bossert, & David J. Donaldson (2005). *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. Cambridge: Cambridge University Press.
- John Broome (2004). *Weighing Lives*. Oxford: Oxford University Press.
- Johan Gustafsson (2020). [Population axiology and the possibility of a fourth category of absolute value](#). *Economics & Philosophy*, 36: 81–110.

## Neutrality Intuition

- Jeff McMahan (2013). [Causing People to Exist and Saving People's Lives](#). *Journal of Ethics*. 17: 5–35.
- Wlodek Rabinowicz (2009). [Broome and the Intuition of Neutrality](#). *Philosophical Issues*. 19(1): 389–411.
- Kryster Bykvist (2007). [The Benefits of Coming into Existence](#). *Philosophical Studies*. 135(3), 335–362.
- John Broome (2005). [Should We Value Population?](#) *The Journal of Political Philosophy*. 13(4): 399–413.
- Elizabeth Harman (2004). [Can We Harm and Benefit in Creating?](#) *Philosophical Perspectives*. 18: 89–113.

## Person-Affecting Views and the Asymmetry

- Gustaf Arrhenius (2009). [Can the Person Affecting Restriction Solve the Problems in Population Ethics?](#), in Roberts, M.A., Wasserman, D.T. (eds.) *Harming Future Persons*. International Library of Ethics, Law, and the New Medicine. 35. Springer, Dordrecht.
- Daniel Cohen (2019). [An Actualist Explanation of the Procreation Asymmetry](#). *Utilitas*. 32(1): 70–89.
- Johann Frick (2014). [‘Making People Happy, Not Making Happy People’: A Defense of the Asymmetry Intuition in Population Ethics](#). Doctoral dissertation, Harvard University.
- Johann Frick (2020). [Conditional Reasons and the Procreation Asymmetry](#). *Philosophical Perspectives*. 34(1): 53–87
- Caspar Hare (2007). [Voices from Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?](#). *Ethics*, 117(3): 498–523.
- Jan Narveson (1973). [Moral Problems of Population](#). *The Monist*. 57(1): 62–86.
- Josh Parsons (2002). [Axiological Actualism](#). *Australasian Journal of Philosophy*. 80(2): 137–147.
- Melinda A. Roberts (2002). [A New Way of Doing the Best That We Can: Person-Based Consequentialism and the Equality Problem](#). *Ethics*, 112(2): 315–350.

## Practical Implications of Population Ethics

- Toby Ord (2020). [The Precipice: Existential Risk and the Future of Humanity](#). Bloomsbury Publishing.
- David Althaus & Lukas Gloor (2018). [Reducing Risks of Astronomical Suffering: a Neglected Priority](#). *Center on Long-Term Risk*.
- Nick Beckstead (2013). [On the Overwhelming Importance of Shaping the Far-Future](#). PhD Dissertation, Rutgers University.
- William MacAskill (2022). [What We Owe the Future](#). Basic Books.

## Impossibility Theorems in Population Ethics

- Gustaf Arrhenius (2000). [An Impossibility Theorem for Welfarist Axiologies](#). *Economics and Philosophy*. 16(2): 247–266.
- Philip Kitcher (2000). [Parfit’s Puzzle](#). *Noûs*. 34(4): 550–577.
- Erik Carlson (1998). [Mere addition and two trilemmas of population ethics](#). *Economics and Philosophy*. 14(2): 283–306.
- Yew-Kwang Ng (1989). [What should we do about future generations? Impossibility of Parfit’s Theory X](#). *Economics and Philosophy*. 5(2): 235–253.

- 
1. When we talk about populations, we mean total populations: not just how many people are alive at a specific time, but consideration of all people across all time. ↩
  2. Other writers, following Parfit (1984), sometimes speak of a “wide person-affecting view” which allows for (non-instrumental) reasons to add happy lives. For ease of expression, in this article we use “person-affecting” in the more distinctive *narrow* sense which rejects this idea. ↩
  3. Throughout this article, we use the terms “quality of life”, “welfare”, and “well-being” interchangeably. These terms are used to describe how well or poorly someone’s entire life goes, not just how well-off someone is at a specific moment in time. Moreover, concepts such as “units of well-being” and “well-being levels” are simplifications used for illustrative purposes, and they do not imply that we can in practice precisely measure well-being. ↩
  4. An alternative method is to add up the well-being levels of all individuals. ↩
  5. The notion of a positive life, which is critical for the total view, only makes sense relative to a zero point on the well-being scale. This zero point is the threshold above which life becomes “worth living”. A “neutral life”, at well-being level 0, is neither “worth living” nor “not worth living”. This may be either a life with no value or disvalue, or a life with exactly as much value as disvalue.

For discussion of the subtleties surrounding the concept of a life “worth living”, see Broome, J. (2004). *Weighing Lives*. Oxford: Oxford University Press, pp. 66–68. ↩

6. Stronger still: on the total view, it would be intrinsically *better* to create a new person at welfare level 100 than to improve an existing person’s well-being from level 1 to 100. ↩
7. For an exploration of whether the world is overpopulated or underpopulated, see Ord, T. (2014). *Overpopulation or Underpopulation?*, in Goldin, I. *Is the Planet Full?*. Oxford: Oxford University Press. ↩
8. Reducing existential risk is a priority not just for the total view but for a wide variety of moral views. However, for the total view there is an especially large amount of value at stake with preserving the long-term flourishing of civilization.  
  
Ord, T. (2020). Chapter 2: Existential Risk, in *The Precipice: Existential Risk and the Future of Humanity*. London: Bloomsbury Publishing. ↩

9. While Parfit deserves credit for raising the philosophical debate around the repugnant conclusion, arguably Henry Sidgwick noticed it much earlier, writing that “the point up to which, on utilitarian principles, population ought to be encouraged to increase, is not that at

which the average happiness is the greatest possible... but that at which the happiness reaches its maximum”.

Sidgwick, H. (1907). *The Methods of Ethics*, 7th edition. London: Macmillan, p. 418. ↻

10. This formulation is taken from Greaves, H. (2017). [Population Axiology](#). *Philosophy Compass*. 12(11).

Parfit’s longer original formulation was as follows: “For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living.”

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press, p. 342.

The total view further implies the *very repugnant conclusion*, according to which enough lives barely worth living can outweigh any number of additional arbitrarily miserable lives.

Arrhenius, G. (2003). The Very Repugnant Conclusion. In Krister Segerberg & Ryszard Sliwinski (eds.), *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Åqvist*. Uppsala, pp. 29–44. ↻

11. Stéphane Zuber et al. (2021). [What Should We Agree on about the Repugnant Conclusion?](#). *Utilitas*. 379–83. ↻

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

Huemer, M. (2008). [In Defence of Repugnance](#). *Mind*. 117(468): 899–933.

Gustafsson, J. (2022). [Our Intuitive Grasp of the Repugnant Conclusion](#). In Gustaf Arrhenius, Krister Bykvist, Tim Campbell, and Elizabeth Finneron-Burns (eds.), *The Oxford Handbook of Population Ethics*. Oxford University Press. ↻

13. Ryberg, J. (1996). [Is the Repugnant Conclusion Repugnant?](#). *Philosophical Papers*, 25: 161–177.

Tännsjö, T. (1992). [Who are the Beneficiaries?](#). *Bioethics*, 6(4): 288–296.

Mackie, J. L. (1985). Parfit’s Population Paradox, in J. Mackie & P. Mackie (eds.) *Persons and Values*. Oxford: Oxford University Press. ↻

14. This was how Parfit sometimes characterized the “repugnant” world Z, for example in his (1986) [Overpopulation and the Quality of Life](#). In Peter Singer (ed.), *Applied Ethics*. Oxford: Oxford University Press. ↻

15. Hutchinson, M. (2014). *The Ethics of Extending and Creating Life*. Unpublished DPhil dissertation, University of Oxford. ↻

16. Tännsjö, T. (2002). [Why We Ought to Accept the Repugnant Conclusion](#). *Utilitas*, 14(3): 339–59.
- Huemer, M. (2008). [In Defence of Repugnance](#). *Mind*. 117(468): 899–933. ↩
17. Parfit, D. (1984). [Reasons and Persons](#). Oxford: Oxford University Press. Chapter 19. ↩
18. At least on standard assumptions. As we will see below, this no longer follows if, in addition to the standard trichotomy of value relations (being *greater than*, *lesser than*, and *precisely equal*), there is a fourth relation of being *on a par*. For then, B might be better than A+, while both A+ and B are merely on a par with A. ↩
19. Gustaf Arrhenius (2000). [An Impossibility Theorem for Welfarist Axiologies](#). *Economics and Philosophy*. 16(2), 247–266.
- Gustaf Arrhenius (2011). [The Impossibility of a Satisfactory Population Ethics](#). In Ehtibar N. Dzhafarov and Lacey Perry (eds.), *Descriptive and Normative Approaches to Human Behavior*. Singapore: World Scientific Publishing Co. 1–26.
- Philip Kitcher (2000). [Parfit's Puzzle](#). *Noûs*. 34(4), 550–577.
- Erik Carlson (1998). [Mere Addition and Two Trilemmas of Population Ethics](#). *Economics and Philosophy*. 14(2), 283–306.
- Yew-Kwang Ng (1989). [What Should We Do About Future Generations? Impossibility of Parfit's Theory X](#). *Economics and Philosophy*. 5(2), 235–253. ↩
20. Greaves, H. (2017). [Population Axiology](#). *Philosophy Compass*. 12(11). ↩
21. E.g. Tännsjö, T. (2002). [Why We Ought to Accept the Repugnant Conclusion](#). *Utilitas*, 14(3): 339–59. Huemer, M. (2008). [In Defence of Repugnance](#). *Mind*. 117(468): 899–933. ↩
22. Note that the average and total views *always* agree on the ranking of outcomes when these outcomes contain the same number of individuals. In such cases, both theories are said to be “extensionally equivalent”. ↩
23. This is a variation on the *Hell Three* case from Parfit (1984), p. 422. ↩
24. Definition adapted from Arrhenius, G. (2000). [An Impossibility Theorem for Welfarist Axiologies](#). *Economics and Philosophy*. 16(2): 247–266.

Note that this label is misleading. Whereas the previous objection noted how the average view sometimes favours the addition of miserable lives (which does seem sadistic), the so-called “sadistic conclusion” instead objects to the claim that adding a small number of negative-welfare lives may be *less bad* than adding a vast number of moderately-positive ones to an otherwise very happy world. Insofar as this comparative judgment stems from evaluating

(some) positive-welfare lives negatively, rather than negative-welfare lives positively, the “sadism” label seems a misnomer. ↩

25. Where  $f$  is a strictly increasing and strictly concave function with a horizontal asymptote. That is, as  $N$  increases,  $f(N)$  increases at a diminishing pace, and never surpasses a certain limit.

Cf. Hurka, T. (1983). [Value and Population Size](#). *Ethics*, 93(3): 496–507.

Ng, Y. (1989). [What Should We Do About Future Generations?](#). *Economics and Philosophy*. 5(2): 235–253. ↩

26. Greaves, H. (2017). [Population Axiology](#). *Philosophy Compass*. 12(11). ↩

27. They may avoid the repugnant conclusion since the total value of a population is subject to an upper limit if the value of additional lives diminishes asymptotically.

Cf. Greaves, H. (2017). [Population Axiology](#). *Philosophy Compass*. 12(11). ↩

28. Parfit, D. (1984). Chapter 18: The Absurd Conclusion, in [Reasons and Persons](#). Oxford: Oxford University Press. ↩

29. Consider a good world with one billion happy people and a single miserable person. Imagine we repeatedly increase the numbers of happy and miserable people by the same factor—ten billion happy people and ten miserable people; one hundred billion happy people and one hundred miserable people, etc. For a sufficiently large population, every time we increase its size in this way, the world gets worse (according to these asymmetric theories) until we eventually reach an overall bad world. ↩

30. For one such attempt, see section 7.2.2 of Chappell, R.Y. (2021). [Parfit's Ethics](#). Cambridge: Cambridge University Press. ↩

31. The critical level is generally assumed to be non-negative, that is either positive or zero. A negative critical level would implausibly attribute positive value to (some) negative lives. ↩

32. Cf. Broome, J. (2004). [Weighing Lives](#). Oxford: Oxford University Press, p. 213–214. ↩

33. Blackorby, C., Bossert, W. and Donaldson, D. (1996). [Quasi-Orderings and Population Ethics](#). *Social Choice and Welfare*, 13 (2): 129–50.

Rabinowicz, W. (2009). [Broome and the Intuition of Neutrality](#). *Philosophical Issues*, 19 (1): 389–411.

Gustafsson, J. (2020). [Population axiology and the possibility of a fourth category of absolute value](#). *Economics & Philosophy*, 36: 81–110. ↩

34. While philosophers do not typically use this colloquial term, it may be helpful to imagine someone who, when asked whether the world is improved by adding a barely-worth-living life, shrugs their shoulders and responds, “meh”. ↩
35. Cf. Chang, R. (2002). [The possibility of parity](#). *Ethics*, 112 (4): 659–688. ↩
36. One may, for example, get this result by thinking of the critical range as representing a range within which it’s *indeterminate where the critical level lies*. Or one may consider it a range of *reasonable pluralism*, such that one could permissibly treat any point in this range as the critical level when forming personal preferences about which lives to add (or not) to the world. On either approach, we can then *supervalue*, and hold that population X is (truly, determinately, or objectively) better than Y just if this evaluation would follow from *all* critical level theories where the critical level falls within the specified range. ↩
37. Gustafsson speaks of “undistinguishedness” in place of our term “value blur”. See:  
  
Gustafsson, J. (2020). [Population axiology and the possibility of a fourth category of absolute value](#). *Economics & Philosophy*, 36: 81–110. ↩
38. The following illustration is adapted from p.92 of Gustafsson, J. (2020). [Population axiology and the possibility of a fourth category of absolute value](#). *Economics & Philosophy*, 36: 81–110. ↩
39. This may prove especially problematic if the critical range is symmetric, such that slightly negative lives also qualify as meh (rather than bad). Most would think the idyllic world A should be strictly better than a world “Z-” containing a vast number of slightly negative lives, but symmetric critical range theories will struggle to deliver this verdict. Cf. Gustafsson (2020), p. 95, who claims that this is not a terrible bullet to bite if the personal/well-being value of such a life is also blurry, and so “meh” (in our terminology) rather than bad *for the person living it*.  
  
Gustafsson, J. (2020). [Population axiology and the possibility of a fourth category of absolute value](#). *Economics & Philosophy*, 36: 81–110. ↩
40. Here we use the term “impersonal” simply to contrast with the (narrow) person-affecting view. So even an impersonal theory (on our usage) can take the value of a state of affairs to be grounded in facts about what is good for particular people, e.g. if one holds that a happy existence can constitute a non-comparative benefit, parallel to how a miserable existence constitutes a non-comparative harm. ↩
41. The critical range theories we’ve discussed are more difficult to capture in a numerical formula. But if we put aside any lives that fall within the critical range, we might approximate the remaining overall value by separately summing the positive and negative welfare levels, to the extent that they exceed their respective ends of the critical range. So, using positive and

negative subscripts to denote the respective numbers (N), averages (Q), and critical boundary points ( $\alpha$ ) for these two subpopulations, we can write the value function as:

$$N_+ * (Q_+ - \alpha_+) + N_- * (Q_- - \alpha_-).$$

Alternatively, we might think of the value of the world as itself indeterminate or corresponding to a numerical range, as given by all the possible critical level theories corresponding to the critical range, i.e.: everything from  $N * (Q - \alpha_-)$  to  $N * (Q - \alpha_+)$ . On this approach, the value of the repugnant world Z, for example, would be indeterminate across an immense range of values stemming from *extremely negative* to *extremely positive*, because N is unimaginably huge whereas the sign of  $(Q - \alpha)$  is indeterminate, given the range of candidate critical levels  $\alpha$ . ↩

42. On the total view, adding a person with positive well-being is always good, all else equal. The same is true for variable value theories, though for a large pre-existing population the marginal value of an additional life being added may be low. On the average view, adding a person is good, other things being equal, if the person's well-being exceeds the existing average; similarly, it is good on critical level (and range) theories, if the person's well-being exceeds the critical level (or range). ↩

43. Narveson, J. (1973). [Moral Problems of Population](#). *The Monist*. 57(1), 62–86, p. 80.

To put pressure on the slogan, consider the limited appeal of its negative-welfare analogue: “We are opposed to making people miserable, but neutral about making miserable people.” As this section goes on to discuss, the viability of person-affecting approaches to population ethics crucially depends upon validating a radical asymmetry between positive and negative lives. ↩

44. For instance, see:

Finneron-Burns, E. (2017). [What's wrong with human extinction?](#). *Canadian Journal of Philosophy*, 47(2-3): 327–343.

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.

Roberts, M.A. (2002). [A New Way of Doing the Best That We Can: Person-based Consequentialism and the Equality Problem](#). *Ethics*, 112(2): 315–350.

Roberts, M.A. (2004). [Person-Based Consequentialism and the Procreation Obligation](#), in J. Ryberg and T. Tännsjö (eds.) *The Repugnant Conclusion*. Library Of Ethics And Applied Philosophy. 15: 99–128. ↩

45. At least, it cannot be better or worse in terms of well-being. ↩

46. Cf. Arrhenius, G. (2000). Chapter 8, [Future Generations: A Challenge for Moral Theory](#). PhD thesis, Uppsala.
- Heyd, D. (1988). [Procreation and Value: Can Ethics Deal With Futurity Problems?](#). *Philosophia*, 18: 151–170. ↩
47. Cf. Parfit's "wretched child" case. Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press, p. 391. ↩
48. McMahan, Jeff (2013). [Causing People to Exist and Saving People's Lives](#). *The Journal of Ethics*, 17: 5–35.
- McMahan, J. (2009). [Asymmetries in the Morality of Causing People to Exist](#). In David Wasserman and Melinda Roberts (eds.), *Harming Future Persons*. Springer. ↩
49. Though one of the coauthors of this chapter has argued elsewhere that "it is thought to be 'intuitive' primarily because it has been implicitly confused with other, more plausible theses." See Chappell, R.Y. (2017). [Rethinking the Asymmetry](#). *Canadian Journal of Philosophy*, 47 (2): 167–177. ↩
50. McMahan, J. (2009). [Asymmetries in the Morality of Causing People to Exist](#). In David Wasserman and Melinda Roberts (eds.), *Harming Future Persons*. Springer. ↩
51. Case description adapted from Beckstead, N. (2013). [On the Overwhelming Importance of Shaping the Far-Future](#). PhD Dissertation, Rutgers University. ↩
52. A major challenge for such a view would be to explain how to render this value blur compatible with the asymmetry, so that miserable lives are appropriately recognized as bad (not merely meh). ↩
53. At least on the assumption that good lives predominate over miserable ones. E.g., a person in world A should be glad that that world exists. ↩
54. Arrhenius, G. (2000). Chapter 8, [Future Generations: A Challenge for Moral Theory](#). PhD thesis, Uppsala.
- Ord, T. (2020). [The Precipice: Existential Risk and the Future of Humanity](#). London: Bloomsbury Publishing. See especially note 25 of the Appendices. ↩
55. Another possibility would be some form of *hybrid* view combining an impersonal theory with additional person-affecting reasons to prioritize the interests of existing individuals. This would help to block "replaceability" worries—i.e., that impersonal views make it too easy to justify (killing and) replacing existing lives with "better" ones—without the extreme implications of pure person-affecting views.

For a fascinating discussion of some of the complexities of adjudicating tradeoffs between ordinary harms and benefits and the noncomparative good of bringing someone into (happy) existence, see:

McMahan, Jeff (2013). [Causing People to Exist and Saving People's Lives](#). *The Journal of Ethics*, 17 (1-2): 5–35. ↩

56. In addition to existential risk reduction, another promising strategy to improve the long-term future is “moral circle expansion”: increasing the moral concern for members of some outlying groups to include, ideally, all sentient beings.

Anthis, J. & Paez, E. (2021). [Moral circle expansion: A promising strategy to impact the far future](#). *Futures*, 130. ↩

57. This assumption is plausible: with continued technological, social, and moral progress, the average quality of life in the future will likely increase further, as it has for hundreds of years. Only an implausibly high critical level—on which even the value of the average life in the present generation is negative—may render invalid the conclusion that existential risk reduction should be a priority. ↩

58. Although the suffering on factory farms might mean that our generation is driving *down* the average well-being among sentient creatures, once you consider all the suffering-filled animal lives we have created. But technological improvements, particularly the development of [cultivated meat](#) and other animal product alternatives, could render factory farming a temporary phenomenon. ↩

59. If what counts as a “large” population is much greater than the current population, this further increases the importance assigned by variable values theories to avoiding existential risks since they approximate the total view for small populations. ↩

60. For instance, Althaus & Gloor (2018) argue that reducing risks of astronomical future suffering should be an ethical priority.

Althaus, D. & Gloor, L. (2018). [Reducing Risks of Astronomical Suffering: a Neglected Priority](#). *Center on Long-Term Risk*. ↩

61. Hilary Greaves and Toby Ord argue that (given a plausible approach for dealing with moral uncertainty), as the expected number of future lives grows, this “systematically pushes one toward choosing the option preferred by the Total View and critical-level views, even if one’s credence in those theories is low”.

Greaves, H. & Ord, T. (2017). [Moral Uncertainty About Population Axiology](#). *Journal of Ethics and Social Philosophy*. 12(2).

See also:

MacAskill, W., Bykvist, K., & Ord, T. (2020). Chapter 8, Practical Ethics Given Moral Uncertainty, in *Moral Uncertainty*. Oxford: Oxford University Press.

Ord, T. (2020). Appendix B: Population Ethics and Existential Risk, in *The Precipice: Existential Risks and the Future of Humanity*. Hachette Books: New York. ↻