Extinction risk and population ethics

Hilary Greaves (Oxford)
Centre for the Study of Existential Risk, Cambridge
29 April 2016

Outline

- Part I: The badness of extinction (risk)
  - Two perspectives: ‘dismissive’ and ‘serious’
  - Population axiology
  - Averagism, totalism and person-affecting views
  - Conclusion: The ‘serious view’ is correct
- Part II: Optimum population size
  - Arguing for population reduction
    - Averagist arguments
    - Totalist/Long-run arguments
  - Arguments from extinction risk
  - Conclusion: *Some* reason, via extinction risk, to think that even the ‘totalist’ should support population reduction

PART I: THE BADNESS OF EXTINCTION (RISK)

Extinction risk

- Inevitable fact: Humanity (and intelligent/sentient life more generally) will *eventually* go extinct.
- But several factors, both natural and anthropogenic, pose some risk of rendering humanity extinct 'prematurely' – e.g. in the next 100-200 years.
  - Henceforth, I'll take "extinction risk" to mean: risk of such 'premature' extinction, from any source.
- Natural vs anthropogenic extinction risks:
  - Natural: Asteroid collision, natural disease pandemic…
  - Anthropogenic: Nuclear war, climate change, artificial intelligence, synthetic biology…
- Evaluative question: How bad would premature extinction be? Two possible sources of badness:
  - Pain and suffering involved in the process of extinction; plus, *perhaps*,
  - Absence of additional future lives thereby averted.

Two perspectives on the badness of extinction

- ‘Dismissive view’
  - Obviously pain and suffering are bad.
  - But that’s *all* that’s bad about premature extinction. In particular: ‘Lives averted’ does not represent any source of badness, because there’s no (actual) victim in a case of life-aversion.
  - The resulting amount of badness would be relatively moderate.
  - Therefore it’s not worth worrying much about very small risks of extinction.
- ‘Serious view’
  - In addition to the (relatively moderate) badness of the pain and suffering involved, the value lost via valuable lives averted, in a premature extinction event, would be astronomical.
    - E.g. Equivalent to 100,000 times the value of improving all lives by 10% for a century.
  - Therefore the *expected* value of reducing extinction risk, even by a tiny amount, can be very large.
  - Reducing extinction risk (even by a tiny amount) is one of the most important things we can do.

Taking a step back: The question of population axiology

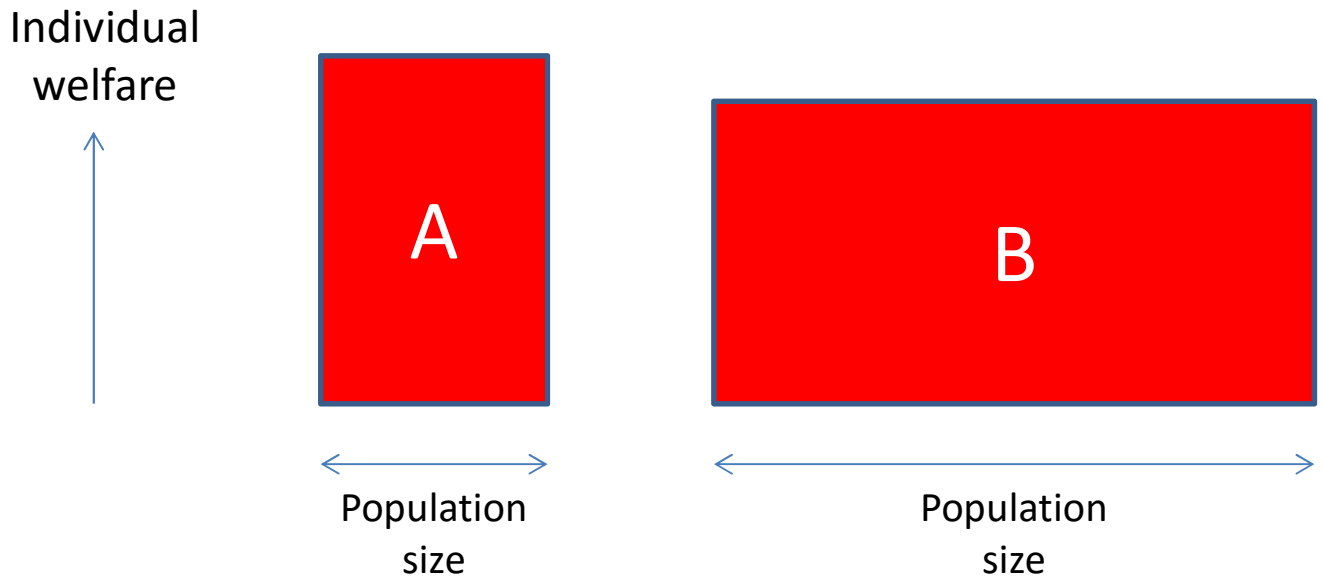
- Axiology: A 'betterness' ranking of states of affairs. (Or, more quantitatively: An assignment of quantities of value to states of affairs, perhaps on a cardinal scale.)
  - Talk of axiology is compatible with:
    - The claim that there is some *vagueness* over which axiology is correct.
    - The claim that the language of *'truth'* and *'correctness'* is in appropriate in evaluative domains.
    - The claim that 'the' 'correct' axiology is an *incomplete* one.
- Fixed-population axiology: An axiology for restricted sets of states of affairs *that agree with one another on* the numbers and identities of persons (and other sentient beings) who ever live.
- Variable-population axiology ("population axiology"): The more ambitious project of ranking states of affairs *that may disagree with one another* over the numbers and/or identities of persons.
  - This more general evaluative theory is crucial for assessing premature-extinction scenarios.
  - Views to be surveyed: 'Averagism', 'Totalism' and 'Person-affecting' views

Fixed-population axiology

- Obviously, a large number of value judgments is involved in any choice of even a *fixed*-population axiology.
  - Which things matter (intrinsically)? The welfare/happiness of sentient creatures, knowledge, friendship, beauty, shared culture, the integrity of natural ecosystems…?
  - *How to aggregate* whichever things do matter?
  - A substantial fragment of moral philosophy deals with these questions.
- I will assume (for the sake of argument) that the right answer in the fixed-population case is the "utilitarian" one: ***goodness = total (or average) welfare***.
  - This sets aside possible non-welfare goods.
    - But we don't really need to be assuming that there *are* no non-welfare goods: the claim could just be that goodness tracks total/average welfare *when all non-welfare goods are equal*.
  - It also assumes that the distribution of a fixed amount of welfare among persons doesn't matter.
    - That is controversial, but for reasons that are largely orthogonal to our present concern.

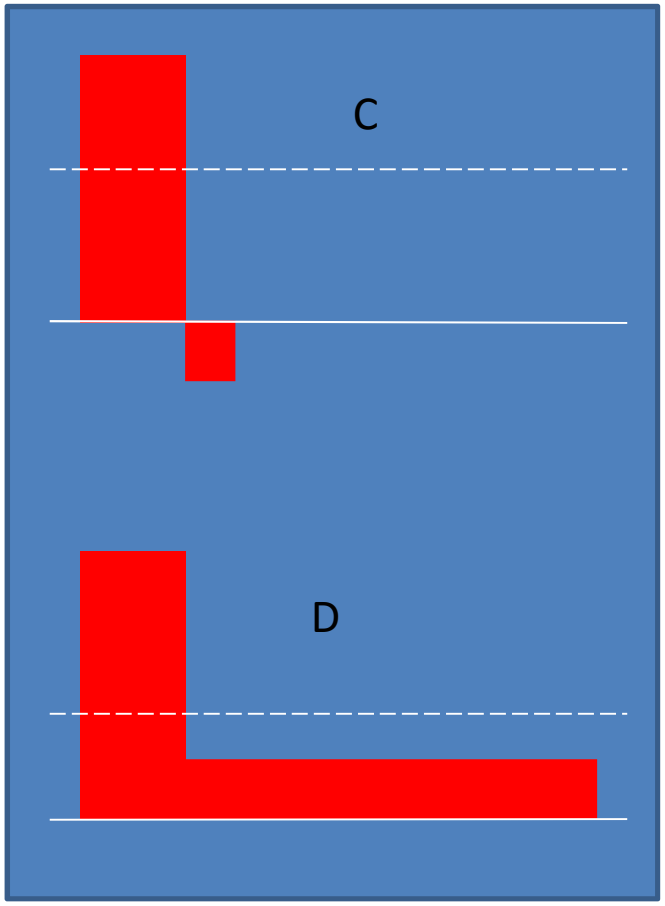
(Variable-)Population axiology

- Fixed-population axiology underdetermines variable-population axiology.
- E.g. settling on the "utilitarian" axiology for the fixed-population case underdetermines whether one is taking goodness to be represented by *total* or by *average* welfare. These come apart in the variable-population case.
  - Which is better, A or B?



Against ‘Averagism’ (I): The ‘Sadistic Conclusion’

- The ‘Sadistic Conclusion’: It is sometimes better to add a smaller number of people with negative well-being (lives not worth living) than a larger number of people with positive well-being (lives worth living). (C>D)
- Averagism entails the Sadistic Conclusion.
- But the Sadistic Conclusion is unacceptable. Therefore Averagism is unacceptable.

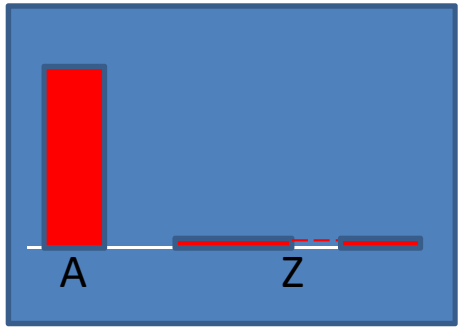


Against Averagism (II): The Egyptology objection

- The relevant average has to be: the average welfare level *of all those who ever live*.
- But in that case, according to Averagism, whether or not I make the world a better place by creating an extra person (with a given welfare level) now depends, in part, on the population sizes and welfare levels in e.g. ancient Egypt.
  - But it's implausible that what I should do now could depend in this sort of way on Egyptology.
  - Therefore Averagism is implausible.

In favour of Totalism

- The ‘totalist’ view does not entail the Sadistic Conclusion, and is not vulnerable to the Egyptology objection.
- It *does* entail the ‘Repugnant Conclusion’: that Z is better than A.



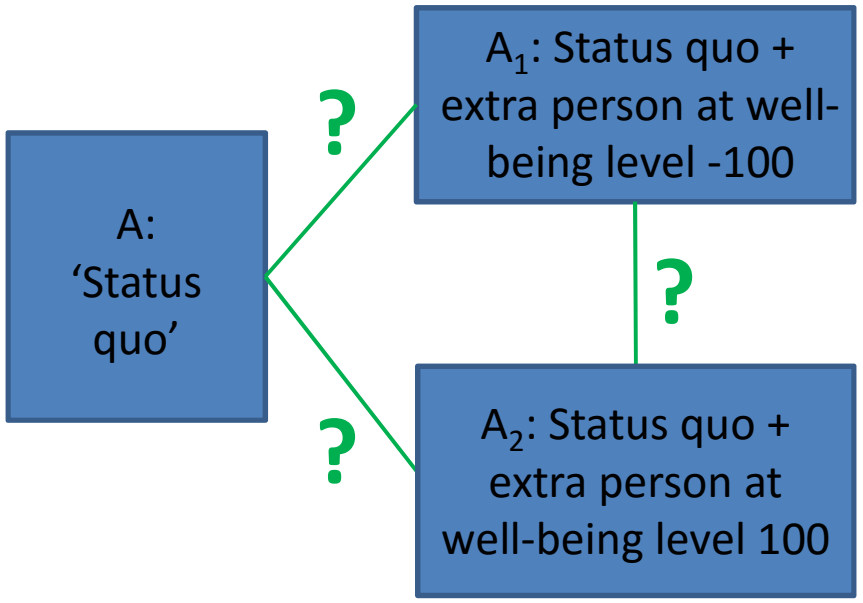
- However:
  - This doesn’t mean that it would be better *in practice* to increase population size indefinitely/to the point at which lives become ‘barely worth living’
  - And it turns out to be remarkably difficult to avoid (even) the (merely hypothetical) Repugnant Conclusion, without falling prey to worse problems.

A third way? ‘Person-affecting’ approaches to population axiology

- Natural thought:
  - What’s bad is if *there are people* who lose most of their lives due to death in childhood. It’s not that ‘non-births’ are tragic, on account of the ‘loss’ of life that is not a loss that befalls any actual person. (Merely possible people don’t matter!)
  - This is because: adding an extra person is in itself *neutral*. (“The intuition of neutrality”)
- Broome (2004): This intuition is extremely natural, *but ultimately incoherent*.
- Three attempts to build a theory capturing the intuition:
  - The ‘equal-goodness principle’
  - Theories of massive incomparability
  - Presentism (and other ‘non-impartial’ theories)

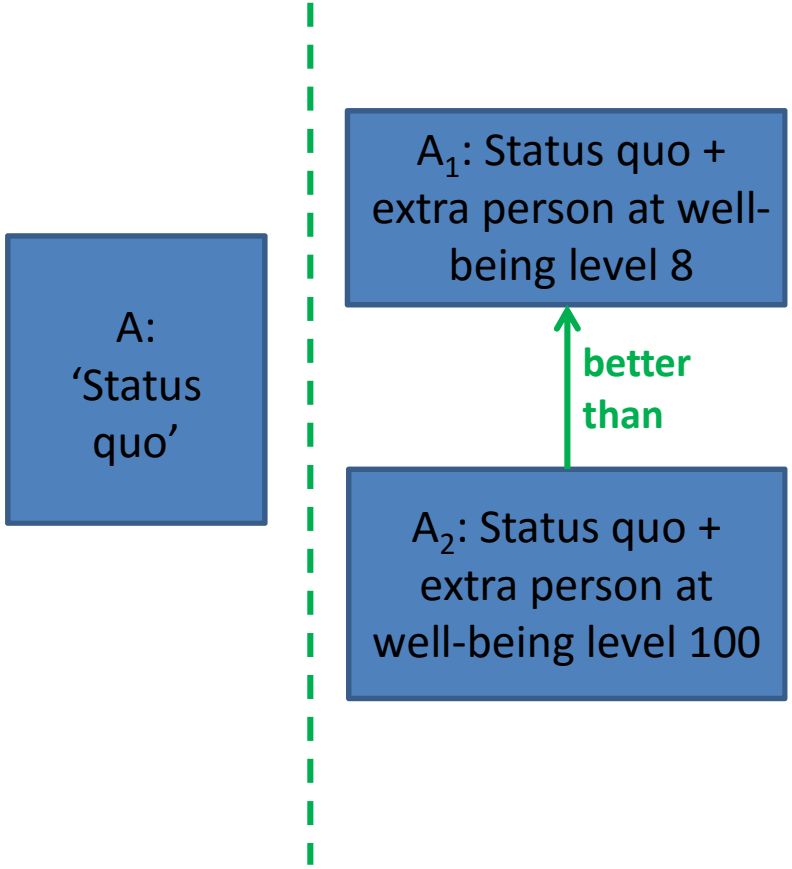
The person-affecting view (I): The equal-goodness principle

- 'Equal-goodness principle': Adding an extra person (others' well-being levels being held fixed) leads to a state of affairs that is *equally as good as* the status quo.
- This principle cannot be true. ($A_2 > A_1$, so we cannot have both $A = A_1$ and $A = A_2$.)



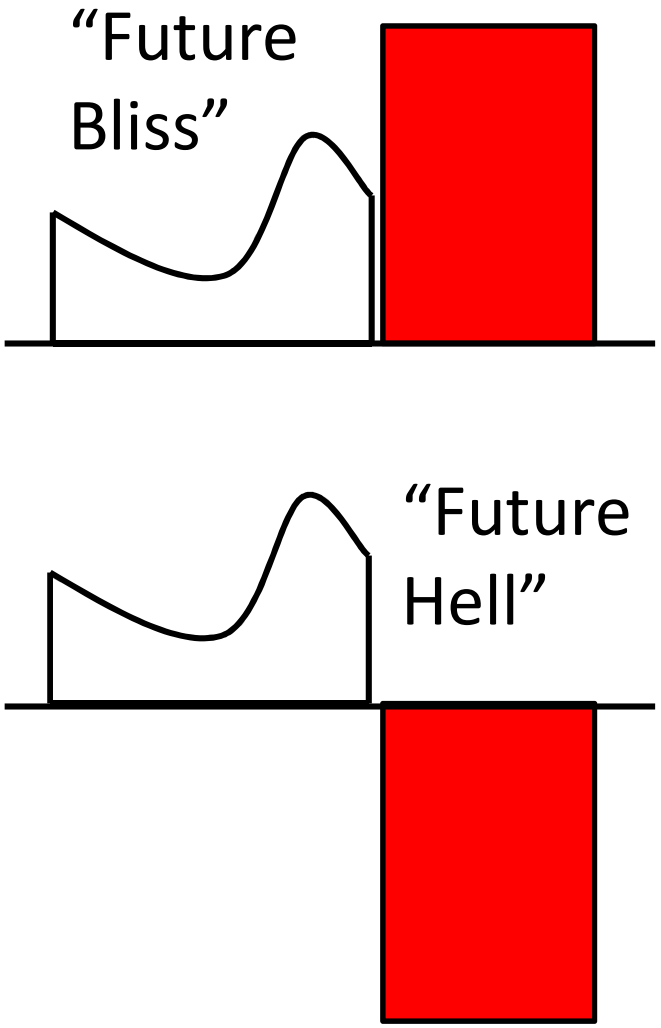
The person-affecting view (II): Theories of massive incomparability

- 'Incomparability principle': (Positive betterness relations hold between same-population states of affairs, but) any two states of affairs according to which different numbers of people live are *incomparable* in terms of goodness.
- But this throws out the baby with the bathwater.
  - Clearly a state of affairs in which *1m people live extremely happy lives* is better than one in which *1m+1 people live extremely miserable lives* (other things equal).
  - So there must be at least *some* positive betterness relations in different-population cases.



The person-affecting view (III): Presentism

- Presentist principle: Only the welfare of *present* people matters morally.
- This view captures the intuition that the loss of welfare directly resulting from ‘non-births’ is not bad (provided the hypothetical births in question are future ones).
- But it too throws out the baby with the bathwater: clearly ‘future bliss’ is better than ‘future hell’ (other things equal).



Back to extinction risk

- The ‘dismissive view’ arises from a person-affecting view on population axiology.
  - But that view is (arguably) discredited, since it turns out to be impossible to make the view precise without embracing some absurdity.
- The ‘serious view’ is entailed by a ‘totalist’ view on population axiology.
  - And this view is (arguably) the front-runner.
- Interim conclusion: Reducing extinction risk, even by tiny amounts, is extremely important. (I.e. the ‘serious view’ is correct.)

PART II: OPTIMUM POPULATION SIZE

Arguing for population reduction

- “Arguing for population reduction”: Arguing that the optimal instantaneous population in the short to medium term, given the actual empirical state of the world, is lower than we are likely to get under ‘business as usual’ scenarios.
  - Population Matters
  - World Overpopulation Awareness
  - Zero Population Growth
  - The Royal Society…

‘Averagist’ arguments for population reduction: From a fixed total resource flow

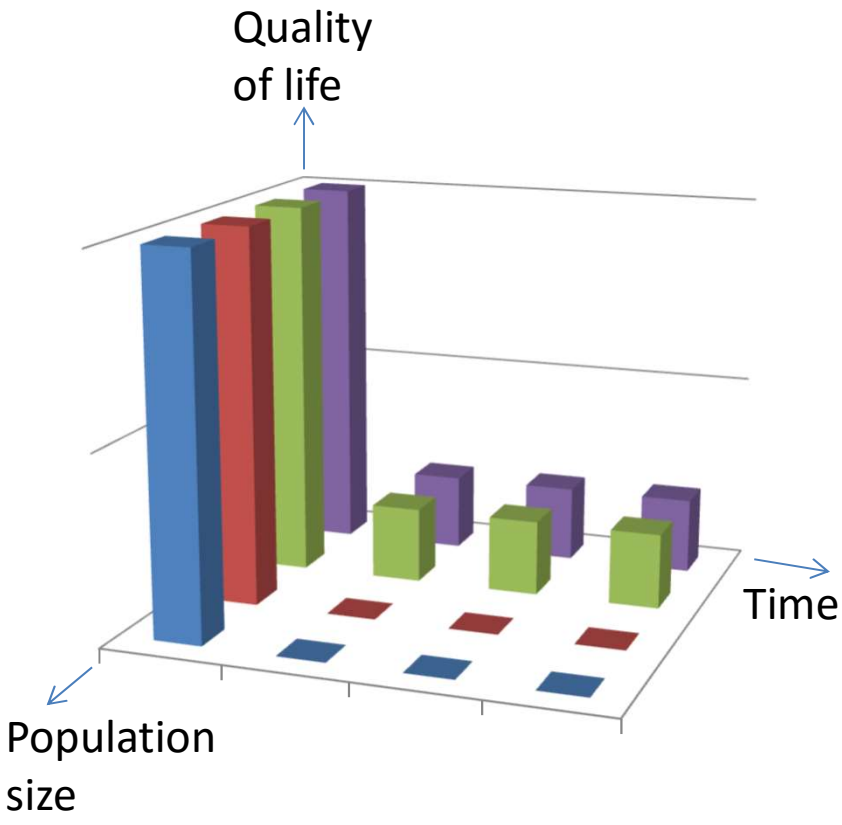
- *Arguably*-fixed [or sublinearly growing] resources per unit time: food, water, physical space, renewable energy…
- *“The more of us there are, the less resources there are for each of us and for the other species with which we share the planet.” Population Matters online article*
- “A diet that includes 40 grams of animal protein per day is probably optimal. … If this criterion is adopted, the world must be considered as overpopulated, as the global average animal protein consumption is 29 grams per capita per day.” *Bernard Gilland (economist)*
- These arguments are controversial even on purely empirical grounds
  - E.g. “Every human being represents hands to work, and not just another mouth to feed” (George W Bush)
- But even if the empirical premises of these arguments are granted, as an evaluative matter *these arguments presuppose Averagism*.
  - In particular, they will not convince a ‘totalist’.
  - In general, it is obviously more difficult to convince a totalist than an averagist of the case for population reduction.

Arguments from long-run considerations

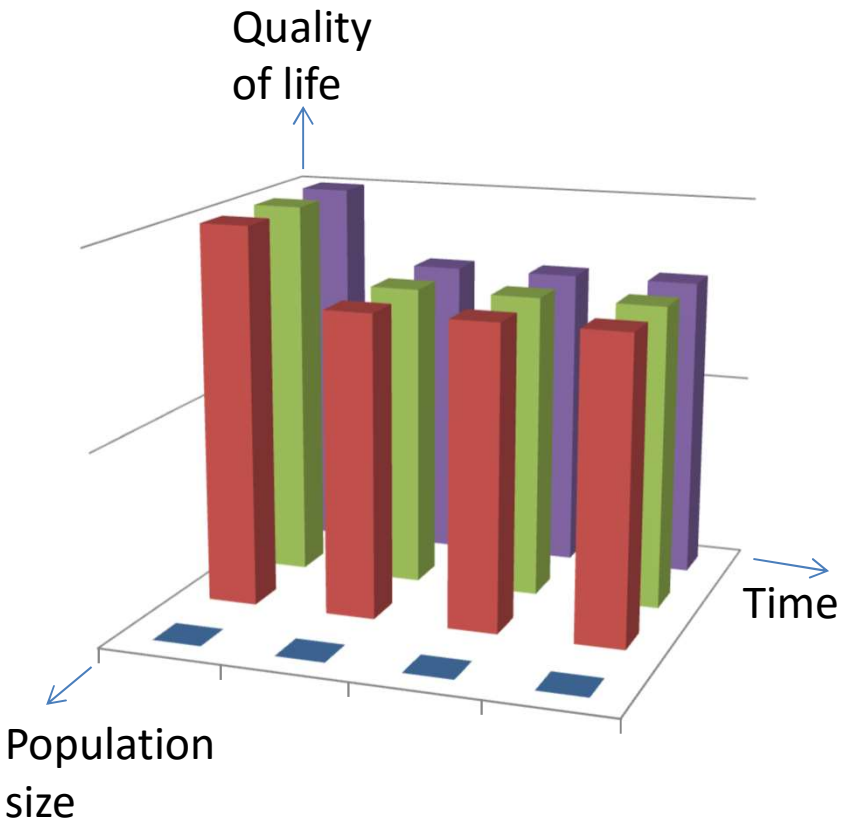
- Basic idea/claim: Overpopulation now would lead to lower quantity/quality of life in the long run
  - Climate change could lead to a loss of “5% of global GDP each year, now and forever” (Stern Review, 2007)
  - “Environmental degradation, including climate change and resource depletion, is steadily reducing the number of people the Earth can indefinitely sustain.” (“Population policy and the environment” position statement, Population Matters, January 2014)
- These arguments, if empirically sound, could convince (even) the totalist...

Totalists care about the long run

Scenario 1: Higher population now, lower quantity and quality of life in future



Scenario 2: Lower population now, higher quantity and quality of life in future



Extinction risk: The long-run argument *par excellence*

- *If* a larger instantaneous population in the short term increases extinction risk, then this could easily be enough to convince totalists of the case for population reduction.
  - (Conversely, if it *decreases* extinction risk, then totalists are likely to favour population *increase*.)
- Million-dollar question: *Would* a larger instantaneous population in the short term increase extinction risk?

Population size and exinction risk (I): The “per capita consumption effect”

- Changes to instantaneous population size would affect per capita consumption.
  - Perhaps with positive correlation, e.g. due to economies of scale.
  - Or perhaps with negative correlation, e.g. due to thinner spreading of limited resources.
- And changes in per capita consumption would affect extinction risk.
  - Perhaps with positive correlation, e.g. due to increased ability of a richer society to develop dangerous technologies.
  - Or perhaps with negative correlation, e.g. due to increased competition for scarce resources leading to war.
- The upshot of combining all these considerations is very unclear.

Population size and extinction risk (II): the “differential development effect”

- The concept of “differential technological development”
  - In the long term, provided that humanity doesn’t become extinct (or civilisation doesn’t collapse, etc.) first, every feasible technology is likely to be developed.
  - And nothing we can do to speed up or slow down the development of any given technology is likely to be very significant *by itself*.
  - However, the *relative* order of the development of different technologies can make an enormous difference, in particular to extinction risk.
    - E.g.: AI vs mechanisms for ensuring that AI is safe
- But the term “technology” here is potentially misleading: the same considerations apply in principle to *any* kinds of progress.
  - In particular, they apply equally to moral, social or institutional progress.

Sketch of a “differential development”-based argument for population reduction

- Suppose (very plausibly) that most existential risks
  - Are anthropogenic, and are related to advances in science and technology.
  - Are mitigated by having good institutional structures (e.g. healthy public decision-making processes).
- In that case, the key question is the pace of scientific/technological progress *relative to* that of institutional progress.
- Increasing instantaneous population size presumably speeds up the relevant scientific/technological progress, in terms of ‘clock time’.
  - Plausibly, this progress rate scales at least close to linearly with population size (is close to constant in ‘population time’).
- The rate of institutional progress, on the other hand, is probably much more limited by ‘clock time’.
  - So that this progress rate scales very sublinearly with population size.
- Therefore increasing instantaneous population size speeds up scientific progress relative to institutional progress.
- Therefore increasing instantaneous population size tends to increase extinction risk (as far as this factor alone is concerned).

Conclusions

- From Part I:
  - Extinction risk is extremely serious according to a “totalist” population axiology, but not very serious according to a “person-affecting” population axiology.
  - But there’s no coherent person-affecting axiology. Meanwhile, the totalist axiology is highly plausible.
  - Therefore it’s highly plausible that extinction risk is extremely serious.
- From Part II:
  - Many extant popular arguments for population reduction presuppose averagism, which is implausible.
  - From the totalist point of view, the key consideration may well be whether population reduction would increase or decrease extinction risk.
  - There’s a differential-development-based reason for thinking that population reduction would decrease extinction risk, and therefore for favouring population reduction (after all).
    - (Although this reason is not on its own conclusive.)