Population Ethics

Johan E. Gustafsson

Populations as Boxes

The size of the population

People's well-being level
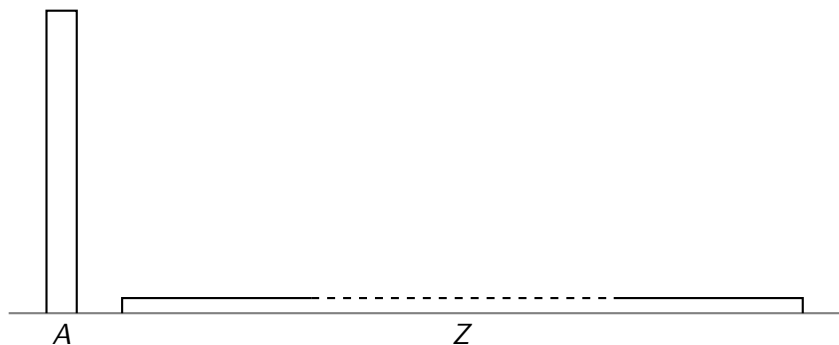
The zero level of well-being
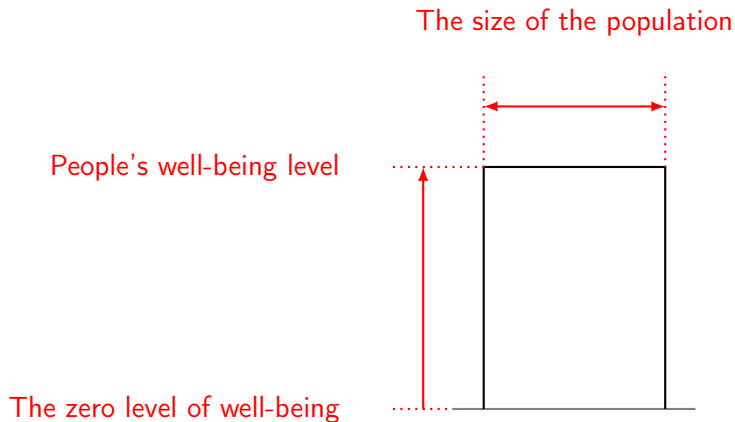
Total Utilitarianism

A first population is at least as good as a second population if and only if the sum total of well-being is at least as great in the first as in the second.

Derek Parfit (1984, p. 388)

The Repugnant Conclusion

For any possible population of at least ten billion people, all with 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 barely worth living.
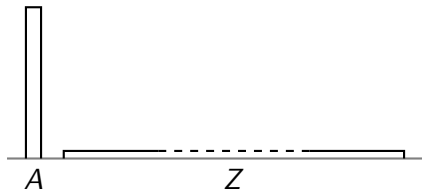
The Repugnant Conclusion

For every population like A, there is a better population like Z.

Average Utilitarianism

A first population is at least as good as a second population if and only if the average well-being is at least as great in the first as in the second.

Average utilitarianism avoids the repugnant conclusion, since it yields that *A* is better than *Z*.
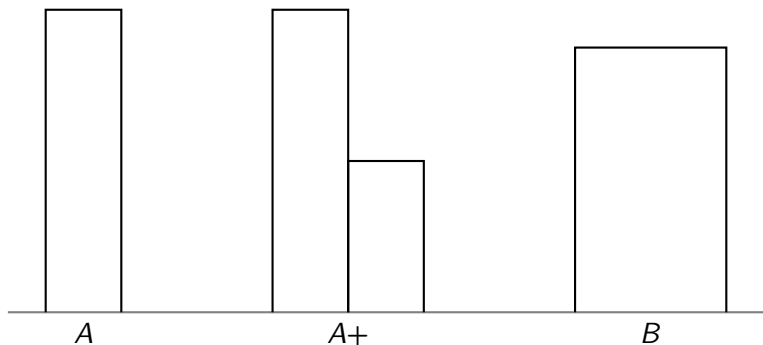

But average utilitarianism has other problems.

It yields that a large population of people with very high well-being is worse than a population with only one person who has just slightly higher well-being than the average in the first population.
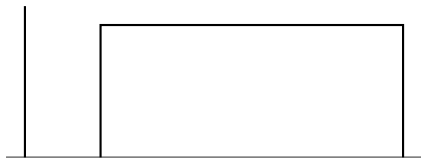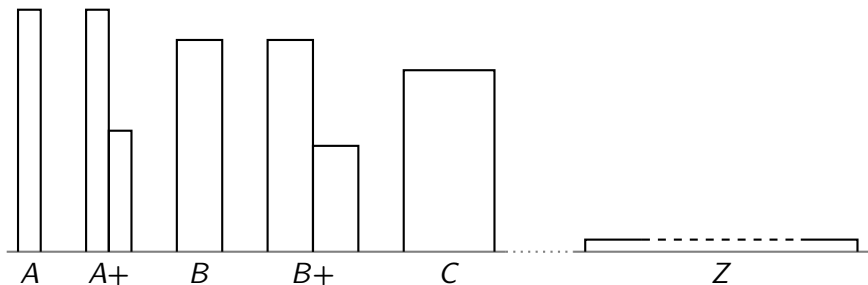
The Mere-Addition Paradox

Intuitively, it seems that

$$A \precsim A+ \prec B.$$

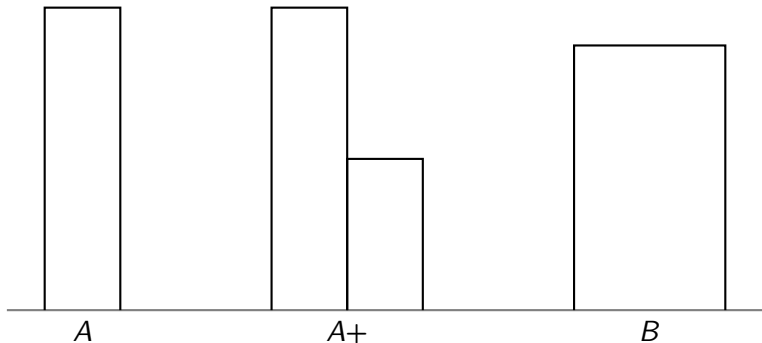Then, by transitivity, we have

$$A \prec B.$$

We can iterate the reasoning in the mere-addition paradox:

$$A \precsim A+ \prec B \precsim B+ \prec C \ldots \prec Z.$$

And then, by transitivity, we get the repugnant conclusion,

$$A \prec Z.$$

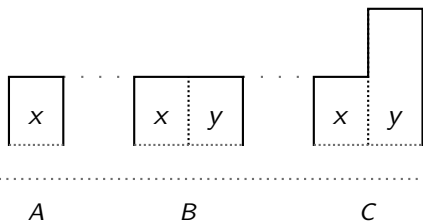Perhaps we should reject the first step in the paradox, that is, reject that $A+$ is at least as good as A.

The Intuition of Neutrality

Jan Narveson (1973, p. 80)

> *We are in favor of making people happy, but neutral about making happy people.*

While this slogan sounds intuitive, it is not obvious what is meant by neutrality.
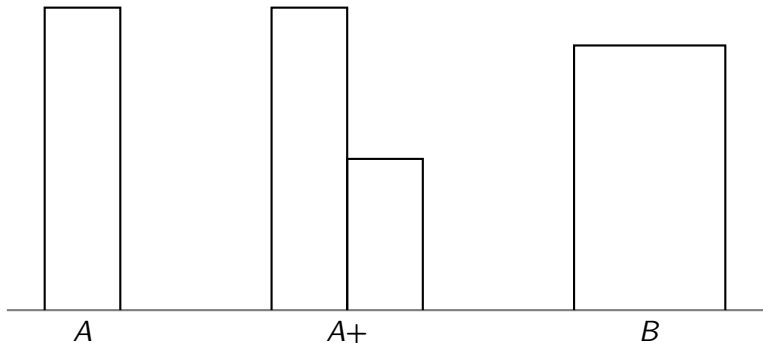
Neutrality as equal goodness

the neutrality range

If neutrality is equal goodness:

$A \sim B$ and $A \sim C$.

From Pareto we get:

$C \succ B$.

This violates transitivity:

$$A \sim B$$
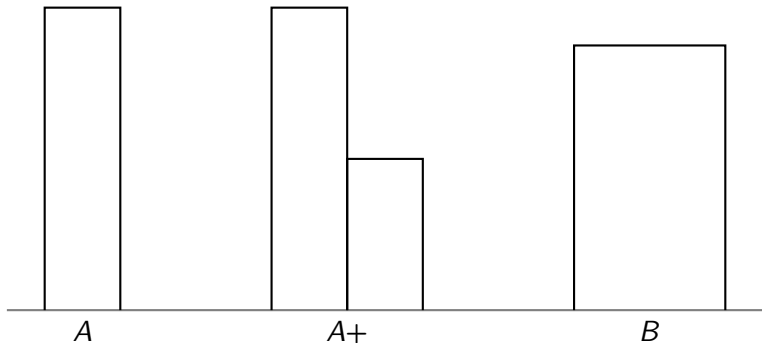$$\text{\textbackslash} \quad C$$

Furthermore, if neutrality is equal goodness, then the neutrality intuition still implies that $A+$ is at least as good as A.

Neutrality as Incommensurability

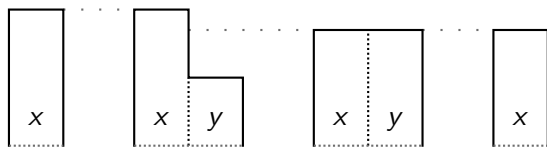the neutrality range
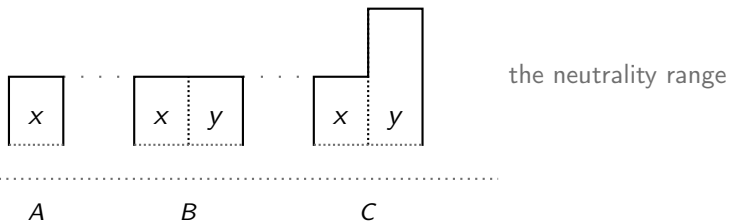
$$A \quad \# \quad B$$
$$\# \quad \prec$$
$$C$$

Interpreting neutrality as incommensurability avoids the difficulty with transitivity, because incommensurability need not be transitive.

If neutrality is incommensurability, then the neutrality intuition rules out that A is at least as good as $A+$.

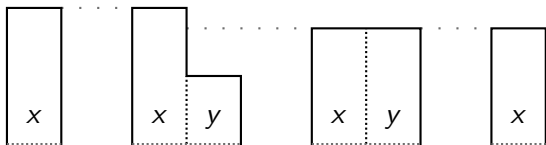# The Objection from Greediness

the neutrality range

If neutrality is incommensurability: $A \# B$.

Since C is more equal and has more well-being in total than B, we should accept: $C \succ B$.

From the transitivity of better, we have $A \not\succ C$, that is, not $A \succ C$.

(Because if $A \succ C$ and $C \succ B$, then $A \succ B$. But we have $A \# B$.)

So a move from A to C is not bad.

the neutrality range

A   B   C   D

From Pareto: $A \succ D$.
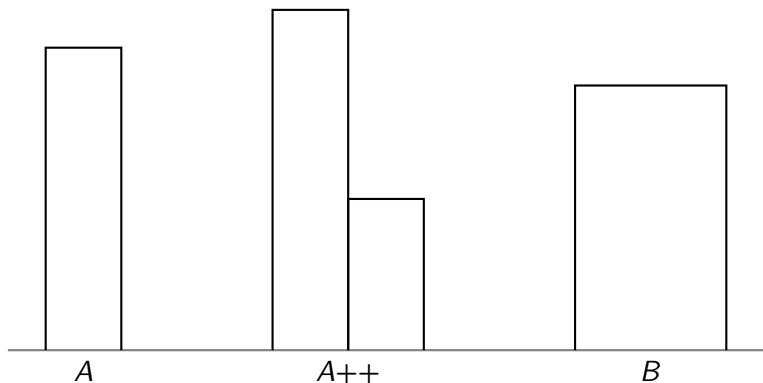
So a move from A to D is bad,

From the intuition of neutrality: a move from D to C is neutral.

Yet we had earlier (last slide): a move from A to C is not bad.

John Broome (2004, p. 170)

> *The net effect of one bad thing and one neutral thing should be bad. But according to our theory it is not bad; it is neutral.*

The Mere-Addition Paradox: Second Version

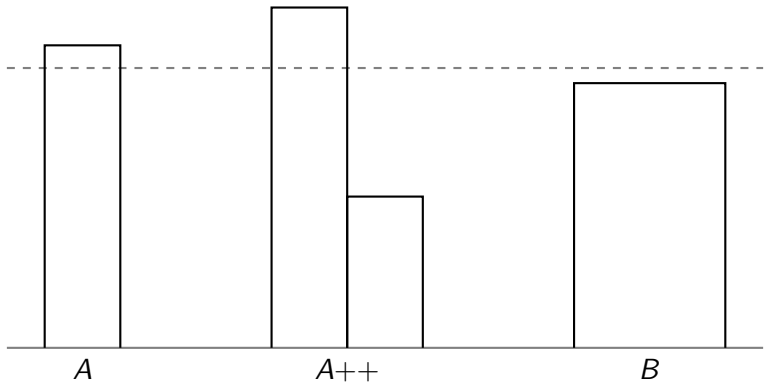Intuitively, it seems that

$$A \precsim A{+}{+} \prec B.$$

Then, by transitivity, we then have

$$A \prec B.$$

Critical-Level Utilitarianism

On critical-utilitarianism, the contributive value of a person p is p's well-being minus a positive critical level.

The value of a population on critical-level utilitarianism is then the sum of the contributive values for all persons in the population.
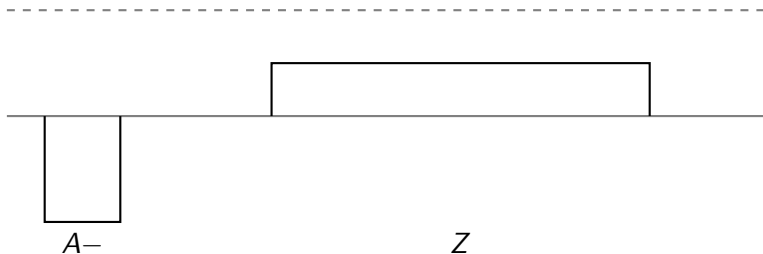
Critical-level utilitarianism lets us deny that $A++$ at least as good as A.

But while critical-level utilitarianism with a sufficiently high critical level avoids the repugnant conclusion, it has other problems.

Arrhenius (2000, p. 256) points out that it implies

The Strong Sadistic Conclusion
For any population of lives with negative well-being, there is a population of lives with positive well-being that is worse, other things being equal.

Critical-level utilitarianism yields that $A-$ is better than Z.

The Person-Affecting Restriction

If an outcome A is better than B, then A is better than B for at least one individual.

The Existential Question

Can it be better or worse for a person to be than not to be, that is, can it be better or worse to exist than to not exist at all?

The person-affecting restriction combined with a negative answer to the existential question yields some strange results.
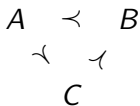
future bliss
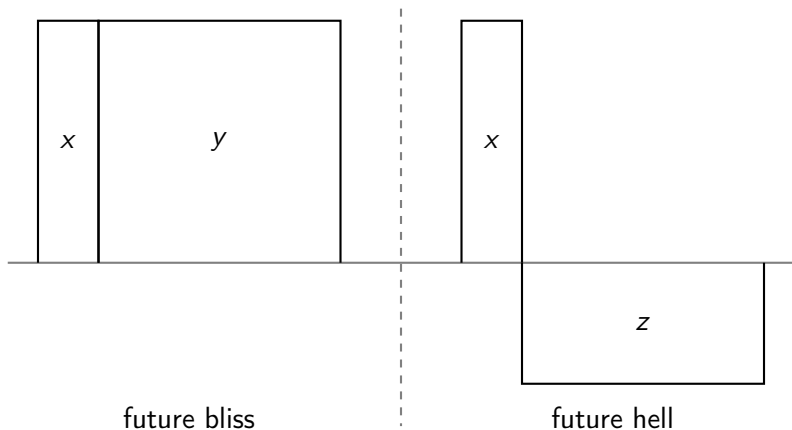
future hell

Surely, future bliss is better than future hell.

But, given the person-affecting view and a negative answer to the existential question, future bliss is not better than future hell.

Moreover, if we also hold that a value comparison between two populations only should only depend on how people are affected between them, we seem to get cyclic value orderings.

$$
\begin{array}{c c c c}
 & P_1 & P_2 & P_3 \\
A & 1 & - & 2 \\
B & 2 & 1 & - \\
C & - & 2 & 1 \\
\end{array}
$$

$$
\begin{array}{c c c}
A & \prec & B \\
 & \prec \quad \prec & \\
 & C & \\
\end{array}
$$

References

Arrhenius, Gustaf (2000) 'An Impossibility Theorem for Welfarist Axiologies', *Economics and Philosophy* 16 (2): 247–266.

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

Narveson, Jan (1973) 'Moral Problems of Population', *Monist* 57 (1): 62–86.

Parfit, Derek (1984) *Reasons and Persons*, Oxford: Clarendon Press.