Population axiology

Hilary Greaves

This is the pre-peer reviewed version of this article. The final version is forthcoming in Philosophy Compass*; please cite the published version.*

Abstract

Population axiology is the study of the conditions under which one state of affairs is better than another, when the states of affairs in question may differ over the numbers and the identities of the persons who ever live. Extant theories include totalism, averagism, variable value theories, critical level theories, and "person-affecting" theories. Each of these theories is open to objections that are at least prima facie serious. A series of impossibility theorems shows that this is no coincidence: it can be proved, for various sets of prima facie intuitively compelling desiderata, that no axiology can simultaneously satisfy all the desiderata on the list. One's choice of population axiology appears to be a choice of which intuition one is least unwilling to give up.

1 Population ethics and population axiology: The basic questions

In many decision situations, at least in expectation, an agent's decision has no effect on the numbers and identities of persons born. For those situations, *fixed-population ethics* is adequate. But in many other decision situations, this condition does not hold. Should one have an additional child? How should life-saving resources be prioritised between the young (who might go on to have children) and the old (who are past reproductive age)? How much should one do to prevent climate change from reducing the number of persons the Earth is able to sustain in the future? Should one fund condom distribution in the developing world? In all these cases, one's actions can affect both *who* is born and *how many* people are (ever) born. To deal with cases of this nature, we need *variable-population ethics*: 'population ethics' for short.

In purely general terms, part of any plausible theory of ethics, fixed- or variable-population, is its axiology: that is, its ordering of states of affairs in terms of better and worse overall. Importantly, this applies also to plausible non-consequentialist theories. Non-consequentialists may well admit side-constraints

(Nozick, 1974) and/or agent-centred prerogatives (Scheffler, 1982) to limit the extent to which they counsel being guided by considerations of the good, but the limitations cannot plausibly be so severe as to leave no room for considerations of overall good.[1] So also for the case of population ethics: part of any plausible theory of population ethics is its population axiology. A population axiology is a betterness ordering of states of affairs, where the states of affairs include ones in which different numbers of persons are ever born.[2]

In addition to this concept of betterness-overall, we also have a concept of one state of afairs being better or worse *for a particular individual* than another state of affairs. A person's *well-being level* in a given state of affairs relates to this latter concept: it is an index of how good or bad that state of affairs is *for that person*. It is usually assumed that the scale of possible well-being levels has at least cardinal structure (that is, that ratios of well-being differences are well-defined; as we will see below, some population axiologies also require there to be additional structure beyond this).

Virtually all of the literature on population axiology represents states of affairs via a specification of how many persons (and, if and where required: *which* persons) exist, together with an assignment of lifetime well-being levels to those persons. This is not to deny that values other than well-being – desert, shared culture, and so on – might have an importance that is independent of their contributions to well-being levels; it is only to assume that all such other values can sensibly be held fixed, so that we can compare states of affairs on the basis of the specified well-being profiles under the stipulation that there are no relevant differences in the extent to which people get what they deserve (etc.) in the states of affairs in question.

Fixed-population axiology underdetermines population axiology. To take the simplest example: even if one has decided that an ordering of states of affairs in terms of aggregate well-being (i.e. 'utilitarianism', in a restricted sense of that term) is the correct axiology for fixed-population cases, this does not settle one's view on whether the betterness ordering in a *variable-population* context is given by total well-being, average well-being, some kind of compromise between the two (as in some of the 'Variable Value theories' discussed below), or anything else that reduces to the total well-being ordering in fixed-population cases. (To put this another way: a fixed-population axiology is a non-trivial equivalence class of variable-population axiologies.) Population axiology therefore opens up new questions. These new questions turn out to be surprisingly difficult.

The structure of this survey article is as follows. Sections 2–5 discuss various concrete proposals for a population axiology: Totalist and Averagist theories, 'Variable Value' theories, 'Critical Level' theories, and various proposals based in one way or another on 'person-affecting' ideology. In each case, we will see that the proposed axiology is open to some serious objection. Section 6 explains

[1] For some dissent from this widely accepted view, see (Taurek, 1977) and (Foot, 1985).

[2] Since population axiology is concerned with the numbers and identities of persons who are ever born, not merely those who exist at some specified time, the 'states of affairs' in question are 'timeless' (i.e., they include covering the whole history of the universe from remote past to remote future), rather than instantaneous.

that this is no accident: here I outline 'impossibility theorems' that show that, in the domain of population axiology (and unlike the fixed-population case), *any* proposed axiology will violate one or more of a number of initially very compelling intuitive constraints. I briefly discuss what the appropriate response to this situation might be.

Throughout, I will assume that the correct axiology for the fixed-population case is a utilitarian one (corresponding to summing or, equivalently in the fixed-population case, averaging well-being across persons). This is of course controversial: in particular, many authors defend instead a prioritarian or egalitarian axiology for the fixed-population case (for discussion, see e.g. (Broome, 1991; L. Temkin, 1993; Parfit, 1997)). The assumption, however, is made only for simplicity: one could easily formulate prioritarian and/or egalitarian versions of the population axiologies discussed below, and much of the discussion would remain essentially unchanged.[3]

2 Totalism and Averagism

We start with the two most obvious population axiologies:

Totalism: A is better than B iff total well-being in A is higher than total well-being in B. If A and B have equal total well-being, then A and B are equally good.

Averagism: A is better than B iff average well-being in A is higher than average well-being in B. If A and B have equal average well-being, then A and B are equally good.

While these two axiologies themselves are straightforward enough to be easily understood from their verbal descriptions, it will be helpful, going forward, to have a uniform more formal notation. To this end, for an arbitrary state of affairs X, let $|X|$ be the number of people in X, and let $\overline{X}$ be the average well-being level in X. In this notation, Totalism and Averagism are represented respectively by the value functions

$$V_{Tot}(X) = |X|\overline{X}; \tag{1}$$

$$V_{Av}(X) = \overline{X}. \tag{2}$$

Note that Totalism is well-defined only relative to a choice of zero point on the well-being scale; this zero point is generally taken to be the threshold at which life becomes 'worth living'.[4]

[3] Non-utilitarian axiologies have been explored in the variable-population context by e.g. (Brown, 2007; Holtug, 2007; Adler, 2009; Fleurbaey & Voorhoeve, 2016; Arrhenius, n.d., chapter 7).

[4] There is some subtlety about precisely how this notion of 'life worth living' is to be understood, if it is not sufficiently intuitive to stand without definition. For discussion, see (Broome, 2004, pp.66-8) and (Arrhenius, n.d., chapter 2).

Each of these population axiologies is, on further reflection, open to serious objection. The key objection to Totalism is that, at least if the well-being scale has a structure like that of the real numbers and provided the set of possible states of affairs is sufficiently rich, Totalism entails

The Repugnant Conclusion: For any state of affairs A, there is a better state of affairs Z in which no-one has a life that is more than barely worth living.

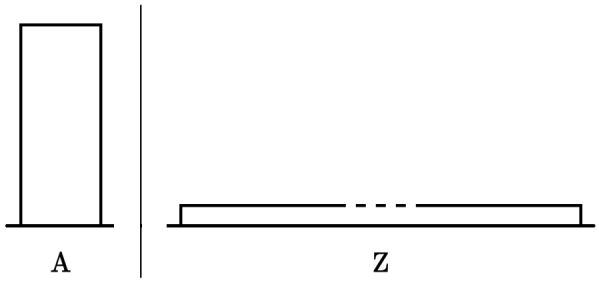

Figure 1: The Repugnant Conclusion.

If Totalism is rejected on the basis of the Repugnant Conclusion, it might be tempting to embrace Averagism. While Averagism avoids the Repugnant Conclusion, however, it faces other problems that are at least as serious. For example, since adding a large number of people with positive but below-average well-being can reduce the average more than adding a smaller number of people with negative well-being would, Averagism entails

The Sadistic Conclusion: A state of affairs resulting from adding persons with negative well-being is sometimes better than one resulting from adding persons with positive well-being, from the same starting point.

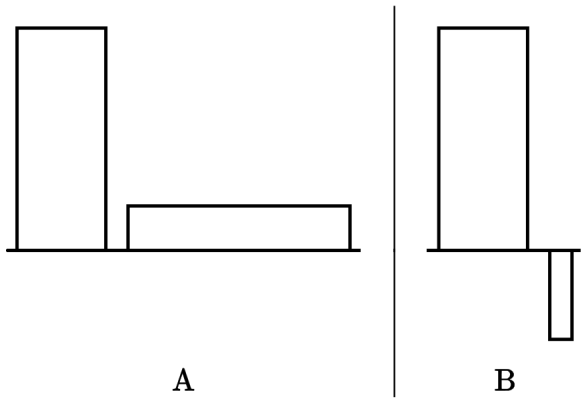

Figure 2: Averagism entails the Sadistic Conclusion: in the example illustrated, A has lower average well-being than B.

In addition (and relatedly), Averagism violates

The Mere Addition Principle: Let A be any state of affairs. Let B be a state of affairs that is just like A except that, in addition, some extra people with lives worth living exist in B who do not exist in A. Then B is not worse than A.

In saying that B is 'just like A', we mean here that everyone who exists in A also exists in B, and has the same well-being level in B as in A (see figure 3).

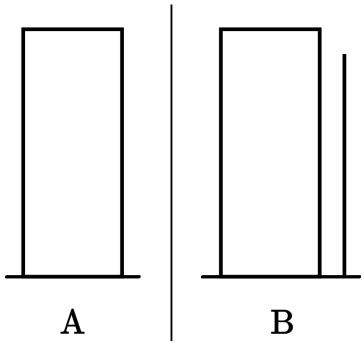

Figure 3: A 'mere addition'. In the case depicted here, B has lower average well-being than A.

Many are willing, on reflection, to reject the Mere Addition Principle. But this is at least *prima facie* puzzling: how can B be not only *not better* than A, but actually *worse* than A, if the change from A to B is only a matter of adding some extra people whose lives are worth living (and who, therefore, are

if anything glad to be alive), and who are not hurting anyone?[5]

3 Variable Value views

In response to the problems faced by Totalism and Averagism, some theorists have investigated axiologies according to which the value (positive or negative) of adding an extra person with a fixed well-being level has a diminishing marginal value, i.e. is of smaller magnitude when the number of pre-existing persons is larger. Call these *Variable Value* theories.

In particular, Variable Value theories are sometimes designed to be compromises between Totalism and Averagism. Arguably, Averagism 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. We might therefore seek a theory that approximates Totalism at small population sizes, but approximates Averagism at large population sizes (and, in particular, thereby avoids the Repugnant Conclusion): for example,

Variable Value theory: The value of a state of affairs X is given by

$$V_{VV}(X) = \overline{X}g(|X|), \tag{3}$$

where g is a strictly increasing and strictly concave function with a horizontal asymptote (T. Hurka, 1983; Y.-k. Ng, 1989).

This view (and its value function) requires some explanation. The qualitative shape of the function g is as depicted in figure 4:

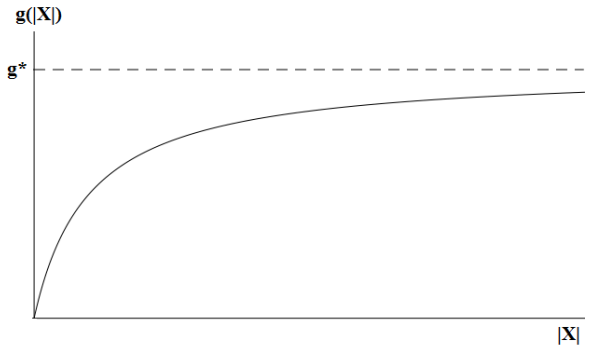

Figure 4: A 'variable value' transform, with horizontal asymptote g^*.

This view then agrees with Totalism that when population size is small, 'mere additions' tend to amount to improvements. Indeed, for comparisons between

[5]For other objections to Averagism, see (T. M. Hurka, 1982a, 1982b; Parfit, 1984, section 143).

small populations, the Variable Value theory represented by (3) approximates Totalism more generally: at sufficiently small population sizes, the function $g(|X|)$ is well approximated by a straight line through the origin ($g(|X|) \simeq \beta|X|$, for some positive constant $\beta \in \Re^+$), and so (3) is approximately equal to the Totalist value function (up to a uniform positive multiplicative constant). However, thanks to the concave shape of the transform g, the same is not true at large population sizes. When the population size becomes sufficiently large, the curve $g(|X|)$ instead approximates the horizontal dotted line (the curve's asymptote), and gets closer and closer to this line the larger the population size becomes: thus, for the purpose of comparisons between states of affairs both of whose timeless populations are large, this Variable Value theory roughly approximates Averagism. In particular, average-reducing 'mere additions' tend to count as disimprovements when the population is already very large.

This Variable Value theory does not entail the Repugnant Conclusion. This is because according to (3), the value of a state of affairs in which no-one has a well-being level higher than the 'barely worth living' level ϵ — no matter how large the population is — can never be more than the *finite* value $g^*\epsilon$. Therefore, provided the 'A' population is such that $\overline{A}g(|A|)$ is greater than the fixed number $g * \epsilon$ — a condition that is easily met by states of affairs with *high* average well-being levels — Z can never be better than A, no matter how large the 'Z'-population.

The Variable Value theory represented by (3), however, both violates the Mere Addition principle and entails the Sadistic Conclusion. This is unsurprising: that view was precisely designed (in part) to mimic Averagism at some population sizes, and as we noted above, Averagism has both these properties.

One might suspect that *any* theory that mimics Totalism at small population sizes, but employs the notion of 'diminishing marginal value' in such a way as to avoid the Repugnant Conclusion, will violate the Mere Addition principle and entail the Sadistic Conclusion. This would of course be true *if* any such theory had to 'mimic Averagism' *in every sense* at large population sizes. But in fact the suspicion turns out to be incorrect, as is shown by the following 'Geometrist' axiology (Sider, 1991):

Geometrism: The value of a state of affairs X is given by

$$V_{Geo}(X) = \sum_{i=1}^{n} \frac{w_i(X)}{r^{i-1}} + \sum_{j=1}^{m} \frac{w_j(X)}{r^{j-1}}, \tag{4}$$

where:

- the index i in the first summation ranges over persons whose well-being level in X is positive, ordered from best off to worst off;
- the index j in the second sum ranges over persons whose well-being level in X is negative, ordered from worst off to best off;
- r is a constant that is greater than 1.

Provided the constant r is chosen to be close to 1, this theory mimics Totalism at small population sizes (because with r close to 1, if both n and m are small, neither $\frac{1}{r^{i-1}}$ with $i < n$ nor $\frac{1}{r^{j-1}}$ with $j < m$ can ever be much different from 1). However, Geometrism also deviates from Totalism at large populations, encoding the idea of 'diminishing marginal value' in such a way as to avoid the Repugnant Conclusion. According to Geometrism, the value of a state of affairs in which everyone's well-being level is ϵ can never, no matter how large the population, be greater than the sum $\sum_{i=1}^{n} \frac{\epsilon}{r^{i-1}}$; but this sum, although it contains an infinite number of terms, has the *finite* value $\epsilon \frac{r}{r-1}$.

The theoretical possibility of Geometrism illustrates that it is *possible* to both satisfy the Mere Addition principle and avoid the Repugnant Conclusion, but not that it is possible to do so *in a desirable way*. Indeed, this view (as its creator was well aware[6]) is profoundly anti-egalitarian. This is easy to see by considering the theory's treatment of persons with positive well-being, even in the fixed-population context. If Jim has well-being level 100 and Bob has well-being level 20, then Jim's well-being will be weighted more heavily than Bob's in the geometric sum (4) (recall that in that sum, persons with positive well-being are ordered from best off to worst off, and that their well-being is then weighted by a factor $\frac{1}{r^{i-1}}$ that is greater for persons whose well-being appears earlier in the sum). It follows that removing some fixed amount of well-being from the better-off Jim and transferring it to the worse-off Bob – for example, transferring 40 units of well-being from Jim to Bob, so that they both end up at 60 — amounts to making things worse, according to this theory. Further, since any such 'equalising' transfer will count as making things worse *by some finite amount*, the result will still count as worse even after we subsequently increase the well-being levels of all persons by a sufficiently small amount (if, perhaps, Jim and Bob both end up at 65 rather than 60). That is, Geometrism violates the following fixed-population-size condition:

Non-Anti-Egalitarianism: Let A, B be states of affairs such that the following three conditions all hold: A and B contain the same number of individuals; all individuals in B have equal well-being; B has higher total (and hence average) well-being than A. Then, B is better than A.

But violation of Non-Anti-Egalitarianism seems unacceptable. While some axiologists (namely utilitarians) are willing to accept that equality of well-being has no intrinsic value, so that inequality-increasing transfers of well-being are matters of indifference, nobody normally thinks that such transfers are positively *good* (as they would have to be if they could sometimes offset or outweigh decreases in total well-being).

[6] Sider outlines Geometrism not as a serious proposal for the correct population axiology, but merely by means of counterexample to the claim that *every* variable value theory will violate the Mere Addition principle.

4 Critical Level theories

The *Critical Level* family of population axiologies (Blackorby & Donaldson, 1984; Blackorby, Bossert, & Donaldson, 1995; Broome, 2004; Blackorby, Bossert, & Donaldson, 2005) consists of variants on Totalism. Like Totalism (and unlike Averagism and Variable Value theories), any Critical Level theory holds that the 'contributive value'[7] of an extra person depends only on that additional person's well-being level, and not on either the number of pre-existing persons nor those pre-existing persons' well-being levels. But while Totalism holds that an additional person's contributive value is simply *equal* to her well-being level, Critical Level theories hold that the contributive value of adding an extra person is equal to that extra person's well-being level *minus* some constant, α:

$$V_{CL}(X) = |X| \left(\overline{X} - \alpha\right). \tag{5}$$

Defenders of Critical Level theories propose that the constant α should be positive ($\alpha > 0$). Heuristically, then, on this view, an additional person increases the value of the world provided her well-being is above the 'critical' threshold α; such an addition decreases the value of the world not only if the extra person's well-being level w is negative, but also if it lies in the range $0 \leq w < \alpha$. A consequence of this is that these Critical Level theories do not entail the Repugnant Conclusion, at least as stated above: if ϵ (the level of a life 'barely worth living') is less than the critical level α, then an increasingly large 'Z-population' of persons with well-being ϵ, on this view, has an overall value that is increasingly *negative,* and certainly less than that of the A-population (which has positive value according to (5) provided only that the average well-being level in A is above the critical level).

Critical Level theories, however, nonetheless entail a *variant* of the Repugnant Conclusion: for any state of affairs A, no matter how good, there exists some better state of affairs Z (involving a very large population) in which no-one has a well-being level that is more than barely above the critical level. Call this the 'Weak Repugnant Conclusion' (figure 5).

How bad this is of course depends on how high the critical level is. If the critical level is sufficiently high that any life with a well-being above α is not just 'barely worth living', but fairly good, then it may not be at all counterintuitive to hold that for any population A and any positive amount of well-being $\epsilon > 0$, there is a sufficiently large population Z′ that is better than A and in which no-one has a well-being level that is higher than $\alpha + \epsilon$.

[7] That is: the amount by which the extra person's existence increases the overall value of the world, relative to a state of affairs in which she does not exist but all other persons' well-being levels are equal.

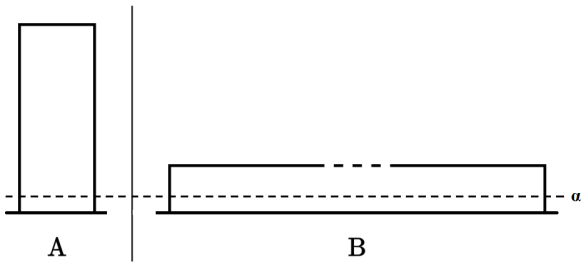

Figure 5: Critical Level theories entail the Weak Repugnant Conclusion.

However, this way of rendering the Weak Repugnant Conclusion intuitively acceptable comes at a price. Note first that any Critical Level theory with a positive critical level both violates the Mere Addition principle and entails the Sadistic Conclusion (figure 6). Secondly, the *way* in which the Critical Level theory violates the Mere Addition principle and entails the Sadistic Conclusion gets intuitively worse, the higher the critical level is. Advocates of Critical Level view therefore face a dilemma: If the critical level is too low then the view entails a worryingly close analogue of the Repugnant Conclusion, but if the critical level is too high then the view instead both entails a particularly worrying version of the Sadistic Conclusion ((Broome, 2004, pp.213-4); (Arrhenius, n.d., section 5.1)), and violates the Mere Addition principle in a particularly egregious way. (See also (Vallentyne, 2009; Broome, 2009).)

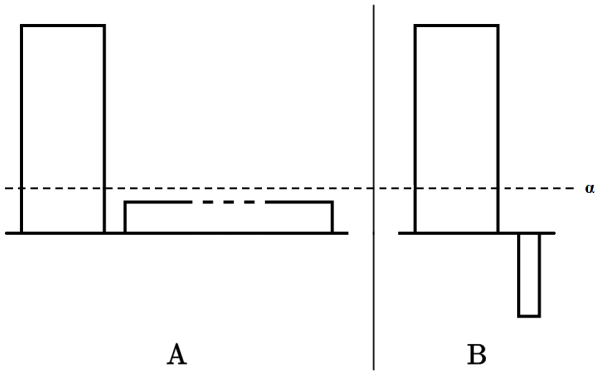

Figure 6: How a Critical Level theory entails the Sadistic Conclusion: with the critical level α as drawn, and with the right-hand subpopulation in A sufficiently large, according to Critical Level theory A is worse than B.

5 Person-affecting views

5.1 Neutrality

Totalism holds that adding an extra person with positive well-being makes a state of affairs better, other things being equal. This is arguably quite intuitive, for reasons similar to those behind the Mere Addition Principle: lives worth living are good things, and, one might think, more of a good thing is in general better (and not merely, as the Mere Addition Principle conservatively holds, not worse). Many people's intuitions, however, strenuously resist this suggestion. The worry is that if a mere addition makes a state of affairs better, it presumably follows that we have at least some moral reason to create extra people, provided that those extra people would have lives worth living. This implication is fairly weak: the moral reason that we have may be one that is easily overridden by other reasons of moral or non-moral type (in particular, there are several reasons why it need not follow that we have *obligations* to have extra children, not least that maximising consequentialism may not be the right account of the connection between axiology and moral obligations). However, to many people, even this weak implication is unacceptable: these people think that there are *no* moral reasons, stemming from the well-being that the additional persons would enjoy if created, to create additional persons. Thus, for example, Narveson writes 'We are in favour of making people happy, but neutral about making happy people' (1973, p.80).

These thoughts are captured in the following

Neutrality Principle: Adding an extra person to the world, if it is done in such a way as to leave the well-being levels of others unaffected, does not make a state of affairs either better or worse.[8]

Note that none of the 'impersonal' population axiologies surveyed in sections 2–4 in general respects the Neutrality Principle. (Totalism and Averagism, for instance, hold that adding an extra person is neutral only in the special case in which that extra person has (respectively) zero well-being, or a well-being level that is equal to the pre-existing average.) This suggests the project of formulating a new 'person-affecting' population axiology that will recover the Neutrality Principle. As we will now illustrate, however, it turns out to be remarkably difficult to formulate any remotely acceptable axiology that captures this idea of 'neutrality'.[9]

[8] A weaker version of this principle would assert 'neutrality' only in cases in which the added person has zero or positive well-being: even those otherwise attracted to the idea of neutrality may well feel that adding a person with a life that is strictly below the threshold for being 'worth living' — a life of uncompensated pain and suffering, say — makes a state of affairs worse. (This difference between positive and negative well-being is often termed 'The Asymmetry'.)

[9] The issues covered in this section are discussed at greater length by Broome (2004, chapter 10).

5.2 The 'Principle of equal existence'

If adding an extra person makes a state of affairs neither better nor worse, perhaps it results in a state of affairs that is *equally as good as* the original state of affairs. That is, one might try to capture the intuition of neutrality via the following principle:

The Principle of Equal Existence: Let A be any state of affairs. Let B be a state of affairs that is just like A, except that an additional person exists who does not exist in A. (In particular, all the people who exist in A also exist in B, and have the same well-being level in B as in A.) Then A and B are equally good.

As Broome (1994, 2004, pp.146-9) points out, however, this principle is all but self-contradictory. This is because there is more than one way of adding an extra person to A — one might add an extra person with well-being level 5, say (leading to state of affairs B_1), or (instead) add the same extra person with well-being level 100 (leading to state of affairs B_2) — and these ways are not all equally as good as one another. In our example, B_2 is clearly better than B_1; but the Principle of Equal Existence would require that B_1 and A are equally good, and that A and B_2 are equally good, in which case (by transitivity of 'equally as good as') B_1 and B_2 would have to be equally as good as one another. The Principle of Equal Existence therefore cannot be correct.

5.3 Non-impartial theories

A second type of attempt to capture the neutrality/person-affecting intuition is explored in detail by Arrhenius (Arrhenius, n.d., chapter 10). According to these 'non-impartial' theories, when comparing two states of affairs A and B in terms of overall goodness, *some* of the possible persons who exist in A and/or in B matter — that is, their well-being is relevant to assessing the goodness of A and B — whereas other such possible persons do not, in this sense, matter.

Within this broad framework, non-impartial theories can be further characterised according to *how* they identify the privileged subset of possible persons. *Presentism* holds that the only persons who matter are persons who presently exist (and, in particular, not those who might or will exist in the future). *Actualism* holds that the only persons who matter are actual persons (and not merely possible persons) Necessitarianism holds that the only persons who matter, in a situation of deciding between A and B, are those who exist both A and in B (i.e., those who, for the purposes of the present decision situation, exist 'necessarily' — who exist regardless of how the decision is settled — and not those whose very existence is contingent on one's current decision).[10]

[10] Although presentist remarks are sometimes made in relatively careless context, this particular form of non-impartial theory appears to have few, if any, serious defenders. (For example, (Heyd, 1988, pp.158, 161) makes some presentist-sounding remarks, but elsewhere in the same article ostensibly the same ideas take on a more necessitarian form.) Necessitarianism is discussed with sympathy, though not strongly advocated, by Singer (Singer, 2011,

Any of these theories, if otherwise satisfactory, would easily be able to capture the intuition that one has no moral reason to create an additional child, via the Neutrality Principle. The hypothetical child does not, at the time of deciding whether or not to create her, *presently* exist; provided one in fact decides not to create her, she does not *actually* exist, even in the future; her existence depends on one's decision, so she is not, in the decision context under consideration, a *necessary* person.

Such theories, however, are on further reflection deeply implausible. A preliminary observation is that it is unclear whether Presentism or Actualism (at any rate) would even help to avoid the problems that we have observed above for 'impersonal' theories. One can, for example, formulate versions of the Repugnant and Sadistic Conclusions using only present persons, or using only actual persons.

In addition, the failure of these theories to consider the well-being of certain possible persons leads to further trouble. Schematically, each theory postulates an 'in-group' (whose interests are taken to matter) and an 'out-group' (whose interests the theory ignores); we can then formulate cases in which each such theory entails wildly implausible implications, in cases in which there is a lot at stake for the 'out-group'. In the examples that follow, for definiteness, we will assume that each of the theories under consideration takes the value of a state of affairs to be equal to the *total* well-being of the relevant 'in-group' (as opposed to the priority-weighted total, or the value function of any other fixed-population axiology), but similar problems would arise on other versions of a presentist/actualist/necessitarian theory.

Presentism, for example, has obviously unacceptable implications for the following 'future bliss or hell' case:

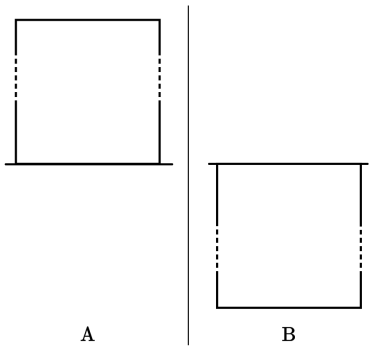

Figure 7: 'Future bliss or hell.'

pp.88–90) (who calls it "the prior existence view"); it is explicitly advocated by Heyd (1992, p. 99). Actualism is perhaps the most popular approach in the 'non-impartial' space (see, in particular, (Parsons, 2002; Cohen, n.d.); it is criticised by Hare (2007). For a more detailed survey of all three types of theory, including further references, see (Arrhenius, n.d., chapter 10).

Here, A and B are equal-sized populations; everyone in A has a blissful life, while everyone in B has a hellish life; but all the people both in A and in B are *future* people. Since presentism accords no importance to the well-being of future persons, that theory implies that A and B are equally good. But that is obviously absurd: clearly, A is better than B.

Actualism has similar implications for a variant of the 'future bliss or hell' case in which the intuitively worse state of affairs is one in which people have lives that are worth living, but much less good than the lives in the intuitively better state of affairs — call this the 'future bliss or mediocrity' case:

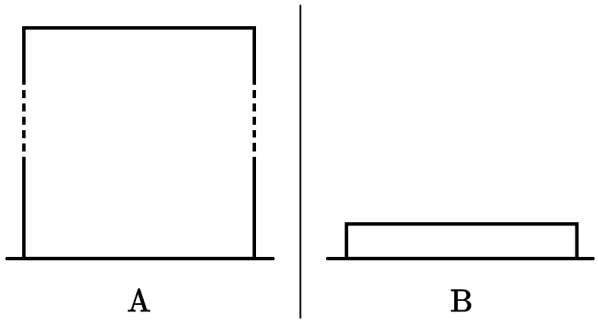

Figure 8: 'Future bliss or mediocrity.'

Suppose, similarly to the 'future bliss or hell' case, that the populations in A and B are disjoint. In this scenario, *if A is actual* then the actualist theory will hold (correctly) that A is better than B. But if B is actual, the theory will hold, implausibly, that B is better than A. This is unacceptable: an adequate theory must capture the *invariant* fact that A is better than B.

To see the problems with necessitarianism, consider the three-option decision scenario in figure 9. Here, person x exists in all three possible outcomes (A, B and C), while person y exists only in A and B:

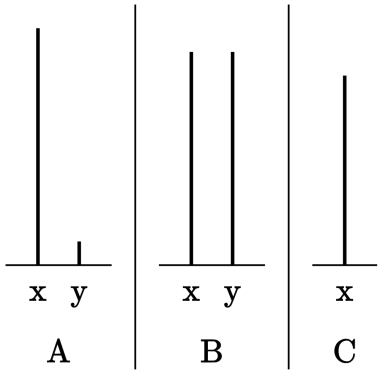

Figure 9: A problem for necessitarianism.

Relative to this decision situation, a necessitarian theory will hold that A is better than B: since y does not exist in C, a necessitarian theory holds that y's well-being is of no moral importance, so that the three options in this decision situation are simply ordered according to the well-being of the single 'necessary person', x. But this is implausible: at least for some variant of this scenario, any plausible theory will hold that A is better than B (and necessitarianism itself would agree *if* A and B were the only two options).

In addition to the problems we have just surveyed — problems arising from the fact that each of these theories ignores the possible interests of some possible persons — actualism and necessitarianism also have structural problems, already visible in the example cases we have considered. Actualism violates a condition of axiological invariance: that the betterness relations between two alternatives should not depend on which alternative is actual (for discussion of this and/or of the related condition of normative invariance, see (Carlson, 1995, pp.100-101), (Broome, 2004, p. 74), (Bykvist, 2007), (Howard-Snyder, 2008)). And necessitarianism is choice-set-dependent: whether or not A is better than B can depend, according to necessitarianism, on which other alternatives are also contained in the choice set (in the above case: whether or not C is included). Each of these structural features is arguably deeply problematic.

5.4 Harm-minimisation theories

A third type of person-affecting theory focusses on the *minimisation of comparative harm.* The basic idea is as follows. Consider any state of affairs A_i, and any person x who exists in A_i. Relative to choice set $\{A_1, \ldots, A_n\}$, say that the amount of (comparative) *harm* that person x occurs in state of affairs A_i is the amount by which x's well-being in A_i falls short of the maximum well-being level that x might have attained:

$$CH\left(x, A_i\right) = \begin{cases} \max_j \left(w_j\left(x\right) - w_i\left(x\right)\right) & \text{if this is positive} \\ 0 & \text{otherwise,} \end{cases}$$

where $w_j(x)$ is x's well-being level in state of affairs A_j. According to the harm-minimisation theory, the best state of affairs is then the one in which there is the lowest total comparative harm, summed over all persons who exist in the state of affairs in question:

$V\left(A_i\right) = \sum_{\{x: x \text{ exists in } A_i\}} CH\left(x, A_i\right)$.

Theories along these lines are discussed by Roberts (2003, 2011), Meacham (2012), Temkin (2012) and Ross (2015).

This type of theory equally succeeds in capturing neutrality as stated above. To see this, note that if we are comparing (only) two states of affairs A_1, A_2 that are identical except that A_2 contains an additional person (with positive well-being level) who does not exist in A_1, then (i) no *other* person suffers any comparative harm in either state of affairs, since other persons' well-being levels are the same in A_1 as they are in A_2. In addition, (ii) the 'added' person suffers no relevant comparative harm in A_1 since she does not exist in A_1, and suffers

no comparative harm in A_2 since she has (if anything) a higher well-being level in A_2 than in A_1.[11]

Harm-minimisation theories are subject, at least prima facie, to three problems. Firstly, like necessitarianism, they are choice-set-dependent: the relative ranking of two states of affairs A_1, A_2 can be altered by the addition or removal of some third alternative from/to the choice set, provided that A_1 and A_2 have non-identical populations. (In the example in the previous paragraph, if a third state of affairs A_3 is added to the choice set, such that the 'added' person has higher well-being in A_3 than in A_2, then *relative to the choice set* $\{A_1, A_2, A_3\}$ the theory judges A_2 worse than A_1.) Secondly, and relatedly, the particular *way* in which they are choice-set-dependent leads to what Ross (2015, section 4) calls the 'problem of improvable life avoidance': it is arguably implausible that we have reasons to avoid bringing additional persons into existence with a given, perhaps very high, well-being level *rather than not bringing them into existence at all*, simply on the grounds that they could have been brought into existence with an even higher well-being level. Thirdly, and most seriously, the theory as stated above yields implausible verdicts in non-identity cases: given any two states of affairs involving disjoint populations (or disjoint subpopulations added to an unaffected pre-existing population), this theory will rank the states of affairs in question as equally good.

The theory developed in (Meacham, 2012) avoids this third problem, via careful specification of the relevant counterpart relations relating persons in different states of affairs. Others (e.g. (L. S. Temkin, 2012, section 12.3) , (Ross, 2015, p.438)) propose that a principle of harm-minimisation forms only part of the correct overall axiology, and that the problem of disjoint populations is to be dealt with by some other part of the overall theory. This latter proposal, however, raises the question of what (if anything) is gained by adding a harm-minimisation element to the 'other part of theory' to which it is coupled: the hybrid theory, for example, may well fail to recover neutrality. However, these matters as this are impossible to assess in abstraction from a detailed specification of the full theory.

5.5 Theories with widespread incomparability

Elsewhere, Narveson comments that 'morality has to do with how we treat whatever people there are'; whether one prefers there to be more or fewer people, he says, is 'a matter of taste' (1973, pp.72-3). This idea is initially supposed to apply to one very particular decision context, viz. that of adding an extra person to an otherwise fixed population. Generalising, however, a fairly natural

[11] The verdicts of the comparative-harm theory in cases in which the added person has a *negative* well-being level depend on whether or not, for this particular purpose, non-existence is treated as akin to having a zero level of well-being. If it is so treated, then a person who exists with negative well-being in state of affairs A suffers a positive amount of comparative harm in A if there is an alternative state of affairs in which she does not exist. Comparative-harm theorists who wish to capture 'the procreative asymmetry' in their axiology take this route. Alternatively, a comparative-harm theorist might uphold neutrality for cases of adding people with negative well-being, as well as for cases of adding people with positive well-being.

development of this idea is that *relations of betterness and worseness hold only in cases of fixed-population comparisons*. That is, on this view, if A and B have non-identical populations, then it is not the case that A is better than B, nor is it the case that B is better than A, and nor is it the case that A and B are equally good: A and B are, instead, *incomparable* in terms of overall goodness. Note that, on the view in question, this incomparability holds not only in cases of disjoint populations, but even if the populations in A and B are *almost* identical: it only requires *one* person to exist in A but not in B, or vice versa, for A and B to be deemed incomparable.

Views along these lines have been defended by David Heyd (1988) and by Ralf Bader (n.d.). The obvious problem with such theories, however, is that in at least some cases of differing populations, one of the states of affairs under consideration is (intuitively) obviously better than the other. The 'future bliss or hell' scenario discussed above is one case in point (a disjoint-population case), but so also is a case in which adding an extra person would transform the lives of all others from lives of bliss to lives of hell (a case of near-identical populations).[12] Theories of widespread incomparability therefore appear to throw out the baby with the bathwater; there may well be *some* incompleteness in the betterness relation, perhaps especially in variable-population cases, but the incompleteness cannot plausibly be as radical as this.

6 Impossibility theorems

In his seminal discussion of population ethics, Parfit (1984) rejected Totalism and Averagism for reasons including those given above, failed to find any alternative axiology that he himself considered satisfactory, but held out hope that this was merely for want of searching hard enough: that, in the future, some fully satisfactory population axiology, called 'Theory X' by way of placeholder, might be found. Much of the subsequent literature has consisted of attempts to formulate such a 'Theory X'. However, as we have seen above, every extant population axiology is open to serious objection: if it does not entail the Repugnant Conclusion then it entails the Sadistic Conclusion, or is anti-egalitarian, or has obviously unacceptable implications concerning future-people cases, or otherwise leads to some similarly serious objection.

If this were where matters rested, one might yet hold out hope that the fully satisfactory 'Theory X' is lurking just around the corner. However, several authors have also formulated *impossibility theorems* for population axiology. These are formal results that purport to show, for various combinations of intuitively compelling desiderata ('avoid the Repugnant Conclusion', 'avoid the Sadistic

[12] On Bader's theory, incomparability holds only when the populations in question fail to be *equinumerous* — that is, when they fail to contain the same number of persons as one another — and not simply when the populations are *nonidentical*. However, this disposes of only a small fraction of the problematic cases, and does not address the fundamental problem: there is still, on Bader's theory, extremely widespread and consequently problematic incomparability. For example, Bader's theory has the same consequence as the more radical incomparability theory for the case of near-identical populations discussed in the main text.

Conclusion', 'respect Non-Anti-Egalitarianism', and so forth), that the desiderata are in fact mutually inconsistent: that is, simply as a matter of logic, *no* population axiology can simultaneously satisfy all of those desiderata (Parfit, 1984; Y.-K. Ng, 1989; Carlson, 1998; Arrhenius, 2000; Kitcher, 2000; Arrhenius, n.d.).[13]

In the light of this, some argue that the Repugnant Conclusion (in particular) may not, at the end of the day, be so clearly unacceptable.[14] According to this line of thought, the Repugnant Conclusion certainly *initially seems* repugnant — that is, it is repugnant *to untutored intuition* — but this 'anti-repugnance' intuition can be debunked as the product of one or more distorting biasses. This line of thought is pressed, in particular, by Torbjörn Tännsjö (2002) and Huemer (2008); see also (Broome, 2004; Pummer, 2013, pp.57-9). Others suggest that one might escape from the impossibility theorems by denying the assumptions that those theorems make concerning the structure of the well-being scale, and/or by appealing to vagueness or incompleteness of the betterness relation (Broome, 2004; Qizilbash, 2005; Rabinowicz, 2009; Parfit, 2016; Thomas, n.d., pp.213-4). Still others argue that the normative force of the impossibility theorems depends on an assumption of choice-set-independence, and then go on argue that the choice-set-dependence exhibited by e.g. necessitarianism and comparative-harm theories may be positively a virtue of those theories, so that we can hope to find an adequate theory in this space (see in particular (Meacham, 2012, section 7); related ideas are pressed by Temkin (1987; 2012)).[15]

One common, but mistaken, response to the impossibility theorems is to turn claim that these theorems are problematic only for consequentialists. It is important to see that this *is* a mistake. For one thing, as noted above, all moral theories need an axiology: while non-consequentialists might spend much of their *theorising time* on the non-axiological part of their theory, the axiological part too must be part of the full normative story at the end of the day. For another, it can be shown (Arrhenius, n.d., chapter 12) that there are straightforward deontological analogues of the impossibility theorems (concerning the relation 'ought to choose rather than' in place of 'is better than'), so that even a moral theory that has no place for axiology faces essentially the same set of

[13]Importantly (since not everyone's reflective judgments deem this principle compelling), not all of these theorems involve any very close cousin of the Mere Addition principle. For example, while Arrhenius' (n.d.) Second and Fifth theorems include a principle of Dominance Addition that is very close to the Mere Addition principle, in his Third, Fourth and Sixth theorems, Dominance Addition is replaced by an anti-Sadism condition.

[14]Denial of either the Repugnant Conclusion itself, or some close relative, *is* one of the axioms of every extant impossibility theorem. Thus, accepting the Repugnant Conclusion (and its close relatives) provides a way out of all impossibility theorems simultaneously. For instance, Totalism does not violate any of the other axioms of any of these theorems.

[15]Temkin often presents his view, deriving from the idea that certain moral ideals are 'essentially comparative', in terms of a single, choice-set-independent *but intransitive* betterness relation, rather than in terms of a multiplicity of choice-set-dependent betterness relations. However, the latter interpretation is arguably at least as well supported by the considerations Temkin appeals to, and preferable overall. For a careful analysis of these two alternatives, see (Cusbert, n.d.) . See also (Ross, 2015, section 6).

issues.

References

Adler, M. D. (2009). Future generations: A prioritarian view. *George Washington Law Review*, *77*, 1478.

Arrhenius, G. (n.d.). *Population ethics: The challenge of future generations.* (Unpublished manuscript)

Arrhenius, G. (2000). An impossibility theorem for welfarist axiologies. *Economics and Philosophy*, *16*(2), 247–266.

Bader, R. (n.d.). *Neutrality and conditional goodness.* (Unpublished manuscript)

Blackorby, C., Bossert, W., & Donaldson, D. (1995, November). Intertemporal population ethics: Critical-level utilitarian principles. *Econometrica*, *63*(6), 1303–1320.

Blackorby, C., Bossert, W., & Donaldson, D. (2005). *Population issues in social choice theory, welfare economics, and ethics.* Cambridge University Press.

Blackorby, C., & Donaldson, D. (1984). Social criteria for evaluating population change. *Journal of Public Economics*, *25*, 13–33.

Broome, J. (1991). *Weighing goods.* Oxford: Blackwell.

Broome, J. (1994). The value of a person. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *68*, 167–185.

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

Broome, J. (2009). Reply to vallentyne. *Philosophy and Phenomenological Research*, *78*(3), 747–752.

Brown, C. (2007). Prioritarianism for variable populations. *Philosophical Studies*, *134*(3), 325–361.

Bykvist, K. (2007). Violations of normative invariance: Some thoughts on shifty oughts. *Theoria*, *73*(2), 98–120.

Carlson, E. (1995). *Consequentialism reconsidered* (Vol. 20). Kluwer Academic Publishers.

Carlson, E. (1998). Mere addition and two trilemmas of population ethics. *Economics and Philosophy*, *14*, 283–306.

Cohen, D. (n.d.). *An actualist explanation fo the procreation asymmetry.* (Unpublished manuscript)

Cusbert, J. (n.d.). *Acting on essentially comparative goodness.* (Manuscript)

Fleurbaey, M., & Voorhoeve, A. (2016, July). Priority or equality for possible people? *Ethics*, *126*, 929–54.

Foot, P. (1985). Utilitarianism and the virtues. *Mind*, *94*(374), 196–209.

Hare, C. (2007). Voices from Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist? *Ethics*, *117*(3), 498–523. doi: 10.1086/512172

Heyd, D. (1988, July). Procreation and value: Can ethics deal with futurity problems? *Philosophia*, *18*(2-3), 151–170.

Heyd, D. (1992). *Genethics: moral issues in the creation of people*. Berkeley, Oxford: University of California Press.

Holtug, N. (2007). On giving priority to possible future people. In T. Rønnow-Rasmussen, B. Petersson, J. Josefsson, & D. Egonsson (Eds.), *Hommage à Wlodek: Philosophical papers dedicated to Wlodek Rabinowicz* (pp. 1–21).

Howard-Snyder, F. (2008). Damned If You Do; Damned If You Donât! *Philosophia*, *36*(1), 1–15. doi: 10.1007/s11406-007-9099-z

Huemer, M. (2008). In defence of repugnance. *Mind*, *117*(468), 899–933.

Hurka, T. (1983). Value and Population Size. *Ethics*, *93*(3), 496–507.

Hurka, T. M. (1982a). Average utilitarianisms. *Analysis*, *42*(2), 65–69.

Hurka, T. M. (1982b). More average utilitarianisms. *Analysis*, *42*(3), 115–119.

Kitcher, P. (2000). Parfit's Puzzle. *NoÃ»s*, *34*(4), 550–577. doi: 10.1111/0029-4624.00278

Meacham, C. (2012). Person-affecting views and saturating counterpart relations. *Philosophical Studies*, *158*(2), 257–287. doi: 10.1007/s11098-012-9884-9

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

Ng, Y.-k. (1989). What Should We Do About Future Generations? *Economics and Philosophy*, *5*(2), 235–253. doi: 10.1017/S0266267100002406

Ng, Y.-K. (1989). What should we do about future generations? Impossibility of Parfit's Theory X. *Economics and Philosophy*, *5*, 235–253.

Nozick, R. (1974). *Anarchy, state and utopia*. New York: Basic books.

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

Parfit, D. (1997, December). Equality and priority. *Ratio*, *10*, 202-2221.

Parfit, D. (2016). Can we avoid the repugnant conclusion? *Theoria*, *82*(2), 110–127.

Parsons, J. (2002). Axiological Actualism. *Australasian Journal of Philosophy*, *80*(2), 137–147. doi: 10.1093/ajp/80.2.137

Pummer, T. (2013). Intuitions about large number cases. *Analysis*, *73*(1), 37–46.

Qizilbash, M. (2005). Transitivity and vagueness. *Economics and Philosophy*, *21*(1), 109–131.

Rabinowicz, W. (2009). Broome and the Intuition of Neutrality. *Philosophical Issues*, *19*(1), 389–411. doi: 10.1111/j.1533-6077.2009.00174.x

Roberts, M. A. (2003). Is the person-affecting intuition paradoxical? *Theory and Decision*, *55*(1), 1–44.

Roberts, M. A. (2011). The Asymmetry: A Solution. *Theoria*, *77*(4), 333–367. doi: 10.1111/j.1755-2567.2011.01117.x

Ross, J. (2015). Rethinking the Person-Affecting Principle. *Journal of Moral Philosophy*, *12*(4), 428–461.

Scheffler, S. (1982). *The rejection of consequentialism*. Oxford: Clarendon Press. (Revised edition 1994)

Sider, T. (1991, October). Might Theory X be a theory of diminishing marginal value? *Analysis*, *51*(4), 265–271.

Singer, P. (2011). *Practical ethics* (5th ed.). Cambridge University Press.

Tännsjö, T. (2002). Why we ought to accept the Repugnant Conclusion. *Utilitas*, *14*(3), 339–359.

Taurek, J. M. (1977). Should the numbers count? *Philosophy & Public Affairs*, 293–316.

Temkin, L. (1987). Intransitivity and the Mere Addition Paradox. *Philosophy & Public Affairs*, *16*(2), 138–187.

Temkin, L. (1993). *Inequality*. Oxford University Press.

Temkin, L. S. (2012). *Rethinking the Good : Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.

Thomas, T. (n.d.). *Some possibilities in population axiology*. (Unpublished manuscript)

Vallentyne, P. (2009). Broome on moral goodness and population ethics. *Philosophy and Phenomenological Research*, *78*(3), 739–746.