No Harm, No Foul: A Person-Affecting Population Principle

Brandt van der Gaast, University of Twente, October 2016, draft.[1]

Abstract: Person-affecting consequentialism is the view that an action's moral status depends solely on how the action affects people. On this approach, notions of relative changes in well-being (i.e. being benefited or being harmed) are more central than notions of absolute amounts of well-being (i.e. being well-off or being poorly off). I develop a person-affecting principle of population ethics, and argue that it requires the cardinal measurability of well-being, and intra- as well as interpersonal comparability of well-being. I show how the theory can be modified to accommodate egalitarian intuitions. I then discuss and reply to three criticisms: the Transitivity Objection, the Non-Identity Problem, and the Asymmetry Objection. I conclude by presenting arguments against two competing views: critical-level utilitarianism and 'no deprivation' views.

1. Introduction

Why do we care about reducing our carbon footprint and building a sustainable economy? About providing structural aid to developing countries? Or about investing in medical research into gene testing? For one thing, because these technologies and policies promise to provide benefits for the future. They are likely to have a positive impact on the lives of future people. But what exactly is the nature of our obligations to future generations? From which moral principles do these obligations derive? The field of population ethics takes up this and other questions.

In some moral dilemmas, it is up to the agent to decide who is harmed and who is not. In others, an agent's choice determines who lives and who dies. But there are also moral dilemmas where the agent's action has an effect on the identities of the people to ever have existed. These effects are likely to be indirect, but they are not less real. Two actions can differ in that one leads to an expansion of the population while the other one does not. Two actions can differ in that one leads to a population consisting of certain people, while the other leads to a population of certain *other* people. Do we require special principles in our moral theory to judge the permissibility of such actions?

Thinking about the impact of population change has a long history. Thomas Malthus famously warned against what he perceived were the dangers of overpopulation in his *An Essay on the Principle of Population* from 1798. Lowering the birthrate, he argued, would result in a higher living standard for all.

[1] This paper has benefited from discussions with Peter Vallentyne, Michael McDermott and Fred Feldman.

Ever since then, writers have used Malthusian ideas as jumping-off points for developing their own views on population issues, some of which now appear clearly morally objectionable. In the early 1900's, for instance, the eugenics movement was popular. This goal of this unsavory movement was to produce a better (happier?) population by stopping certain people from pro-creating.

The goal of this paper is to formulate a plausible principle of population ethics. This principle concerns the moral rightness of actions—especially actions where the population increases or changes. In the search for this principle, axiological questions about value will rear their head. The most important of these is: How does the overall value (or 'goodness') of a situation depend on the value for the individuals that exist in it? In addressing these questions, our topic veers into the field of welfare economics. This branch of economics studies so-called 'social welfare functions': mathematical functions that characterize how social welfare depends on individual welfare. The issue of equality plays a key role here, as we will see below.

Many of the theories discussed here are versions of consequentialism. If you, the reader, are not inclined towards consequentialism, you might see this as a reason to toss this paper aside. That would be premature, though. This paper ought to be interesting to anyone who believes that there are situations where the goodness of an action's consequences can make that action more worthy of choice. Moral rightness is most likely a highly complex property, consisting of various elements that all constrain the moral choiceworthiness of actions. It is hard to deny that in some situations the overall value of an action's outcome can give the agent a moral reason in favor of choosing that action. That is all that you, the reader, have to agree with in order for the arguments and theories from this paper to be of interest.

The plan for the paper is as follows. The next section covers traditional forms of utilitarianism and their drawbacks. In Section Three I introduce a person-affecting form of consequentialism. This theory as formulated is not egalitarian enough, so in Section Four I propose a modificaton. Section Five discusses some of the theoretical commitments about the extent to which well-being or utility can be measured. I then move on to three criticisms of the person-affecting approach: in Section Six, the Transitivity Objection; in Section Seven, the Non-Identity Problem; and in Section Eight, the Asymmetry Objection. Competing views are discussed in Section Nine. The first of these is 'critical level utilitarianism', forms of which are defended by John Broome, Charles Blackorby, Walter Bossert and David Donaldson. The second competing view is one that adopts a so-called 'no deprivation' principle. Melinda Roberts and Peter Vallentyne defend versions of such a principle.

2. Totalism and Averagism

Let us define ‘consequentialism’ as the view that a person morally ought to perform the action that out of all the available actions has the best consequences. Simply put: one ought to bring about the best available outcome. ‘Hedonism’, in turn, can be defined as the view that what is intrinsically good is pleasure, and that something is intrinsically valuable to the extent that it ‘contains’ pleasure. The best outcome, on hedonism, is the one that contains the most pleasure. Traditional utilitarianism can then be regarded as the combination of a normative theory (consequentialism) and an axiological theory (hedonism).

Jeremy Bentham held the view that one ought to maximize “the greatest good for the greatest number.” But what is the greatest good for the greatest number? Is it simply the sum of all individual well-being? Or the average? Bentham’s remarks on the issue do not clearly answer that question; neither do John Stuart Mill’s. Henry Sidgwick did consider this question and chose for a principle of summation. He endorsed a principle that says to maximize “the product formed by multiplying the numbers of persons living into the amount of average happiness” (1947: 415-6). Sidgwick also anticipated certain questions about population expansion. What does an ethical theory recommend we do, he wondered, given that “we can to some extent influence the number of future human (or sentient) beings[?]” (1847: 414).

Utilitarian theories, in order to be complete theories, require principles of aggregation that determine overall utility on the basis of individual utility. Some utilitarians endorse what I will call ‘totalism’. On totalism, the overall value of an outcome is simply the sum of the utilities of all the individuals that exist. Other utilitarians endorse what I will call ‘averagism’, where the overall utility is the average of the utilities of all the people that exist.

Before moving on, a terminological wrinkle. I just used the phrases ‘overall value’ and ‘overall utility’. These denote a property of outcomes or alternatives. Now, one has to be careful how one understands this ‘term of art’, because it can be approached from two directions. One option is to understand overall value as directly connected to what morally ought to be done. Such an approach sees it as: the-thing-to-be-maximized according to the consequentialist, or as the property of outcomes that provides agents with a moral reason to bring them about. This can be called a *deontological* conception. Another option is to understand the overall value of an outcome as a direct function of the value for the people that exist in the outcome.[2] This can be called an *axiological* conception. Starting out with the axiological understanding of overall value is likely to lead one to a different principle of aggregation than adopting the deontological understanding.

Some ethicists, for example, claim that all value is value *for* someone, that there is no impersonal value. On the face of it, this is a plausible claim. However, it might lead one to the idea that what is valuable in an outcome simply is a matter of what is valuable for the people that exist in that outcome. Once this is established, the search can begin for the relation between overall value and what morally ought to be done. Chances are, this relation will turn out to be not so straightforward. Conversely, if one starts with a deontological understanding of overall value, the relation between overall value and what morally ought to be done is more straightforward. It simply is the property of outcomes that provides agents with a moral reason to bring them about. Approaching overall value from this direction might mean that the relation between overall value and individual value will be less straightforward.

This paper operates with the deontological understanding: an outcome's overall or social value is that property of outcomes which provides agents with a moral reason to realize the outcome. However, this does not mean this paper's claims are incompatible with the idea all value is value *for* someone.

Throughout this paper, I will be using tables to represent choice situations (see below). Each column in a table corresponds to an individual ('first', 'second', etc.). These individuals can be already-existing individuals, but also newly-created individuals. Each row in a table corresponds to an available action ('A', 'B', etc.). The available actions are mutually exclusive and jointly exhaustive. Whenever I say 'outcome A', this is short for: the outcome that results from action A. The squares in the table represent the utilities of the different individuals in the outcomes resulting from the different actions ('6', '7', etc.; I use a utility scale from 1 to 10). These utilities are understood as lifetime utilities, not as time-period utilities. And finally: The views discussed here are compatible with a wide range of theories about what utility or well-being consists in, such as hedonism, desire-satisfaction theories, but also objective list-theories.

	first	second
A	6	6
B	6	6

Oftentimes, agents find themselves in same-*people* choice situations. In such situations, none of the

[2] This amounts to what Sen 1979 calls 'welfarism'. See also Broome 2004: 30-5, 62.

agent's available actions change who exists and who does not. Other situations are same-*number* choice situations. Here, none of the agent's actions change how many people exist. The table above is a same-people choice situation and therefore also a same-number choice situation. But not every same-number choice situation is a same-people choice situation. Totalism and averagism are not sensitive to the identities of the individuals in the different outcomes, so they do not discriminate between same-number choice situations that are same-people situations and ones that are not. In the next section, we will see views that are sensitive to people's identities.

Total utilitarianism and average utilitarianism generate the same judgments of moral rightness in same-number choice situations. It is easy to see why. The totalist says: for each outcome, take the sum of the utilities of the different individuals that exist, and rank these sums from highest to lowest. The averagist says: for each outcome, take the sum of the utilities of the different individuals that exist, divide them by the number of individuals, and rank them. Since all outcomes have the same number of people existing, the averagist ends up with the same ranking as the totalist. In many different-number choice situations, however, totalism generates a different ranking than averagism.

It is well-known that both totalism and averagism have certain problems when it comes to such different-number choice situations. Indeed, many think that standard totalism and averagism are untenable in light of these. First, consider totalism. Here are two outcomes that differ only in that one contains an additional individual with positive utility.

	first	second
A	6	
B	6	5

Totalism ranks outcome B above outcome A. Intuitively, however, B is not better than A. There exists no *prima facie* moral reason for 'adding people to the world'. Many authors consider this a strike against totalism. Jan Narveson, for instance, coined the slogan, "We ought to make people happy, not happy people" (1973: 73). Derek Parfit seems to share this intuition as well. He writes, "if [a] couple do decide not to have [an] extra child, it would not be clear that they are open to moral criticism" (1982:140). Intuitively, failing to add a person to the world is not morally wrong.[3]

John Harsanyi is an averagist; her writes that, "every possible social arrangement… [is to be evaluated] in terms of the average utility level likely to result from it" (1975: 45). It is unclear whether Harsanyi spent much time thinking about different-people choice situations, but it is well known that averagism also faces a serious difficulty when it comes to these type of situations. This can be illustrated with the same case as above. On averagism, outcome B is worse than outcome A because adding the second individual lowers the average utility. But intuitively, outcome B is not morally worse than outcome A.

The argument against totalism relies on the intuition that outcome B is not better than outcome A. The argument against averagism relies on the intuition that outcome B is not worse than outcome A. If outcome B is neither better nor worse than outcome A, that means that they are morally on a par. Adding people to the world, in other words, seems to be *morally neutral*. John Broome calls this 'the intuition of neutral existence' and admits that this intuition "grips one strongly" (2004: v). In what follows, we will consider population principles that try to respect this moral intuition.[4]

3. Person-Affecting Principles

Why is outcome B not better than outcome A? A possible answer is: because the action that leads to outcome B *benefits* no one. Similarly, perhaps outcome B is not worse than outcome A because the action that leads to B *harms* no one. If we take this tack, the concepts of harming and benefiting will take center stage.

Let us start with this plausible principle concerning harm:

[3] J.J.C. Smart famously endorsed this aspect of totalism. He wrote, "Would you be quite indifferent between (a) a universe containing only one million happy sentient beings, all equally happy, and (b) a universe containing two million happy beings, each neither more or less happy than any in the first universe? Or would you, as a humane and sympathetic person, give a preference to the second universe?" (1961 in Smart/Williams 1973: 27-8). Smart seems to be operating with an axiological understanding rather than with a deontological understanding of total value. What to say about the intuition that Smart is expressing here? First off, it might turn on suppressing an 'everything else is equal'-clause. The addition of persons to the world often improves the well-being of existing people. But in order for Smart's claim to be true, everything else ought to be kept equal, including the utilities of existing people. Secondly, perhaps a preference for a more populated world is aesthetic or otherwise non-moral. Jonathan Bennett has addressed this issue. When imagining a far-away future when there might or might not be a human species, Bennett writes, "I don't regard [my pro-humanity stance] as part of my morality or, therefore, as a source of moral obligations" (1976: 67). Narveson makes a similar point. He writes, "we might prefer… a universe containing people to one that does not contain them…, but is this… a moral preference? It seems to me that it is not, and that the effort to make it one is a mistake" (1967: 72).

[4] He does not accept it, though. He has "grudgingly concluded it has to be abandoned" (2004: v).

> No Foul: An action that does not harm anyone is not morally wrong.

This seems in line with the intuition appealed to a few paragraphs ago: that all value is value *for* people. Whether No Foul is plausible or not depends on what exactly we mean by 'harm'. Consider the following definition:

> Harm: Action A harms person P if and only if there is an action B available that results in an outcome in which P is better off than P is in the outcome resulting from A.

'Benefit' can be defined in analogous fashion (just replace 'better off' with 'worse off'). Combining the two concepts, we can say: to *affect* someone is to either harm of benefit that person. The person-affecting view, then, is the view according to which the moral status of an action depends solely on the people affected by the action.

Our definition of 'harm' implies the following: A person can only be harmed by an action if he or she exists in the outcome resulting from the action, *and* in the outcome resulting from *at least one* alternative action. In other words, a person cannot be harmed if he or she exists only on one outcome. Another consequence of our definitions—a harmless one—is that in situations where there are at least three alternatives, one and the same action can both harm and benefit some person.

A number of ethicists are initially drawn to such a person-affecting approach—even ethicists who later abandon it in favor of a version of the totalist view. Larry Temkin writes about the person-affecting view that, "many think it expresses the *essence* of morality" (1987: 168; italics original). According to Peter Singer the view contains, "what is fundamentally sound about utilitarianism" (1976: 84). Even Parfit says that "most of our moral thinking" is in terms of the view (1984: 370). Yet all these people abandon the view because of a number of criticisms to be discussed and replied to below.

One option is to work out the person-affecting approach is by using a principle like:

> Harm Minimization: An action is morally right if and only if it minimizes total harm.

The total harm on an outcome is the sum of all the harm done to individuals on that outcome. Harm Minimization is stated in terms of moral rightness, but it can also be formulated in axiological terms. Then it reads: An outcome is better than another outcome if and only if it contains less total harm than that other outcome.

Benefit maximization can be defined analogously:

> Benefit Maximization: An action is morally right if and only if it maximizes total benefit.

For same-person choice situations, the two are equivalent. But for many different-person scenarios, the two are not equivalent, as we will see below.

If we want to respect the ‘intuition of neutral existence’, Benefit Maximization is not a useful principle. In the case below, Benefit Maximization says to realize option C. But intuitively, both options A and C are permissible. (The column ‘extant’ represents existing people; their utilities are not relevant.)

	extant	first
A	x	
B	x	4
C	x	6

Harm Minimization by itself also seems not to be correct. For consider the choice situation below,

	extant	first	second
A	x		
B	x	4	6
C	x	6	4

Intuitively, outcomes A, B and C are equally good. Yet Harm Minimization selects A as the best outcome, because it is the only outcome with zero harm. It seems that we need a principle that combines the two principles. I suggest the following:

> Harm Minimization Or Benefit Maximization (HB Minmax): An action is morally right if and only if it either minimizes the total harm or maximizes the total benefit.

On HB Minmax, the outcomes A, B and C above are equally good.

One side effect of adopting this disjunctive principle is that we no longer get a full ranking of outcomes. To see this, consider the following choice situation:

	first	second
A	5	5
B		7
C	4	

Here, outcome B is the best one according to HB Minmax. It minimizes total harm (0) and maximizes total benefit (2), so it ranks highest. But which action comes in second? A or C? On Harm Minimization, outcome C comes in second (it causes 1 unit of harm as opposed to A which causes 2 units of harm). On

Benefit Maximization, outcome A comes in second (it causes 1 unit of benefit, as opposed to C which causes 0 benefit). HB Minmax leaves it open which action comes in second. However, if our goal is merely to come up with a procedure for selecting optimal outcomes, then the fact that this principle generates a partial ranking poses no problem.

HB Minmax is the theory I will adopt throughout the remainder this paper, with one important modification to be discussed in the next section.

4. Inequality Aversion

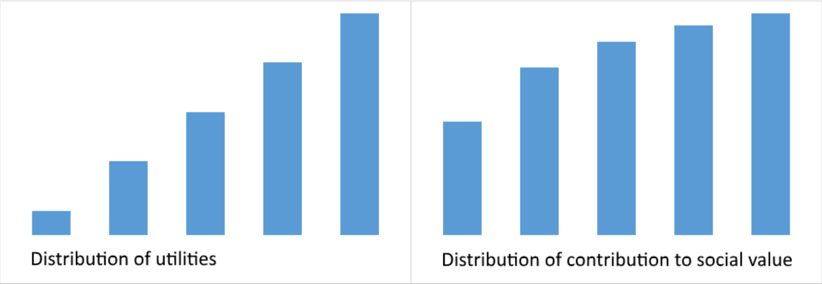
John Rawls famously endorsed a so-called 'minimax principle'. On minimax, one outcome is better than another just in case the well-being of the worst-off on the former outcome is higher than it is on the latter. In contrast to totalism and averagism, minimax focuses solely on the worst-off. Rawls writes, "Inequalities are permissible when they maximize, or at least all contribute to, the long-term expectations of the least fortunate group in society" (1971: 151). A drawback of minimax is that it makes no discriminations among situations where the worst-off are tied in terms of utility. Leximin is an improvement upon minimax. It breaks such ties by recommending changes where the well-being of the second worst-off is increased. If these are tied as well, it moves on to the third worst-off, and so on.[5]

Minimax and leximin are inequality-averse. Totalist utilitarianism is not and this is one of the drawbacks of the theory. A number of authors have addressed this issue by formulating 'generalized utilitarianism'.[6] The idea here is that the aggregation principle (summation, averaging, or some other principle) operates on *transformed* utilities, where the transformation in question is given by a function that is strictly increasing and strictly concave. A function is strictly increasing just in case its slope is positive. And a function is strictly concave just in case its slope is decreasing. The exact shape of the function represents the particular way in which inequality is to be avoided.

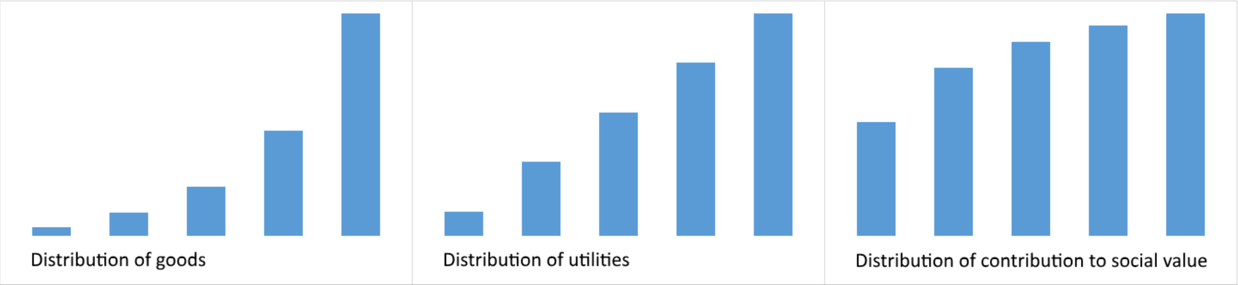
The contribution of an individual's utility to the overall utility is weighed, and individual utilities at the lower end of the spectrum are weighed more heavily than utilities at the higher end of the spectrum. In this manner, generalized utilitarianism can accommodate egalitarian intuitions. The view implies that a more egalitarian distribution of utilities is better than one that is not, *ceteris paribus*. In the diagram

[5] Both minimax and leximin seem to have the same problem with different-number choice situations as averagism: if the utility of a newly-added person is lower than that of all existing persons, the resulting outcome is worse than the status quo. Consider e.g. the example we used in the previous section to argue against totalism and averagism.
[6] E.g. Blackorby and Donaldson 1984; Blackorby, Bossert and Donaldson 2005.

below, on the left we see a linear distribution of individual utilities, while on the right we see a concave distribution of what these individual utilities contribute to social value:



It is worth pointing out that using transformed utilities in the calculation of social value is not the same as accepting the diminishing marginal utility of *goods.* The fact that a good (e.g., money) has diminishing marginal utility means that this good's contribution to an individual's utility diminishes as the person possesses more. The first dollar counts for more than the thousandth dollar, so to speak. Theorists wanting to capture this phenomenon usually apply an increasing, concave transformation function from goods to individual utilities. This function captures how the contribution of a good to an individual's utility diminishes as the individual possesses more of the good in question.



Generalized utilitarianism, on the other hand, applies different weights to different utility levels in determining their contribution to an outcome's social value. If we combine the diminishing marginal utility of goods with generalized utilitarianism, the value of goods is adjusted *twice*. In the left side of the picture above, we see a convex distribution of goods among a number of individuals. Applying the principle of the diminishing marginal utility of goods, their contribution to individual utilities is displayed in the middle, where we see a linearly increasing line. A generalized utilitarian will then apply another weighing procedure, where increases in utility on the lower end of the spectrum count more towards social value than similarly-sized increases in utility at the higher end. So on the right, there is a concave shape. Welfare economists have shown that if this second transformation function is very concave, generalized utilitarianism is almost as inequality-averse as minimax (Blackorby and Donaldson 1984: 22).

Our principles involving harm and benefit can also be made inequality-averse by using transformed utilities instead of plain utilities. Consider the negative social value of harm. Let us define the negative social value of the harm done to an individual on an outcome as the difference between the transformed utility of the individual on that outcome and the transformed utility of the individual on the outcome where he or she is best off.[7] Similarly for benefit. If we re-define the social value of harm and benefit in this way, we arrive at inequality-averse versions of our principles.

To see the theory in action, consider the following example:

	first	second
A	4	6
B	7	3

In this choice situation, action A causes 3 harm to the first person and 3 benefit to the second. Action B causes 3 harm to the second person and 3 benefit to the first. On standard HB Minmax, the two choices are on a par. But if we calculate total harm and benefit using transformed utilities, then the negative social value of the harm on A decreases while the positive social value of the benefit on A increases. Similarly, the negative social value of the harm on B increases, while the positive social value of the benefit on B decreases. So if we use transformed utilities, outcome A is better than B. And this is in line with our egalitarian intuitions.[8]

By using a person-affecting principle with transformed utilities, we can also address cases like following:[9]

	extant	first	second
A	x		
B	x	3	3
C	x	1	6

[7] We could also define ‘harm’ directly in terms of these two transformed utilities, but this would have the drawback that the harm to an individual depends upon the utility levels of his peers—a somewhat implausible implication. So instead of re-defining ‘harm’, we modify the principle that calculates the social disvalue of harm.

[8] However, if we change the current same-people choice situation into the following same-number situation, the verdict changes:

	extant	first	second
A	X	4	
B	X	7	
C	X		3
D	X		6

Here, the harm on outcomes B and D is zero. The transformed benefits on outcome B is larger than the benefit on outcome D. If it is morally right to *either* minimize harm or maximize benefit, then both B and D are permissible. This strikes me also as being supported by intuition.

[9] A variation on this case was suggested to me by McDermott.

As far as Harm Minimization is concerned, A is the winner, because both B and C cause harm. So we only have to consider benefit: outcome B causes 2 units of benefit, whereas C causes 3 units of benefit. So it appears that regular HB Minmax judges the actions leading to A and C to be morally right. But, intuitively, action B is also morally right (perhaps action C is even morally wrong). This is where the transformation function comes in handy. If our transformation function of utilities is concave enough, the positive social value of the benefit on B will be equal to or greater than that on C. HB Minmax with transformed utilities will then judge both actions A and B as morally permissible, which is as it should be.

5. Comparisons of Utility

There are different views on how *commensurable* utilities are across time, across possibilities, and across people. These views differ in the type of comparison of utility that they allow. One important distinction is between the weaker view that allows only ordinal measurability and the stronger view that also allows cardinal measurability. Another important distinction is between the weaker view that allows only in*tra*personal comparability and the stronger view that also allows in*ter*personal comparability. I will discuss these two distinctions in turn. After that, I will briefly consider the concept of the zero point, and the notion of summation.

On any theory that accepts ordinal measurability, any finite set of utilities can be ranked from lowest to second-lowest to third-lowest… all the way to second-highest and highest. Mere ordinal comparability does not imply that there is an answer to a question like, “Is the increase from lowest to second-lowest larger than the increase from second-highest to highest?” Consider a group of siblings, ordered in a line from the youngest to the oldest with the youngest one on the left. From merely looking at the line-up, you know each sibling’s place in the ranking. But you do not know how much older each one is than his or her left neighbor, or how much younger than his or her right neighbor.

A cardinal ordering, on the other hand, is informationally richer. Continuing our illustration: knowing the siblings’ precise ages means knowing the exact differences in age among them. Similarly, if utilities are cardinally measurable, then an individual’s utility on one alternative can be represented with a number, and his utility on another alternative with some other number. On a cardinal scale, the size of the difference between these two numbers is significant. Mere cardinal measurability does not attach importance to the choice of unit, or to the choice of zero point. It is similar to the way temperature can be expressed; the Celsius, Fahrenheit and Kelvin scale are cardinally equivalent, even though they differ in their unit and in their zero point.[10]

In addition to ordinal and cardinal comparability, there is also ratio-scale comparability. Here, the choice of zero point does have significance. The kilometer and mile scale, for example, are ratio-scale equivalent because they coincide on their zero point. Totalist utilitarianism requires a zero point on the utility scale. Earlier, we quote Sidgwick who considered the issue of population expansion. He wrote that on utilitarianism such an expansion is only permissible if the well-being of the newly-added people outweighs the loss in well-being of the existing people. Sidgwick writes, "if we foresee… that an increase in numbers will lead to a decrease in average happiness or *vice-versa*… we ought to weigh the amount of happiness gained by the extra number against the amount lost by the remainder" (1874: 415). ('Gained' is not really the right word for Sigdwick to use here, as these extra individuals cannot be said to have any happiness on the alternative where they do not exist.)

But how can a zero point on the utility scale be calibrated? What would be a good reason for choosing one calibration over another? Some utilitarians hold that a life with zero individual utility is a life such that living it is as good for a person as not living it (Broome 2004: 234, 254; Blackorby/Bossert/Donaldson 2005: Ch. 2). A life with zero utility is on the borderline of being worth living for the person living it. Any life with negative utility, they say, is a life better not lived than lived.[11]

But does it make sense to draw a distinction between lives that are worth living and lives that are not worth living? It is not easy to get a good grasp of what exactly this amounts to. If this distinction cannot be clearly understood, then this calibration procedure leads nowhere. My point of criticism is not that all lives are worth living, or that no life is worth living.[12] My criticism is also not that this distinction differs from person to person or that it is vague. My view is that it is nigh impossible to get a handle on the notion that some lives are, and some lives are not, worth living for the people living them. For instance, I

[10] See Blackorby, Bossert and Donaldson 2005, Chapter 2. In technical terms, the Celsius, Fahrenheit, and Kelvin scale are 'increasing, affine transformations' of each other.

[11] Note that this concerns *individual* utility, not social utility. In a later section, we will consider a view where there is also a zero point when it comes to social value. Lives with a certain level of utility, proponents of this view say, have zero social value (i.e. they do not make the outcome any better).

[12] See Benatar 2006 for a defense of this latter claim.

do not think the notion of a life worth living can be straightforwardly connected to the concept of a person's willingness to continue with their life.[13]

Luckily, views like Harm Minimization or HB Minmax do not require any zero-point calibration. They only require relative comparisons of utility. This means that they do not need the zero point on the utility scale to represent anything meaningful. In other words, these views do not require ratio-scale measurability; they can get by with mere cardinal measurability. If it is tricky it is to make sense of ratio-scale measurability for utilities, as I have suggested, then this is a strike in favor of the current view.

Why believe that utility is cardinally measurable? The most important argument for cardinal measurability is the famous Von Neumann/Morgenstern theorem from 1947. The key idea in this theorem is that the strength of people's preferences can be figured out by looking at their willingness to take certain gambles. Consider a lottery. For an individual, the value of participating in a lottery depends on the value he attaches to the prizes in the lottery, and the likelihood that he will win those prizes. The Von Neumann/Morgenstern theorem shows the following: On the basis of two sets of facts (viz. the value for the individual of participating in the lottery, and the likelihoods of winning the different prizes), we can infer another set of facts (viz. the value he attaches to the different prizes in the lottery).

A simple example can serve as demonstration. Suppose I prefer a burrito to a slice of pizza, and a slice of pizza to a hamburger. Let us now construct a lottery—a simple coin flip—where heads = burrito and tails = hamburger. Would I prefer pizza over playing in this lottery? If so, then the pizza's utility is closer to that of burrito than to that of hamburger. Would I be indifferent? Then pizza's utility is right in between that of burrito and that of hamburger. Of course, preferences for food items are a whole different ball game than the lifetime utilities we have been talking about. But the Von Neumann and Morgenstern theorem shows that there is no in principle obstacle to measuring utilities cardinally.

Next up, there are views that accept intra- but not in*ter*personal comparability and views that accept both. On mere intrapersonal comparability, the well-being of an individual can be compared across times and across alternatives. A person's well-being at one time can be said to be higher than his well-

[13] Broome believes that the value of a life (for the person living it) depends upon the value of the stretches of time within that life (2004: Ch. 15). A stretch of time within a life can have zero value, on his view. This happens when the well-being at that stretch of time is at "the level such that a person's continuing to live through an extra period at that level is equally good for her as dying" (235). A life consisting of only such moments, then, is a life of zero personal value. "[A] life that is, throughout, just on the borderline of being worth continuing is, taken as a whole, just on the borderline of being better lived than not lived" (256). It strikes me that this claim relies on an 'intra-life' aggregation principle that is too simplistic to be plausible.

being at some other time; and his well-being on one alternative can be said to be higher than his well-being on some other alternative. On in*ter*personal comparability, one person's well-being can be said to be higher or lower than some *other* person's well-being. The Von Neumann/Morgenstern theorem says nothing about interpersonal comparability.

Does our view require interpersonal comparability? The calculation of harm or benefit to a single person requires only intrapersonal comparability across alternatives. However, once we sum the individual harms on an outcome to the total harm, we need interpersonal commensurability. Interpersonal comparability is also appealed to in the transformation functions discussed in the previous section. Giving a certain weight to higher levels of utility and giving a different weight to lower levels of utility requires comparing utility levels across people, and so it requires cardinal measurability as well as interpersonal comparability.

Finally, the issue of summation. Our proposed theory differs from totalist utilitarianism in that it does not simply sum the utilities of individuals. Instead, summation is applied elsewhere: Our principle sums the harm and benefit done to individuals on an outcome in order to determine the total harm and total benefit done on that outcome. If summation as an aggregation principle in moral theories is objectionable, then our proposed theory faces a problem. Our theory would then be committed to this problematic procedure, just like totalist utilitarianism, even though it applies summation in a different way.

The most famous attack on summation principles in ethics is John Taurek's 1977 article 'Should the Numbers Count?' In this article Taurek discusses trolley-style scenarios and famously argues that in such scenarios there exists no obligation to save the greater number. In these situations, each person deserves your help equally; it does not follow, he argues, that that you ought to save the larger number. If one is persuaded by Taurek's argument or by other anti-summation arguments, one could modify the person-affecting view in certain ways. For instance, one could adopt a version of a harm minimization view where one ought to choose the alternative on which the biggest harm is the smallest. This version could be called 'Minimax Harm'.[14] This theory requires interpersonal comparability, but does not use summation anywhere.

[14] Minimax Harm is structurally closely related to the decision-theoretic principle of *minimax regret*. See for instance Resnik 1987: 28.

In this paper I argue that HB Minmax generates moral judgments that in many cases agree with our moral intuitions. This provides us with reasons to accept the theory and so also with reasons in favor of any of the principles that make up the theory. In this section I claimed that HB Minmax is committed to cardinal measurability, intra- and interpersonal measurability, and a principle of summation. Any evidential support in favor of HB Minmax is indirect evidential support for these commitments as well. As usual, the proof of the pudding is in the eating.

6. The Transitivity Objection

Now, for the first objection. Broome presents the following example where an agent has the choice of bringing about one of these three outcomes:

	extant	John
A	x	
B	x	5
C	x	7

Broome then proceeds to make the following pairwise comparisons: If we compare A to B using the person-affecting principle, they are morally on a par, because neither on A nor on B is anyone harmed or benefited. Outcomes A and C are also on a par, for the same reason. However, C is a better outcome than B, because on B John is worse off than he would have been on C. Broome writes, “The principle implies, then, that [C] is equally good as [A], [A] is equally as good as [B], but [C] is better than [B]. This is a contradiction. As a matter of logic, the relation ‘equally as good as’ is transitive, and the… principle implies that it is not” (1994: 170).

Another transitivity argument can be found in Gustav Arrhenius 2003:

	first	second	third
A	5		7
B	7	5	
C		7	5

Using pairwise comparisons, we seem to obtain the following results: B is better than A, C is better than B, yet, A is better than C! On the basis of transitivity, one would expect A to be worse than C. Many philosophers consider such arguments to be an insurmountable problem for any person-affecting view. Broome says that they show the view to be “ultimately incoherent” (1994:168) and Parfit that they show the view to have “self-contradictory premises” (1976: 102).

Both Broome and Arrhenius use pairwise comparisons to generate a ranking between A, B and C. But this is not a procedure that our theory recommends. In order to determine harm and benefit, *all* the available alternatives must be taken into account. Applying Harm Minimization to Broome's example, we get the following results. Outcomes A and C tie for first place (because they both involve zero harm); B comes in second. On Benefit Maximization, we get the following ranking: C is first (because it is the only outcome on which someone is benefited); A and B tie for second place. On our hybrid view, HB Minmax, outcomes A and C are ranked highest.

Views like these are not *end-state views* (Michael McDermott 1984: 175). They evaluate the outcomes of actions relative to the choice situation in which they are available. This relates to the distinction made earlier between the deontological and the axiological approach. The deontological approach sees the overall value of outcomes as related to the choice situation that can give rise to them. On our view, the value of an outcome cannot be determined independently of such a choice situation. The axiological approach does not relate the value of outcomes to these choice situations. Larry Temkin labels the axiological approach the 'intrinsic aspect view'; on such an approach, "how good [a] situation is all things considered… will be based solely on the internal features of the situation" (1987: 159).

Our theory HB Minmax rejects a condition known in decision theory and welfare economics as the *independence of irrelevant alternatives* (Amartya Sen 1977, 1993). Formulating this principle in terms of the 'better than' relation, it reads as follows:

> Independence of Irrelevant Alternatives: If A is better than B relative to a choice set X, then A is also better than B relative to choice set Y, where Y is a proper subset of X.[15]

The person-affecting view we are considering does not satisfy this principle. To see an illustration, consider the choice situation below. In the first case, A is better than B. But in the second choice situation, A is not better than B (they are equally good). The Independence of Irrelevant Alternatives is violated.

[15] This principle can also be formulated in terms of a choice function, but then its formulation requires two parts. It requires: *contraction consistency* and *expansion consistency*. Contraction consistency says: If A is to be chosen from a choice set X, then A is also to be chosen from choice set Y, where Y is a proper subset of X. Expansion consistency says: If A is to be chosen from choice sets X_1 and from X_2 and from X_3… X_n, then A is also to be chosen from choice set Y, where Y is the union of X_1 and X_2 and X_3… X_n (Sen 1993).

	extant	John
A	x	
B	x	5
C	x	7

	extant	John
A	x	
B	x	5

Is this a drawback? A number of authors have presented examples designed to show that the principle does not need to be adhered to, because it is too demanding. Sen, for instance, discusses an example of a guest who is offered cake. Will he take the biggest slice from the plate? Suppose the guest has good manners and takes the second-largest slice. He considers the second-largest slice better than all others. Now, what if he had been offered that same plate *minus* that largest slice? Then the slice he actually picked (the second-largest one) would be the largest slice. But it would no longer be better than all others. So the removal of an option changes the agent's ranking of the remaining options. Sen claims that this reveals no irrationality on the part of the agent.

Michael Resnik provides another example involving food (1987: 40). A customer is looking over the menu in a cheap and shabby-looking restaurant. He sees two items: hamburger and roast duck. The customer fears that the kitchen is not very good, so orders the burger. The waiter then informs the guest that they also have sautéed frog legs. The guest now thinks the cook might have skill and goes for the roast duck because he likes duck. Again, this violates the Independence of Irrelevant Alternatives. The addition of an option changes the agent's rankings: what was previously a sub-optimal choice (the roast duck) now becomes the best choice. It seems there is no irrationality on the part of the customer.

However, these analogies have their limits. Whether the agents in these two examples meet the standards of rationality depends on how their options are *described*. In Sen's example, the choiceworthiness of the different slices is not just a function of their size. The goal of the agent is not simply to maximize the amount of cake eaten, but to make a good impression on the host. And in Resnik's example, the addition of the third option changes the nature of the first two options. As Resnik himself says, "the old acts were *order hamburger at a seedy place*, *order roast duck at the same seedy place*. But the new acts do not include these since you no longer think of the restaurant as seedy"(40). So Sen and Resnik's remark do not settle the issue.

Summing up our reply to the Transitivity Objection: Views like Harm Minimization and HB Minmax are not consistent with the Independence of Irrelevant Alternatives. This is because the view adopts a deontological conception of overall or social value, instead of an axiological one. It is not obvious that disagreeing with this principle is a big cost to the theory, however. The burden of proof is on those who insist that a plausible population principle must satisfy this independence principle.

7. The Non-Identity Problem

Parfit's *Reasons and Persons* introduced the Non-Identity Problem.[16] His example:

> *The 14-Year-Old-Girl*. This girl chooses to have a child. Because she is so young, she gives her child a bad start in life. Though this will have bad effects throughout this child's life, his life will, predictably, be worth living. If this girl had waited for several years, she would have had a different child, to whom she would have given a better start in life.[17]

Would it be wrong for the girl to have the earlier child? According to the person-affecting view, the woman acts wrongly in having the early child only if this child is the same person as the happier child that would result if the woman waited. In that case, the woman would inflict more harm than she needed to. But, argues Parfit, the early child would not be the same person as the later child. Why not? Because the different timing of conception would result in a different person. He writes, "on all the plausible views… it is in fact true that, if you had not been conceived within a month of [the time you were actually conceived], *you* would never have existed" (1984: 355; italics original).

If that is true, then on the person-affecting view the woman is not harming anyone in having the early child. Following our principle that doing wrong requires harm, the woman is not doing anything wrong in having the earlier child. But, intuitively, the woman *is* acting wrongly in having the early child. Says Parfit, "it would have been better if this girl had waited, so that she could give to her first child a better start in life" (1984: 359).

Does Parfit's argument undermine the person-affecting approach? I claim that it does not. The problem with the non-identity argument is that it is suspect, because it makes use of an *nonintuitive premise*. Intuitively, we think that the *same* person results if only the timing of the conception is changed. The premise that a different person will result if the timing of conception is changed seems initially as nonintuitive as the conclusion of the Non-Identity Problem. Off the bat, we think that the same person

[16] Boonin 2014 is a book-length treatment of the Non-Identity Problem.
[17] Parfit 1984: 358.

results if the woman has the child now or later. As a result, we think that the woman who has the not-so-happy child harms the child. To echo McDermott, who arrives at the same judgment, “It is no objection to [the theory] that it yields an anti-intuitive conclusion when combined with an anti-intuitive judgment of identity” (1982: 166).

Instead of,

	extant	child	child
A	x	5	
B	x		7

we imagine the situation to be like so:

	extant	child
A	x	5
B	x	7

In the first of these two situations, neither A nor B cause any harm. Both are permissible. In the second of the two situations, action A is not permissible because it inflicts more harm than B.

Many proponents of the Non-Identity Problem do find the premise about non-identity intuitive or uncontroversial. Parfit, for instance, writes that it is “not controversial and easy to believe” (1984: 351). Kavka writes that it is a “plausible premise” and a “basic fact” (1981: 95). But when it comes to philosophical methodology, one should tread carefully. It strikes me that theories should be judged on their own merits. It should not count against a theory that it is incompatible with a widely-held, yet non-intuitive claim implied by a certain theory of personal identity.

But consider again the situation with two different newly-added people. Is it so obvious that it is permitted to bring into existence a not-so-happy person, when the possibility exists to bring into existence a different, very happy person? I have pretended it to be morally permissible, but not everybody agrees. Narveson, for instance, writes, “if [you] have a choice of which to produce, [you should] produce the happier one, other things being equal” (1978: 56). Singer appears to arrive at a similar verdict. He writes, “there will be a minimum number of lives being lived [regardless of what we choose], and it is by its effects on the happiness of that number of lives that [an action] should be judged” (1976: 88). It is not clear, however, whether these authors are reporting a pre-theoretic intuition or whether they are spelling out an implication of their view.

However that may be, this view runs into other serious problems. For consider the following situation:

	extant	first	second
A	5		7.1
B	7	5	

In this same-number choice situation, there is one person that exists on both outcomes, and there are two newly-added persons, a different one for each outcome. The difference in utility between the two newly-added people is slightly larger than the difference in the extant person's utility on the two outcomes. Totalist utilitarianism takes into account the sum total of utilities, so it judges action A as morally right. Intuitively, however, outcome A is not better than outcome B.

Now, a totalist might deny the intuition that action A is not better than action B; he might insist that A *is* better. This is not a promising strategy, however, because the result that A is not better than B follows from the 'intuition of neutral existence' that we discussed in section II, together with some fairly minimal assumptions. To see this, consider the following situation where we add a third outcome C:

	extant	first	second
A	5		7.1
B	7	5	
C	5		

It appears that the following things hold: First, outcome A is not better than C (this is a result from the intuition of neutral existence). Second, outcome C is not better than B (if anything, it is worse). But if A is not better than C, and C is not better than B, it follows that A is not better than B. So we are back to square one.

The intuition of neutral existence—together with some very minimal principles—implies that situation A in the preceding example is not better than situation B. So, again, it seems that the main trouble with totalism is that it denies the plausible intuition of neutral existence. (It is worth noting that this line of reasoning relies on the Independence of Irrelevant Alternatives principle. However, since we are criticizing totalism, and since standard forms of totalism include this principle, we can use it in arguments against the position.)

In short, our response to the Non-identity Problem is the following. The argument relies on a premise that is as counterintuitive as the conclusion, meaning that the premise is to be blamed and not the person-affecting theory. I then considered choice situations where two different persons, with different levels of well-being, can be added to the world. I argued that there exists no obligation to choose for the happier of the two persons, by using an argument that relies on not much more than the intuition of neutral existence.

8. The Asymmetry Objection

On person-affecting utilitarianism, there is no *prima facie* obligation to create persons that fare well. But what about people that fare very poorly? Suppose there are lives that are so bad that they are not worth living. Does the person-affecting view imply that it is permissible to bring into existence individuals with such lives? A number of authors have considered this problem for the person-affecting view.

	extant	child
A	x	-4
B	x	

Suppose a woman can have a child whose quality of life is guaranteed to be extremely low. Her alternative course of action is not to have the child. Assume, unrealistically, that everything else is equal. On person-affecting utilitarianism, the woman would not be inflicting harm if she decided to have the child. Our earlier principle No Foul says that an action that does no harm is not morally wrong. Bringing about outcome A does no harm, so according to this principle action A is not morally wrong. We seem to have arrived at an important asymmetry, for intuitively her action seems wrong.

Many writers on population ethics agree that choosing action A would be morally wrong. Jonathan Bennett writes that “it is wrong to bring into existence someone who will be miserable” (1976: 61). Parfit concurs; he says, “it would be wrong to have the wretched child” (1984: 391).

The standard reply for defenders of the person-affecting approach is to modify the theory. McDermott, for instance, writes, “the following people [are also] relevant…: anyone alive on *one* alternative, if he is miserable on that alternative” (1982: 165; italics original). Earlier, when we formulated the person-affecting view, we said that the well-being of non-affected people does not contribute or detract from an outcome’s overall value. On the proposed modification, this is no longer true: a life can detract from an outcome’s overall value if the well-being in this life is very low. Other proponents of person-affecting theories that modify the theory in this fashion include Christopher Meacham and Melinda Roberts.[18]

[18] Meacham 2012, Roberts 1998.

McDermott admits in so many words that this move is *ad hoc*. He writes that the theory "is not *deep* enough. It offers no explanation for the difference… in its treatment of newly-created happy people and newly-created miserable people" (169). Being *ad hoc* is not the final straw for a theory, but it is nevertheless a drawback.

But there is a bigger problem. In what way do we take into account the lives of newly-added people who are very poorly off? To what degree do such lives detract from the social or overall value of an outcome? The issue is that the 'special harm' that can be done to newly-created people must be commensurable with the 'regular harm' that can be done to existing people. But, intuitively, it is difficult to wrap one's head around such a comparison. Consider for instance the situation depicted below. Is the harm done on outcome A bigger or smaller than the harm done on B? It is unclear.

	first	second
A	8	-4
B	4	

I submit to have no moral intuitions about situations where relative decreases in well-being (i.e. instances of harm) are weighed against the absolute levels of well-being of people who are very poorly off. To me, this suggests that our judgment that it is wrong to bring such people into existence has a *different origin*. What I am suggesting is that these intuitions flow from a different aspect of our moral thinking. In Section Five, I made some critical remarks about views that have a zero point on the individual utility scale. A drawback of totalist utilitarianism, I argued there, is that is it requires absolute quantities of utility to be comparable with relative changes in utility. The current proposal shares this aspect with totalism. If I am correct that our intuitions about such comparisons are weak to non-existent, then we better not adopt this proposal.

The issue is not that the Asymmetry Objection relies on the distinction between lives that are and lives that are not worth living. The objection can be formulated without that distinction. Everyone has to admit that there are lives that are very short and filled with nothing but misery and suffering. Our search is for a theory that implies that it is morally wrong to bring individuals into existence that are guaranteed to live such lives. The Asymmetry Objection cannot simply be dismissed by refusing to accept the distinction between lives that are worth living and lives that are not.

Singer approaches the Asymmetry Objection from a different angle. He claims, "[I]f for [a certain] reason it is not obligatory to bring a happy person into the world, then by a symmetrical form of reasoning it cannot be wrong to bring a miserable being into the world either" (1976: 93). Singer goes on to say that once such a person comes into existence, there exists a moral obligation to carry out an act of euthanasia on this person. It needs no argument that this is a very extreme and also implausible view. We better not take this path.

What could be the origin of the intuition that bringing people into existence whose lives are guaranteed to be of extremely low quality is morally impermissible? There are a number of different possible routes here. In the introduction, I suggested that moral rightness might be a complex property, with one component being a consequentialist one. So it is open to maintain that a different, a non-consequentialist component of our concept of moral rightness gives rise to this particular judgment. Since this paper is only concerned with consequentialist aspects of moral rightness, I will not pursue this suggestion here any further.

But perhaps the asymmetry can be grounded in consequentialist grounds, after all. Let me briefly and inconclusively sketch how this might go. There is a version of consequentialism where not only outcomes have intrinsic value, but the *acts* leading to those outcomes themselves as well. G.E. Moore seems to have held this view. In *Principia Ethica*, he writes, "In asserting that the action is the best thing to do, we assert that *it together* with its consequences presents a greater sum of intrinsic value than any possible alternative." (1903, §17; italics added). A morally right act, on this view, can be an act that "has greater intrinsic value than any alternative [act], whereas both its consequences and those of the alternatives are absolutely devoid either of intrinsic merit or intrinsic demerit" (Ibid.)

Now, if acts can have positive intrinsic value, then presumably they can have intrinsic negative value as well. For instance, the act of failing to keep a promise might be said to have some intrinsic negative value. Let me now return to the case of newly-added people who are very poorly off. The important difference between newly-added happy people and newly-added miserable people is that when the latter come into existence, there exists a moral obligation to improve their lives. But often such obligations to care for others cannot be fulfilled. This means that on the outcome where a person with a very low quality of life is created, there will exist unfulfilled obligations. And such failures to act might have negative intrinsic value. A situation where there exist unfulfilled obligations seems worse than one where such things do not exist, *ceteris paribus*.

Whether this can be developed into a satisfactory reply to the Asymmetry objection remains to be seen. But the main conclusion from this section is the following. The person-affecting theory can be modified so that the total harm on an outcome not only depends on relative changes in the well-being of people who exist on multiple outcomes, but also on the absolute well-being of very poorly-off people who exist on one outcome only. However, I have argued that this is not a promising approach, because it requires that relative changes in well-being be comparable with absolute levels of well-being. I claimed that there is little intuitive support for such comparisons.

9. Two Competing Views

This section discusses two competing approaches. The first of these is 'critical-level utilitarianism' and is defended by Broome and by Blackorby/Bossert/Donaldson (Broome 2004, Blackorby/Donaldson 1984, Blackorby/Bossert/Donaldson 2003, 2005). The second competing approach employs a so-called 'no deprivation' principle. Versions of this approach are defended by Vallentyne and Roberts (Vallentyne and Tungodden 2007, Roberts 1998). I will discuss these two approaches in turn.

Critical-level utilitarianism is a variation on totalist utilitarianism. First off, the view adopts an individual utility scale with a zero point. The zero point represents lives of neutral value: lives such that living them is as good as not living them. Secondly, the theory proposes that for every population size, there exists a *critical level*. This critical level is positive on the individual utility scale, meaning it represents lives worth living for the people living them. The key feature of the critical level is this: adding to the population an individual whose lifetime well-being is at the critical level is a *neutral* addition. It does not add to the overall value of the outcome. Adding to the population an individual whose lifetime well-being is above (or below) the critical level is a positive (or negative) addition.

An argument in favor of critical-level utilitarianism is that it avoids the repugnant conclusion. The repugnant conclusion is Parfit's well-known thought experiment where he compares a population consisting of well-off people with a population that is many, many times larger, but that consists of people whose well-being is much, much lower (Parfit 1984). The lives in the second population are of very low quality. Totalism implies that if this second population is very large, their sheer number can make up for their very low levels of well-being, resulting in higher overall value. Critical-level utilitarianism does not face the problem of the repugnant conclusion, because the individual utilities of the people in the second population are below the critical level, and so do not make the outcome better.

Because critical-level utilitarianism is a version of totalism, it does not face certain issues discussed in this paper. The theory does not face the Transitivity objection that we discussed in Section Six, because it implies the Independence of Irrelevant Alternatives. The theory does not face the Non-Identity Problem, because it is an anonymous theory that is not sensitive to the identities of people on the different outcomes. And finally, the theory also does not face the Asymmetry Objection. This is because it assigns to newly-added individuals whose lives not worth living a negative overall or social utility.

A drawback that critical-level utilitarianism has in common with totalism is that it denies the intuition of neutral existence. Of two outcomes that differ only in that the second has one additional person whose lifetime utility is above the critical level, the outcome with the extra person has a higher overall value. As I have argued in Section Two and elsewhere in this paper, there is not much intuitive support for such a judgment. Now, it is open to proponents of the theory to deny that there is a straight-forward connection between overall value and what morally ought to be done. On such an axiological as opposed to deontological approach, the higher value of an outcome does not translate into a moral injunction to realize it. Paraphrasing Narveson: There is no moral reason to create happy people, even though their addition makes the world better.

This is a valid approach, but it runs the risk of turning the notion of 'overall value' into a term of art. If this notion is not connected to what morally ought to be done, then what is its content? There is a contrast here with the notion of individual value. The notion of individual value is grounded in axiological intuitions about what well-being consists in. The notion of overall value, on the other hand, does not have such a secure footing. Do we have axiological intuitions about overall value, when this concept is wholly separated from deontological questions about what ought to be done? I doubt it.

Another downside to critical-level utilitarianism is that it implies that absolute amounts of well-being can be compared with relative changes in well-being. Broome illustrates this process with an example. He envisions a scenario where a couple can choose to have a second child, but where this causes a decrease in well-being for their existing child. He writes that the outcome where they have the second child, "is better [than the one where they do not] if the second child's lifetime well-being is sufficiently above the neutral level. The difference between it and the neutral level must be greater than the loss of wellbeing suffered by the first child" (2004: 259).

Let us see this procedure in practice:

	first	second
A	8	
B	5	7 (4+3)

Suppose the critical level is at a utility level of 3. The utility of the second child is 7, so it is 4 above the critical level. On outcome B, the extent to which the second child is above the critical level (4) is greater than the harm done to the first child (3). So outcome B ranks above outcome A.

Is making comparisons of this type part of our moral thinking? I cannot conceive of any real-life situation where moral agents would weigh the harm done to an individual against a certain level of well-being experienced by a newly-added person. This is in contrast with thinking about harm. As I have been arguing throughout this paper, it is part of our morality to compare harm done to certain people with harm to others. It is part of our moral thinking to compare total amounts of harm done on different outcomes. In other words, the person-affecting approach jives with the way we ordinarily think through the consequences of our actions, whereas critical-level utilitarianism does not.

The second competing theory I want to discuss employs a principle that Vallentyne calls 'No Gratuitous Deprivation'. Roberts also defends a theory that incorporates such a principle (Roberts 1998, Vallentyne and Tungodden 2007). Roberts uses the term 'wronging' for her central notion, but it amounts to the same thing as what Vallentyne calls 'gratuitous deprivation'. For both of these authors, this principle is merely a part of their theory of moral rightness, since it only disqualifies certain actions by labelling them as morally wrong. But even so, we can criticize their theory by criticizing this particular component of their overall theory.

The principle can be formulated in two steps, building off of our concept of harm. First off, there is a principle connecting moral wrongness to deprivation:

> No Gratuitous Deprivation: An action is morally wrong just in case it gratuitously deprives at least one person.

Second, there is a definition of 'gratuitous deprivation':

> An action gratuitously deprives a person just in case it results in an outcome where this person is harmed by being better off on some outcome where nobody is harmed.

Gratuitous deprivation is a type of harm, but it is narrower. Every instance of deprivation is an instance of harm, but not all instances of harm are instances of deprivation. In a nutshell: Being gratuitously deprived is being harmed by being better off on an outcome where there exists zero total harm.

To provide an illustration of this somewhat tricky concept, consider the choice situation below. In this situation, action A harms the first individual and action B harms the second individual. However, only action B gratuitously deprives someone. Outcome B gratuitously deprives the second individual, because he is harmed by being better off on an outcome where there exists no harm (viz. outcome C).

	first	second
A	5	
B	7	5
C		7

Is No Gratuitous Deprivation a plausible principle? I will now argue that the principle does not provide enough guidance. I will argue that it is too weak because it does not take into account the *size* of harms. Consider the following choice situation:

	first	second	third	fourth
A	6	3	5	
B	5	6		3

Here, the first and the second individual exist on both outcomes. The first individual is slightly harmed on outcome B and the second individual is more seriously harmed on outcome A. On outcome A, a third individual exists. His level of well-being matches that of the first individual on the outcome where he is worst off. On outcome B a fourth individual exists and his level of well-being matches that of the second individual on the outcome where he is worst off.

Is anyone in this choice situation gratuitously deprived? No. Both the first and second individual are harmed, but they are not harmed by being better off on an outcome with zero harm. So the deprivation principle does not judge an action as morally wrong. Intuitively, however, outcome A is worse than outcome B. In order to generate this intuition, a theory that incorporates the deprivation principle needs an additional principle. But what principle can that be? The total amount of well-being on both outcomes is the same. The number of people on both outcomes is the same. The distribution of well-being (anonymously considered) on both outcomes is the same. It seems, then, that we need a person-affecting principle to generate the judgment that A is worse than B.

Our theory delivers exactly this verdict. Outcome B is worse than outcome A because the amount of harm done on outcome B is larger than on outcome A. The preceding argument might not win over proponents of the No Gratuitous Deprivation principle. First of all, they might not accept the cardinal measurability of utility. A view that does not include cardinal measurability cannot appeal to ‘sizes’ of harm in the way that Harm Minimization does. Secondly, it is also possible that someone does not share the moral intuition that outcome A is worse. To such a critic, I do not know what to say. When offering up a moral intuition, I hope of course to be doing something more than merely spelling out an implication of a certain theory. But intuitions do diverge.

This concludes our discussion of person-affecting consequentialism. Summing up: In this paper, I have formulated a version of a person-affecting approach to population ethics, viz. HB Minmax. According to this theory, an action is morally permissible if and only if it either minimizes harm or maximizes benefit. In Section Four, I described how HB Minmax can be turned into an inequality-averse theory by using transformed utilities. In Section Five, I showed how the theory is committed to cardinal measurability, and to intra- as well as interpersonal comparability of utility. I then discussed and replied to three objectons: the Transitivity Objection, the Non-Identity Objection, and the Asymmetry Objection. I concluded by comparing the view to two competitors. The theory of harm minimization, I hope to have shown, is an approach to population ethics that has serious promise.

Literature

Arrhenius, Gustav 2000. *Population Axiology*. Doctoral Dissertation. Available at: http://www.iffs.se/media/2285/future-generations-for-homepage.pdf

Arrhenius, Gustav 2003. The Person-Affecting Restriction, Comparativism, and the Moral Status of Potential People. *Ethical Perspectives* 10, 3-4: 185-95.

Benatar, David 2006. *Better Never to Have Been: The Harm of Coming into Existence.* Oxford: Oxford University Press.

Bennett, Jonathan. 1978. On Maximizing Happiness. In Sikora, Richard I. and Brian Barry (eds.) 1978. *Obligations to Future Generations*. Philadelphia: Temple University Press.

Blackorby, Charles and David Donaldson 1984. Social Criteria for Evaluating Population Change. *Journal of Public Economics* 25: 13-33.

Blackorby, Charles, Walter Bossert and David Donaldson 2003. Population Ethics and the Value of Life. *Cahiers de Recherche en 2013, Centre Interuniversitaire de Recherche en Economie Quantitative.* Obtained from http://www.cireqmontreal.com/cahiers-de-recherche

Blackorby, Charles, Walter Bossert and David Donaldson 2005. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. Cambridge: Cambridge University Press.

Boonin, David 2014. *The Non-Identity Problem and the Ethics of Future People.* Oxford: Oxford University Press.

Broome, John and Adam Morton 1994. The Value of a Person. *Proceedings of the Aristotelian Society* supp. 68: 167-98.

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

Harsanyi, John C. 1975. Can the Minimax Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory. In John C. Harsanyi 1976. *Essays on Ethics, Social Behavior, and Scientific Explanation*. Dordrecht: D. Reidel Publishing Company.

Narveson, Jan 1967. Utilitarianism and New Generations. *Mind* vol. 76, 301: 62-72.

Narveson, Jan 1973. Moral Problems of Population. *The Monist* vol. 57, 1: 62-86. Reprinted in Sikora, Richard I. and Brian Barry (eds.) 1978. *Obligations to Future Generations*. Philadelphia: Temple University Press.

Narveson, Jan 1978. Future People and Us. In Sikora, Richard I. and Brian Barry (eds.) 1978. *Obligations to Future Generations*. Philadelphia: Temple University Press.

Kavka, Gregory 1981. The Paradox of Future Individuals. *Philosophy and Public Affairs* 11, 2: 93-112.

Meacham, Christopher 2012. Person-Affecting Views and Saturating Counterpart Relations. *Philosophical Studies* 158: 257-87.

McDermott, Michael 1982. Utility and Population. *Philosophical Studies* 42: 163-77.

McMahan, Jeff 1981. Problems of Population Theory. *Ethics* 92: 96-127.

Moore, G.E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Parfit, Derek 1979. On Doing the Best for Our Children. In Michael D. Bayles (ed.) 1979. *Ethics and Population*. Cambridge: Schenkman Publishing Company.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Rawls, John 1971. *A Theory of Justice (Original Edition)*. Cambridge: Harvard University Press.

Resnik, Michael D. 1987. *Choices*. Minneapolis: University of Minnesota Press.

Roberts, Melinda A. 1998. *Child versus Childmaker*. Oxford: Roman and Littlefield Publishers.

Sen, Amartya 1977. Social Choice Theory: A Re-Examination. *Econometrica* Vol. 45, no. 1: 53-89.

Sen, Amartya 1979. Personal Utilities and Public Judgments: Or, What's Wrong With Welfare Economics? *The Economic Journal* 89: 537-58.

Sen, Amartya 1993. Internal Consistency of Choice. *Econometrica* Vol. 61, no. 3: 495-521.

Singer, Peter 1976. A Utilitarian Population Principle. In In Michael D. Bayles (ed.) 1979. *Ethics and Population*. Cambridge: Schenkman Publishing Company.

Smart, J.J.C. and Bernard Williams 1973. *Utilitarianism: For and Against.* Cambridge: Cambridge University Press.

Taurek, John M. 1977. Should the Numbers Count? *Philosophy and Public Affairs* vol. 6, 4: 293-316.

Temkin, Larry 1987. Intransitivity and the Mere Addition Paradox. *Philosophy and Public Affairs* vol. 16, 2: 138-87.

Tungodden, Bertil and Peter Vallentyne 2005. On the Possibility of Paretian Egalitarianism. *Journal of Philosophy* 102: 126-54.

Tungodden, Bertil and Peter Vallentyne, 2007. Person-Affecting Paretian Egalitarianism with Variable Population Size. In John Roemer and Kotaro Suzumera (eds.) 2007. *Intergenerational Equity and Sustainability*. New York: Palgrave Macmillan.

Sigdwick, Henry. 1894. *The Methods of Ethics.*

Von Neumann, John and Oskar Morgenstern 1947. *Theory of Games and Economic Behavior (2nd edition).* Princeton: Princeton University Press.