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 the 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 discussing and criticizing an alternative person-affecting view.

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, the agent 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?

In 1798 Thomas Malthus famously argued that we ought to slow down population growth, to ensure a high enough level of well-being for all. Many writers since then—politicians, economists, environmentalists—have addressed the issue of population change.

The goal of this paper is not to wade into these murky political waters, but instead to formulate a plausible principle of the morality of actions that involve population change. In the search for this principle, 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 this question, 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 theories discussed here are versions of consequentialism. Does that mean that this paper is only of interest to people inclined to endorse this view? No, for even many non-consequentialists ought to believe 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

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

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 some traditional forms of utilitarianism and their drawbacks. In Section Three I introduce a person-affecting principle. This principle as formulated turns out to be not egalitarian enough, so in Section Four I propose a modification. Section Five discusses some of the theoretical commitments about the extent to which well-being can be measured. Then I move on to three criticisms: in Section Six, the Transitivity Objection; in Section Seven, the Non-Identity Problem; and in Section Eight, the Asymmetry Objection. An alternative person-affecting view is discussed and criticized in Section Nine.

2. Totalism and Averagism

Jeremy Bentham believed that one ought to maximize "the greatest good for the greatest number." But what is the greatest good for the greatest number? Is it the sum of all individual well-being? Or the average? Bentham did not discuss this question in much detail; neither did John Stuart Mill. Henry Sidgwick did address the question. He endorsed a principle that says to maximize "the product formed by multiplying the numbers of persons living into the amount of average happiness" (1847: 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[?]" (414).

Utilitarian theories, in order to be complete theories, require principles of aggregation that determine overall well-being (or utility) on the basis of individual well-being. Some utilitarians endorse what I will call 'totalism'. On totalism, the overall utility 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.

	first	second
A	6	6
B	6	6

Throughout this paper, I will be using tables like the one above to represent choice situations. Each column in a table corresponds to an individual ('first', 'second', etc.). Each row in a table corresponds to an available action ('A', 'B', etc.). The 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. Finally: The views discussed here are compatible with a wide range of theories about what utility or well-being consists in.

Oftentimes, agents find themselves in same-people choice situations. In such situations, none of the 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. A same-number choice situation need not be a same-people choice situation. Yet other situations are different-number choice situations.

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 a positive life-time 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 shares the 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.[2]

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 here. 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 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).[3] In the following section, we will consider population principles that try to respect this moral intuition.

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:

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

Whether No Foul is plausible or not depends on what exactly we mean by 'harm'. Consider the following definition:

---

[2] 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). But is a preference for a more populated universe really a moral preference? Jan Narveson thinks not; 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" (1967: 72). Jonathan Bennett concurs: "I don't regard [my pro-humanity stance] as part of my morality or, therefore, as a source of moral obligations" (1976: 67).

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

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.[4] 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 the same person.

A number of ethicists are initially drawn to such a person-affecting approach—even ethicists who later abandon it in favor of totalism or some variant. 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 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.

Benefit maximization can be defined analogously:

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

In axiological terms: An outcome is better than another outcome if and only if it contains more total benefit. For same-person choice situations, Harm Minimization and Benefit Maximization 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 judges outcome C to be the best (for it contains 2 benefit). But intuitively, outcomes A and C are equally good. (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 correct. For consider the choice situation below,

[4] For better readability, I will use male pronouns to refer to nameless individuals or persons.

	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, actions A, B and C are all three morally right.

Due to the disjunctive nature of HB Minmax, the axiological version is slightly more complicated. It reads: i) An outcome is better than another if it contains less harm *and* more benefit; ii) an outcome is as good as another if it contains less harm and less benefit, or more harm and more benefit; iii) an outcome is worse than another if it contains more harm and less benefit.

For an example of this second option, consider:

	first	second
A	5	5
B		7
C	4	

Here, outcome B is judged the best. It minimizes total harm (0) and maximizes total benefit (2). But is A better than C or is it other way around? Harm Minimization says that C > A, for 1 unit of harm is better than 2 units of harm. Benefit Maximization says that A > C, for 1 unit of benefit is better than 0 units of benefit. HB Minmax does not have a preference either way and judges A and C to be equally good.

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

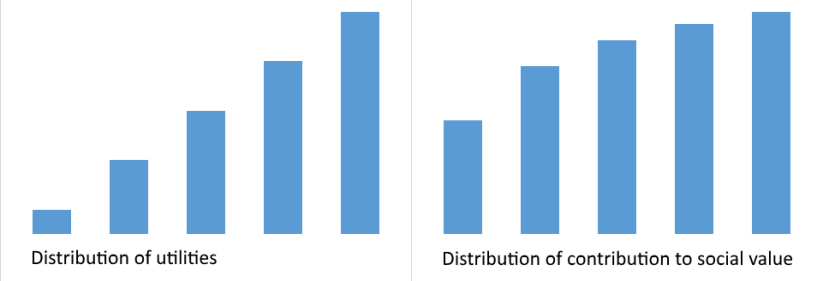
John Rawls famously endorsed a so-called 'minimax principle'. On minimax, one outcome is better than another just in case the well-being of the worst-off on the former outcome is higher than it is on the latter. In contrast to totalism and averagism, minimax focuses solely on the worst-off. Rawls writes, "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).[5]

Minimax is inequality-averse. Averagism and totalism are not and that is one of their drawbacks. A number of authors have addressed this issue by formulating 'generalized utilitarianism'.[6] The idea here is that the principle that aggregates individual utilities operates on *transformed* utilities. Before being aggregated, individual utilities are fed through a transformation function that is strictly increasing and

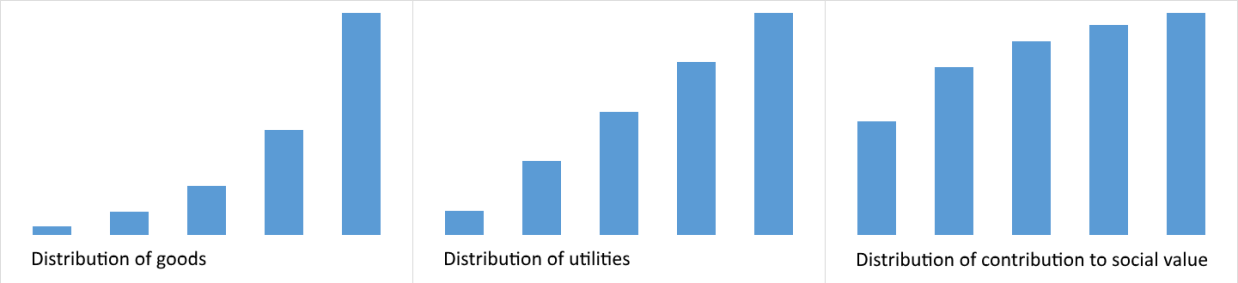
[5] Minimax has the same problem with different-number choice situations as averagism: if the utility of a newly-added person is lower than that of all existing persons, the resulting outcome is worse than the status quo.
[6] Blackorby and Donaldson 1984, Blackorby, Bossert and Donaldson 2005.

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.

Applying a transformation function to individual utilities amounts to a weighing procedure, where increases in utility at the lower end of the spectrum count more heavily towards social value than increases in utility at the higher end of the spectrum. In the diagram below, the left side shows a linear distribution of individual utilities, while the right side shows a concave distribution of their contributions to social value. On generalized totalism, for instance, all transformed individual utilities are summed up to determine the outcome's overall value. This view implies that a more egalitarian distribution of utilities is better than one that is not, *ceteris paribus*.



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 hundredth dollar, so to speak. Theorists wanting to capture this phenomenon usually apply a strictly increasing, strictly concave transformation function from goods to individual utilities. The left panel in the picture below shows a convex distribution of goods; the middle panel shows a linear distribution of individual utilities:



Generalized utilitarianism, in contrast, applies different weights to different individual utility levels in determining their contribution to an outcome's social value. So if we combine the diminishing marginal utility of goods with generalized utilitarianism, the value of goods is adjusted *twice*. Firstly, the value of goods is adjusted in the calculation of individual utility; secondly, the value of individual utilities is adjusted in the calculation of social utility. 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 person-affecting principles can also be made inequality-averse by using transformation functions. This requires re-defining harm and benefit as follows. The negative social value of an instance of harm is the difference between the transformed utility of the individual on the current outcome and the transformed utility of the individual on the outcome where he is best off. Similarly, the positive social

value of an instance of benefit is the difference between the transformed utility of the individual on the current outcome and the transformed utility of the individual on the outcome where he is worst off.

To see this modified 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 use 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. This is in line with our egalitarian intuitions.[7]

By using transformed utilities, we can also address cases like the following:[8]

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

As far as Harm Minimization is concerned, A is the winner, because both B and C cause harm. Outcome B causes 2 units of benefit, whereas C causes 3 units of benefit. So regular HB Minmax judges outcomes A and C as better than B. But, intuitively, outcome B is not worse than C (perhaps outcome C is actually worse than B). This is where the transformation function comes to the rescue. If our transformation function is concave enough, the positive social value of the benefit on B will be equal to or greater than that on C. If the positive social value of the benefit on B is equal to that on C, HB Minmax with transformed utilities will judge all three outcomes are equally good. If the positive social value of benefit on B is greater than that on C, HB Minmax with transformed utilities will judge A and B as the best outcomes.

I will adopt this inequality-averse version of HB Minmax in the remainder of this paper.

5. Comparisons of Utility

There are different views on how *measurable* well-being or utility is. Some views accept only the weaker ordinal measurability, while others also accept the stronger cardinal measurability. Another distinction is that between the weaker intrapersonal comparability and the stronger interpersonal comparability. In

[7] 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.

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

this section, I consider which types of measurability and comparability HB Minmax is committed to. I also briefly discuss the notion of summation.

Suppose we have a complete ordinal ranking of alternatives for an individual. This means that any two alternatives are such that, for the individual, the one is better than the other, as good as the other, or worse than the other. Consider a ranking where A > B > C. On ordinal measurability, there is no answer to a question like, 'Is the difference between A and B bigger or smaller than the difference between B and C?' To give an analogy: Consider a group of siblings, ordered in a line of increasing age, with the youngest one on the left. From merely looking at the line-up, you know each sibling's rank, but you do not know the exact differences in age among them. The line-up is an ordinal representation.

A cardinal ordering, on the other hand, is informationally richer. Imagine lining up the siblings in such a way that the distance between any two neighboring siblings represents the gap in age between them. This is a cardinal representation. Applied to individual utilities, it means that an individual's utility on one alternative can be represented with a number, and his utility on another alternative with another number. The size of the difference between these two numbers is significant. Cardinal measurability does not attach importance to the choice of unit, or to the choice of zero point. Compare temperature: the Celsius, Fahrenheit and Kelvin scale are cardinally equivalent, even though they differ in their unit and in their zero point.[9]

Ratio-scale measurability is yet stronger. Here, the choice of zero point *does* have significance. On ratio-scale measurability, it is meaningful to speak of absolute quantities of the thing measured. Take the kilometer scale, for instance. It is meaningful to speak of an absolute quantity in distance, expressed in a number of kilometers. The kilometer scale is ratio-scale equivalent to the mile scale, because the two coincide on their zero point.

But how can a zero point on the individual utility scale be calibrated? Some utilitarians maintain that a life with positive utility is worth living, whereas a life with negative utility is not worth living. A life with zero 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).

But does it make sense to draw a distinction between lives that are worth living and lives that are not? I do not want to suggest that all lives are worth living, or that no life is worth living.[10] Instead, my point is that it is nigh impossible to get a handle on the distinction between lives worth living and lives that are not. For instance, I do not think the notion of a life worth living can be straightforwardly connected to the concept of a person's willingness to continue on with their own life.[11] Theories committed to ratio-scale measurability face the challenge of finding a meaningful way to calibrate the individual utility scale.

---

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

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

[11] 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.

Totalism requires a zero point on the utility scale. Earlier, we quoted Sidgwick who considered the issue of population expansion. Sidgwick wrote, "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; italics original).[12] ('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.) In the totalist calculation that Sidgwick proposes, absolute levels of well-being must be comparable to relative decreases in well-being.

But is it part of our moral thinking to make such comparisons? On totalism, the absolute utility of a newly-added person can make an outcome better, and the relative decrease in the utility of an existing person can make an outcome worse. But how do these two types of change stack up against each other? This comes back to the intuition of neutral existence we appealed to earlier. If one shares this intuition of neutral existence, one will also be skeptical of such comparisons of absolute and relative utilities.

Views like Harm Minimization, Benefit Maximization, or HB Minmax do not require ratio-scale measurability, but only cardinal measurability. This is because they center around notions like harm and benefit; notions that involve *differences* between individuals' utilities on different alternatives. If it is problematic to conceptually justify the notion of absolute levels of utility, as I have been suggesting, then that is a strike in favor of the current view.

But why believe that utility is even cardinally measurable? An important argument here is the 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 gambles. Consider a standard 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).[13] If this is correct, then there is no in principle obstacle to representing preferences in terms of cardinal utility.

Next up, there are views that accept intra- but not in*ter*personal comparability and views that accept both. On mere intrapersonal comparability, the utility of an individual can be compared across times and across alternatives. A person's utility at one time, or on one alternative, can be said to be higher or lower than his utility at some other time, or on some other alternative. On in*ter*personal comparability, one person's utility or well-being can be said to be higher or lower than some *other* person's utility. The Von Neumann/Morgenstern theorem says nothing about interpersonal comparability.

Does HB Minmax require interpersonal comparability? The calculation of harm or benefit to a single person only requires the 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

[12] I am ignoring Sidgwick's "vice-versa". The principles discussed in this paper take into account the utilities of the people to ever have existed, and it is impossible for the number of people to ever have existed to decrease.

[13] 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.

section. Giving a diminished weight to higher levels of utility and an increased weight to lower levels of utility requires comparing utility levels across people, and so requires interpersonal comparability.[14]

Finally, the issue of summation. HB Maxmin uses a procedure of summation to calculate total amounts of harm and total amounts of benefit. 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 there exists no obligation to save the greater number. In these situations, each person deserves your help equally; but it does not follow, he argues, that you ought to save the greater number. If Taurek is correct, then many moral theories that incorporate summation principles are in trouble—including ours.[15]

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 have claimed that HB Minmax is committed to cardinal measurability, to intra- and interpersonal measurability, and to 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 an example where the agent has the choice of bringing about one of these three outcomes:

	extant	John
A	x	
B	x	5
C	x	7

Broome argues as follows: 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 is 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).

Many philosophers consider such arguments to present 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).

Broome uses pairwise comparisons to generate a ranking among 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

[14] It also requires cardinal measurability (but not ratio-scale measurability).

[15] Unusual versions of the person-affecting approach are still viable, even without any summation principle. For example, consider the view that says to choose the outcome on which the biggest individual harm is the smallest. This version could be called 'Minimax Harm'. This theory still requires intra- and interpersonal comparability, but does not use summation anywhere. Such a view is structurally similar to the well-known decision-theoretic principle of minimax regret. See e.g. Resnik 1987: 28.

Benefit Maximization, we get the following ranking: C is first (because it contains 2 benefit); A and B tie for second place. On our hybrid view HB Minmax, outcomes A and C are both ranked as best.

HB Minmax is an approach to overall or social value that is sometimes called *deontological*.[16] On the deontological conception, judging one outcome better than another is only meaningful in relation to a particular choice situation. A different approach is the *axiological* conception. Here, an outcome's being better than another only depends on the intrinsic overall value of both, and these values can be considered independently of any choice situation. The view that Amartya Sen dubs 'welfarism', for instance, is an example of the axiological approach.[17] On welfarism, an outcome's overall utility is a function *only* of the utilities of the people that exist in that outcome. Larry Temkin calls such a view the 'intrinsic aspect view'; on such an approach, he says, "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* (Sen 1977, 1993).[18] 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.[19]

HB Minmax does not satisfy this principle. Broome's example can serve as an illustration. In the example, A is judged better than B. But in the choice situation below, A is not better than B. The Independence of Irrelevant Alternatives is violated.

	extant	John
A	x	
B	x	5

Is this a drawback? A number of authors have presented examples designed to show that the Independence of Irrelevant Alternatives is too demanding. Sen discusses an example of a guest who is offered cake. Will he take the biggest slice from the plate? Suppose the guest is hungry but also has good manners and therefore takes the second-largest slice. He considers this slice better than all others. Now, what if he had been offered that same plate *minus* the largest slice? Then the slice he actually picked 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 says 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 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 the restaurant also offers sautéed frog legs. The guest now thinks the cook might have skill and goes for the

---

[16] E.g. in Tungodden and Vallentyne 2007.

[17] Sen 1979 contains a critical discussion of welfarism. See also Broome 2004: 30-5, 62.

[18] But only for overall value, and not for individual value. Incidentally, The Von Neumann/Morgenstern theorem relies on the Independence of Irrelevant Alternatives—but for individual value.

[19] 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).

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.

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 merely to satisfy his hunger, it is also to make a good impression. And in Resnik's example, the addition of the third option changes the nature of the first two. 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; italics original). So Sen and Resnik's remarks do not settle the issue.

Summing up our reply to the Transitivity Objection: HB Minmax is inconsistent with the Independence of Irrelevant Alternatives. This is because it relies on a deontological conception of overall or social value, instead of an axiological conception. 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 the principle.

7. The Non-Identity Problem

Parfit's *Reasons and Persons* introduced the Non-Identity Problem.[20] He gives an example of a 14-year-old girl who can have a child now or have a child later. If she has a child now, she will not be able to give the child a good start in life; if she waits several years, her child's life will turn out significantly better. Would it be wrong for the girl to have the earlier child?

According to the person-affecting view, the girl acts wrongly only if she inflicts more harm than is necessary. Is she inflicting harm in having the earlier child? That depends upon whether the early child will be the same person as the later child. Parfit argues that it will not be the same person, because of the different timing of conception. "[O]n 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).

On the person-affecting view, then, the girl is not harming anyone in having the early child. Following our principle that doing wrong requires harm, the girl is not doing anything wrong in having the earlier child. But, intuitively, the outcome where the girl has the early child is worse. 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).

Is this a problem for the person-affecting approach? The first thing to note is that the argument is suspect, because it makes use of an nonintuitive premise. The premise that a different person will result if the girl waits seems initially as nonintuitive as the claim that she's not doing anything wrong in having the early child. Off the bat, we think that the *same* person results if only the timing of conception is changed. As a result, we think that if the girl has the early child, she 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,

[20] 1984: 358. Boonin 2014 is a book-length treatment of the Non-Identity Problem.

	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 outcomes are equally good. In the second of the two situations, outcome A is worse because it contains 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 runs into trouble when combined with a widely-held, yet nonintuitive claim implied by a certain theory of personal identity.

In order for argument to be a convincing one, it needs a premise with two *obviously* different people. But in a situation with two obviously different people, we lose the intuition that the outcome with the happier person is better than the outcome with the less happy person.

Or do we? Not everyone 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 arrives 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).

Such a view, however, runs into trouble because it conflicts with the intuition of neutral existence discussed earlier. For consider the following situation:

	extant	first	second
A	5		7.1
B	7	5	

Here, 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. Some views might judge outcome A as better than B. Intuitively, however, A is not better than B.

Now, someone like Narveson or Singer might deny that outcome A is not better than B; he might insist that A *is* better. This is not a promising strategy, however, because this leads to a conflict with the intuition of neutral existence.

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		

Intuitively, it appears that the following things hold: First, outcome A is not better than C (this follows 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.

The conclusion of this line of reasoning is that outcome A is not better than B relative to the choice situation where the alternatives are: A, B and C. Many consequentialist views accept the Independence of Irrelevant Alternatives. And if we apply this principle, we can conclude that outcome A also is not better than B relative to the choice situation where the alternatives are only A and B. So, again, the view runs into trouble when it comes to the intuition of neutral existence.

Summing up, our reply to the Non-Identity Problem is that the argument relies on a premise that is no less counterintuitive than the conclusion; the premise is to be blamed, in other words, and not the person-affecting theory. I then argued against claims endorsed by Narveson and Singer. They cannot respect the intuition of neutral existence while also accepting the Independence of Irrelevant Alternatives.

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? Does the person-affecting view imply that it is permissible to bring into existence individuals that will live lives with very low levels of well-being?

	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 to not have the child. Assume—unrealistically—that everything else is equal. On person-affecting utilitarianism, the woman is not inflicting harm if she has the child. Bringing about outcome A does no harm, so action A is not morally wrong. We seem to have arrived at an important asymmetry, for intuitively her action *does* seem wrong.

Many writers on population ethics agree that choosing action A is 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 a 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 formulating the person-affecting view, I 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.[21]

[21] Meacham 2012, Roberts 1998: 152 ff.

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. For in what way do we take into account the lives of newly-added people who are very poorly off? How do such lives detract from the overall value of an outcome? The issue is that the 'special harm' that can be done to poorly-off individuals that exist on one outcome must be commensurable with the 'regular harm' that can be done to individuals that exist on multiple outcomes. 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 beliefs 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 totalism, 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. The current proposal denies the intuition of neutral existence when it comes to people who are very poorly off.

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 filled with nothing but misery and suffering. The search is for a theory on which 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.

What could be the origin of the intuition that bringing people into existence whose lives are guaranteed to be of extremely low quality is 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 negative intrinsic value as well. For instance, the act of failing to keep a promise could be said to have negative intrinsic value. Now, consider a newly-added person who is 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. A situation where there exist unfulfilled obligations seems worse than one where these 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. I 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.

9. A Competing View

Let me finally discuss an alternative person-affecting view. This view employs a principle that Vallentyne calls 'No Gratuitous Deprivation' (Tungodden and Vallentyne 2007). Roberts also defends a theory that incorporates such a principle (Roberts 1998). Roberts uses the term 'wronging', but her notion amounts to the same thing. For both of these authors, this principle is merely a small part of their theory of moral rightness, since it only excludes certain actions by labelling them as morally wrong. Even so, we can criticize their theory by criticizing this component of it.

The principle can be formulated in two steps. 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 the central notion:

> An action gratuitously deprives a person if it results in an outcome where this person is harmed by being better off on an outcome where nobody is harmed.[22]

[22] Tungodden and Vallentyne 2007 (Tungodden does not accept No Gratuitous Deprivation). In her 1998 book, Roberts formulates a principle called 'D*' (1998:63). In a footnote, she suggests a slight modification. This version of D* says that a person is 'wronged' (gratuitously deprived) on outcome A if there exists an outcome B such that 1) this person is better off on B, 2) other individuals that exist on both outcomes are at least as well-off on B, and 3) individuals existing only on B are as well-off as is possible. Our simplification in the main text captures the same principle.

Gratuitous deprivation is a special type of harm. Every instance of deprivation is an instance of harm, but not every instance of harm is an instance 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 concept, consider the choice situation below. In this situation, action A harms the first individual and action B harms the second individual. But 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 argue that the principle does not provide enough guidance. The principle is too weak, I claim, because it does not take into account the *size* of harms. Consider the following 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 agree with this intuition, a person-affecting 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 our view 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 total harm or maximizes total 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. Finally, I compared the theory with a competitor.

Literature

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

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.