Multidimensional Consequentialism and Population Ethics

Martin Peterson

Is it better to bring about a world in which a small number of people enjoy high levels of wellbeing, or would it be better to bring about a world with a much larger population in which everyone has a life barely worth living? Much work in population ethics has focused on a series of negative results. Parfit argues in *Reasons and Persons*[1] hat none of the theories he discusses is morally acceptable. Ng[2] and Arrhenius[3] generalize Parfit's conclusion by showing that *no* theory of population ethics is consistent with a number of intuitively reasonable normative desiderata. They do this by proving a set of formal impossibility theorems.

In this chapter, I argue that multidimensional consequentialists have reason to reject some of the key premises of Parfit's Mere Addition Paradox, as well as Arrhenius' sixth impossibility theorem. The latter is the most general and far-reaching impossibility theorem in the literature on population ethics. My main point is that multidimensional consequentialists can reasonably maintain that the mere addition of people who have lives worth living is not always entirely right. To add what Parfit calls "extra people" is right with respect to one moral aspect (the number of lives worth living) but wrong with respect to another (the average quality of life.)

Multidimensional consequentialists believe that an act's rightness or wrongness depends on several irreducible moral aspects.[4] If no act is optimal with respect to all aspects, then no act is entirely right. On the multidimensional view, moral rightness and wrongness literally come in degrees. An act is right to the highest degree if and only if it is optimal with respect to all applicable aspects, meaning that if two or more aspects clash even the best alternative is somewhat wrong.

The claim that an act's deontic status depends on several irreducible aspects can be clarified by an analogy with geometry:[5] The area of a triangle depends on two variables (its height and the length of its base) but the area of a circle depends on only one variable (its radius). In a similar vein, multidimensional consequentialists believe that an act's deontic status depends on at least two irreducible aspects or variables (for instance, equality and wellbeing), whereas one-dimensional consequentialists believe that an act's deontic status depends on a single aspect or variable.

For a straightforward example of a one-dimensional view, consider total hedonistic act utilitarianism. According to this theory, the only aspect that determines an act's deontic status is whether it maximizes happiness. The fact that both the intensity and duration of a pleasurable experience matter (as well as the intensity and duration of painful ones) does not make the theory multidimensional. These factors can be reduced to a single aspect: the total amount of happiness brought about by the act. What matters for the classification of a theory as multidimensional is whether it is *possible* to characterize an act's deontic status as a function of a single aspect. By definition, theorists who believe that we should assign weights to different aspects and then maximize the weighted sum, or aggregate them in some other way, defend one-dimensional views. If it is possible to aggregate two or more aspects into a new, composite aspect, then those aspects could obviously be reduced to a single composite aspect.

This chapter has five further sections. Section 3.2 recapitulates the Repugnant Conclusion and the Mere Addition Paradox. Section 3.3 explains the basic tenets of multidimensional consequentialism. Section 3.4 examines multidimensional consequentialism in greater detail and responds to a couple of objections. Thereafter, in sections 3.5 and 3.6, the multidimensional analysis of what I take to be the core issues in population ethics are discussed at length.

<1> **3.1. The Repugnant Conclusion and the Mere Addition Paradox**

The Repugnant Conclusion is one of the key results in Parfit's (1984) discussion of population ethics. It is derived by comparing a series of future worlds A, B, C …, Z. The width of each rectangle in Figure 3.1. shows the number of people living in that world. The height represents their level of wellbeing. In all worlds, all people have lives worth living.

[INSERT FIGURE 3.1. HERE]

Many traditional consequentialists believe that B is better than A, given that there are sufficiently many more B-people than A-people, and given that the quality of life for the B-people is just a little bit lower than that of the A-people.[6] (Note that there is no inequality in neither of these worlds.) For the same reason, C is better than B, and the same applies, *mutatis mutandis*, to every intermediary world between C and Z. However, as famously pointed out by Parfit, this line of reasoning leads to the following Repugnant Conclusion:

> For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population [Z] whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living.[7]

If we accept the premise that "better than" is transitive, then Z is better than A. However, if we compare Z and A, we may find this conclusion hard to accept. Intuitively, A is better that Z. But it cannot be true that Z is better than A *and* that A is better than Z.

The Repugnant Conclusion is an important element of the Mere Addition Paradox. In order to state the Mere Addition Paradox, imagine a world A+ in which there is one group of

people of the same size as the only group in A, with the same high quality of life. There is also another group of people in A+ that Parfit calls "extra people". They all have lives worth living and affect no one else. See Figure 3.2. Now compare Divided B to A+. Because the average level is higher in Divided B, this world seems to be better than A+. The extra people in Divided B benefit more than what is lost by the best-off people in A+. Moreover, Divided B is exactly as good as B. However, if we put all these moral judgements together we arrive at a problematic conclusion: A+ is not worse that A, Divided B is better than A+, and B is equally as good as Divided B. Therefore, B is not worse than A, and by repeating this argument sufficiently many times it follows that Z is not worse than A. But we also believe that Z is worse than A, because we reject the Repugnant Conclusion. We have now contradicted ourselves.

[INSERT FIGURE 3.2. HERE]

A possible response to the Mere Addition Paradox is to give up the premise that "better than" and "not worse than" are transitive relations.[8] Parfit has recently expressed sympathy for this solution.[9] However, Arrhenius (2000, 2016) shows that analogous contradictions arise if the premises are formulated as normative adequacy conditions about the rightness and wrongness of acts, rather than as evaluative claims about better or worse worlds.[10] This indicates that Parfit's solution is not as general as one might have hoped.

<1> **3.2. Multidimensional Consequentialism with Fixed Populations**

Before the multidimensional theory is applied to cases with variable population size it is helpful to consider how it deals with cases in which the identity of all individuals is fixed.

Consider the example in Table 3.1. The numbers denote each individual's wellbeing. Act 1 brings about the largest sum total of wellbeing (100 units). Utilitarians believe that Act

1 is right, while the other two alternatives are wrong. Strict egalitarians, who care about nothing but equality, believe that Act 3 is the only right alternative. The prioritarian analysis depends on which priority function is applied. A reasonable and frequently mentioned priority function is the root function.[11] By calculating the square root of all the numbers in Table 3.1., it can be easily verified that Act 2 is the only right alternative according to this and many other priority functions.

[INSERT TABLE 3.1 HERE]

Multidimensional consequentialists analyze the alternatives in Table 3.1. differently. For expository purposes it is helpful to imagine that equality and total wellbeing are two irreducible moral aspects, as suggested in the introduction. I will not try to defend this assumption here. My goal is just to explain how the multidimensional analysis differs from one-dimensional ones. In sections 5 and 6 I will adopt a slightly more complex version of the multidimensional view, according to which each individual person counts as a separate moral aspect.

None of the alternatives in Table 3.1. is optimal with respect to both total wellbeing and equality. Act 1 is optimal with respect to total wellbeing but scores poorly with respect to equality. Act 3, on the other hand, is optimal with respect to equality but scores poorly with respect to total wellbeing. Act 2 is not optimal with respect to any of the two aspects, although it scores well with respect to both of them. Multidimensional consequentialists, therefore, believe that because none of the available alternatives is optimal with respect to all applicable aspects, no alternative is entirely right. All three alternatives are somewhat right and somewhat wrong, although some of the somewhat right acts are right to a higher degree than the others.[12] Because Act 2 is almost right with respect to each of the two aspects, and thus also a little bit wrong with respect to each of them, it is *all things considered* much closer to being entirely right than to being entirely wrong. Act 1 is, all things considered,

right to a much lower degree because it scores so poorly with respect to equality. The same holds *mutatis mutandis* for Act 3.

The non-binary notions of moral rightness and wrongness adopted by multidimensional consequentialists are technical concepts, which may or may not reflect the ordinary meaning of those terms. There is little reason to think that our existing vocabulary is always the most appropriate one. The scientific meaning of "heat" and "one meter" differ significantly from the original, ordinary meanings of those terms. Just like in the sciences, our willingness to accept the technical vocabulary proposed by multidimensional consequentialists should depend on whether this is useful for describing phenomena the moral theorist has reason to care about, rather than on what most people ordinary mean when they use those terms.[13]

An act's degree of rightness can be represented by a real number between 0 and 1. From a technical point of view, this is unproblematic. As long as the number of degrees is not larger than the unaccountable infinity of real numbers in the unit interval this representation will not restrict the multidimensional consequentialist. However, the technical details of a numerical account can be spelled out in different ways.[14] In what follows I will state what I believe to be one of the most attractive numerical versions of multidimensional consequentialism for fixed populations. According to this version of the theory, an act's degree of rightness with respect to total wellbeing is determined by calculating the ratio between the total amount of wellbeing brought about by the act (on a normalized positive scale) and the maximum amount obtainable by any alternative act.[15]

Equality can be measured in different ways. One of the most widely used measures is the Gini index. Because the Gini index of a perfectly equal distribution is zero, and one represents maximal inequality, an act's degree of rightness with respect to equality equals one

minus its Gini index.[16] By calculating the Gini index for each distribution of wellbeing in Table 3.1. we obtain the numbers in the rightmost column of Table 3.2.

[INSERT TABLE 3.2. HERE]

Table 3.2. summarizes the three acts' degree of rightness with respect to each of the two aspects: total wellbeing and equality. Each act's degree of rightness, all things considered, is a function of these variables. But how should we aggregate rightness with-respect-to-a-moral-aspect to rightness all-things-considered?

One possibility could be to maintain that the overall degree of moral rightness is determined by the moral aspect with which the act scores least well, such that the act's overall degree of rightness equals the degree to which the act is right with respect to that aspect. This is the maximin account of all-things-considered rightness. It holds that Act 2 is right to degree .99 while Act 1 is right to degree .52, and Act 3 is right to degree .02. This is the view I find most plausible myself.

That said, one could also assign weights to each aspect and then calculate the weighted average degree of rightness and wrongness of each alternative. This is the weighted average account. Its main weakness is that it is hard to understand where the weights would come from. A third possibility, which I believe is inferior to both the first and second, is to maintain that the overall degree of moral rightness is determined by the moral aspect with respect to which the act scores best. This is the maximax account. The chief objection to the maximax account is that it makes it "too easy" for an act to become entirely right. For instance, according to the maximax rule, both Act 1 and Act 3 in Table 3.2. would be entirely right, but not Act 2. This is not very plausible.

<1> **3.3. Moral Aspects and Degrees of Rightness**

In the previous sections I offered two examples of moral aspects: equality and total wellbeing. These examples were chosen for expository purposes. According to the version of multidimensional consequentialism I believe to be most plausible myself, each person's individual wellbeing level, not the total wellbeing in the population, counts as a moral aspect.[17] I shall return to my preferred version of the multidimensional theory in section 3.5. The aim of the present section is to spell out the distinction between multidimensional and one-dimensional theories in more general terms, without referring to any particular examples of moral aspects.

Let me first say a few words about the relation between aspects and dimensions. Generally speaking, a dimension is a conceptual space in which an aspect can vary. For instance, the height and width of a rectangle are two different aspects that belong to the same physical dimension (length), but the mass and velocity of a falling object are examples of different aspects belonging to different dimensions. In previous work, I have argued that the wellbeing enjoyed by each person, rather than the total wellbeing enjoyed by the group, forms a set of different moral aspects belonging to the same moral dimension (persons).[18] However, even if one thinks that it is the wellbeing experienced by each person that counts, not the wellbeing enjoyed by the group, it does not follow that one is a multidimensional consequentialist. This is because utilitarians, for instance, acknowledge that there is a sense in which one could say that an act's deontic status depends on the wellbeing enjoyed by each person: individual wellbeing affects the total wellbeing. So in *that* sense utilitarians could say that an act's deontic status depends on more than one aspect.

However, utilitarians also believe that we can *reduce* all those aspects (the wellbeing enjoyed by each person) to a single composite aspect: the total wellbeing enjoyed by all individuals together. It was in order to steer clear of this type of maneuver I pointed out in the

introduction that multidimensional consequentialists believe that an act's deontic status depends on several *irreducible* aspects.

That said, it is worth keeping in mind that multidimensional consequentialism, technically speaking, allows for views in which all the different moral aspects belong to the same dimension. Some multidimensional views may, in fact, turn out to be merely "multiaspectual". However, because this term is a bit clumsy I prefer to use the term multidimensional, even when I discuss views holding that all aspects belong to the same dimension.

Let me elaborate a bit on how I use the term *moral aspect*. Consider the following definition:

> something counts as a moral aspect if and only if it directly influences an act's deontic status, irrespective of how other aspects are altered. That something directly influences the deontic status of an act should be understood as a claim about functional relationships: an aspect, *a*, directly influences the deontic status, *d*, of an act if and only if *d* is a function of *a*.[19]

A potential problem with this definition is that the term "direct influence" can be interpreted in more than one way, even when we add the condition that it must be possible to capture this influence by some mathematical function. For instance, a possible but not very plausible interpretation of direct influence is to read it as a causal notion. Utilitarians do not believe that an act's rightness is caused by the fact that it brings about the greatest amount of wellbeing. Many utilitarians rather believe that the act's property of being right supervenes on its property of bringing about the largest amount of wellbeing. However, note that if we read "direct influence" as "supervenes on", then we open up for a new set of metaphysical

worries. This is because our understanding of supervenience is almost as unclear as our understating of causation. Many technical and slightly different notions of supervenience have been proposed in the literature, but it does not help much to say that direct influence means (some particular form of) supervenience until we have been able to explain what exactly supervenience is.[20]

So how should we understand the term direct influence? I believe the best we can do, without giving up the claim about functions, is to appeal to the analogy with geometry mentioned in the introduction. The area of a triangle is a function of, and is thus directly influenced by, two aspects (its height and the length of its base) but the area of a circle is a function of, and thus directly influenced by, just one aspect (its radius). How do we know this? Well, if we increase the length of the base while holding the height of the triangle constant, its area increases. The same applies *mutatis mutandis* to circles. As far as I am aware, there is no philosophically uncontroversial way of explaining in what sense the variables of a function directly influence the output of a function that goes beyond this.

In response to this, Ralf Bader has objected that the notion of direct influence captured by the notion of a function is implausible, because:

> Understanding what it is for something to directly influence something else in terms of the latter being a function of the former is not particularly plausible. This is because functional relationships can be chained since functions can be composed, whereas (any intuitive understanding of) direct influence does not allow for chaining. If x is a function of y, and y is a function of z, then x will be a function of z…This, however, does not mean that z directly influences x.[21]

Bader is right that there is *some* way of understanding "direct influence" that would make the notion of a function misplaced, for the reason he mentions. For instance, if you directly influence my political opinions by having a long conversation with me (perhaps you point out that 91 prisoners were executed in Texas between January 1, 2010 and November 15, 2016, which you think is 91 people too many) and I then directly influence the political opinions of my neighbor by telling her what you told me, it seems to be a mistake to conclude that you have directly influenced the political opinions of my neighbor. It appears that your influence on my neighbor's political opinion was indirect.

However, it does not follow from this there is *no* intuitive understanding of direct influence that allows for chaining. The best response to Bader's objection is thus to point out that we should understand the notion of direct influence *such that* it allows for chaining. Consider, for instance, average utilitarianism. Is this a multi- or one-dimensional view? Well, average utilitarianism can be described in more than one way. We could say that average utilitarianism is the view that an act's deontic status is a function of a single moral aspect: its average utility. However, we could also say that average utilitarianism is the view that an act's deontic status is a composite function in which each individual's wellbeing is a separate moral aspect. The second view can then be characterized by a composite function in which all individual wellbeing levels are first transformed to an average value by the first function, and the second function assigns a deontic predicate such as "right" or "wrong" to the act depending on how high this average value is.

This suggests that there *is* a sense in which the average utility directly influences the act's deontic status, and because it is possible to *reduce* all aspects to a single aspect, average utilitarianism is a one-dimensional view, not a multi-dimensional one. So the fact that a function can be chained with other functions is not a threat to the distinction between one- and multidimensional consequentialism.

Doug Portmore has proposed another type of objection to multidimensional consequentialism. He asks us to consider a situation in which the multidimensional consequentialist is faced with a choice among three alternative distributions of wellbeing:

> My considered moral intuition is that Act 2 [in Table 3.3.] is entirely right. And this is what utilitarianism implies. MDC [multidimensional consequentialism], by contrast, implies that Act 2 is somewhat wrong, for MDC is suboptimal both in terms of equality and in terms of Alice's wellbeing, each of which are moral aspects on MDC. Indeed, MDC implies that all three alternatives are somewhat wrong. So utilitarianism seems to do better in this case.[22]

[INSERT TABLE 3.3. HERE]

Portmore correctly points out that utilitarians and multidimensional consequentialists analyze the situation in Table 3.3. differently. Utilitarians believe that Act 2 is entirely right in the binary sense. Multidimensional consequentialists, on the other hand, believe that Act 2 is right to a high degree, although not entirely right. This is because another alternative, Act 3, brings about more equality.

The gist of Portmore's objection is that it is *absurd* to maintain that Act 2 is somewhat wrong. On his view, Act 2 (but not Act 1 or Act 3) is entirely right. Before I discuss what multidimensional consequentialists have to say in response to Portmore's objection, it is worth calculating exactly *how* right and wrong the three alternatives in his example actually are according to this theory. A reason for doing this is that the calculation may affect our considered intuition about the case.[23]

Let us suppose, once again, that equality and wellbeing are two separate moral aspects. Now, if an act's all-things-considered rightness is calculated according to the maximin rule, it can be easily verified that Act 1 is, all things considered, right to degree

.499950, while Act 2 is right to degree .999999 and Act 3 is right to degree .000001 (see Table 3.4.).

[INSERT TABLE 3.4. HERE]

The numbers in Table 3.4. indicate that the disagreement between utilitarians and multidimensional consequentialists about Act 2 is, from a practical point of view, quite minuscule. Utilitarians believe that this act is entirely right in the binary sense (that is, right to degree 1 in the multidimensional vocabulary), while multidimensional consequentialists believe that Act 2 is "merely" right to degree .999999. So what is at stake in this discussion?

I believe the numbers in Table 3.4. reveal that utilitarians and multidimensional consequentialists have roughly the same, but not exactly the same, intuitions about the three alternatives in Portmore's example. The key difference is that multidimensional consequentialists, unlike utilitarians, can capture the fact that Act 2 brings about a tiny bit of inequality in their final moral verdict. So by concluding that Act 2 is right to degree .999999 instead of 1, multidimensional consequentialists can articulate a *more nuanced* verdict than utilitarians. Act 2 is almost entirely right, but because the act produces some inequality it is also a tiny bit wrong.

Let me now explain where I believe Portmore's reasoning goes wrong. As I understand his objection, he believes that the morally *best* alternative is always entirely *right*, even if the best alternative is not optimal with respect to all moral aspects. Otherwise, he cannot explain why he thinks Act 2 is entirely right. However, it is a mistake to believe that the best alternative is always entirely right, for at least two reasons.

Firstly, Portmore's assumption prevents us from expressing the type of non-binary moral nuances that feature in the multidimensional analysis described above. If we believe that these moral nuances are important moral phenomena that a plausible moral theory should be able to account for, then we ought to reject Portmore's assumption.

Secondly, there is an important *practical* difference between the verdicts reached by utilitarians, who accept Portmore's assumption, and multidimensional consequentialists, who reject it. In another paper, Portmore formulates the following principle, which he thinks, "is, perhaps, the most fundamental and least controversial normative principle concerning action"[24]:

> we ought to perform the alternative that we have most reason (all things considered) to perform. Of course, there could be more than one alternative that is tied for first place. So, to be a bit more careful, I should say that we ought to perform one of the alternatives that we have optimal reason to perform. More formally, the principle that I have in mind holds:
>
> PYBO1 A subject, S, is permitted to perform an action, j, if and only if (and because) j-ing is an alternative for S and there is no other action that is an alternative for S that S has more reason to perform.[25]

Multidimensional consequentialists believe PYBO1 is false. In their view, this is not "the most fundamental and least controversial normative principle concerning action". Consider, for instance, a case in which some alternative (Act 1) is morally right to degree .8 and some other alternative (Act 2) is right to degree .9. There are no other relevant differences. So, all things considered, subject S has more reason to perform Act 2 than Act 1. However, it does not follow that S is therefore permitted to perform only Act 2. On the contrary, multidimensional consequentialists claim that the agent is permitted to carry out each alternative with a probability that reflects its degree of rightness (and possibly some other relevant considerations). The higher degree to which the act is right, the higher should

the probability be that the act is performed, everything else being equal. Alternatives that are right to degree 1 should be carried out with probability 1 (unless more than one alternative is right to degree 1, as explained above), and alternatives that are right to degree 0 should be performed with probability 0.[26]

Why should we accept this view? Why not always perform the *best* alternative, which we have *most reason* to perform? I admit that my argument for preferring the view I propose is far from conclusive, but I nevertheless think it is the best argument one can give. Here it is: The fact that Act 1 is morally right to some positive degree should be *properly reflected* in the action performed by a morally motivated and fully informed and rational decision maker, even if Act 2 is right to a somewhat higher (but non-maximal) degree. We can think of this as a type of *fittingness relation* between actions and reasons. My objection to Portmore is thus that it is more fitting to randomize between two alternatives that are somewhat right and somewhat wrong, because if you perform the alternative you have most reason to perform with probability 1, you thereby neglect the complexity of the normative landscape. Other means of doing justice to the complexity of the normative landscape, such as expressing regret about the shortcomings of all available options, would not be "visible" in the option eventually selected by the agent. Such attempts to do justice to the complexity of the normative landscape are merely cosmetic. If you perform Act 2 with probability 1 because this option right to a higher degree than Act 1, your action would not accurately reflect the fact that Act 1 was also right to a fairly high degree. Your *behavior* would have been the same if Act 1 had been right to degree 0.

<1> **3.4. The Repugnant Conclusion from a Multidimensional Perspective**

Let us now discuss how multidimensional consequentialists analyze the Repugnant Conclusion.

The verdict reached by the multidimensional consequentialist will, of course, be sensitive to what he or she takes the relevant moral aspects to be. As mentioned earlier, I believe that each and every *person* counts as a separate moral aspect, rather than the total amount of aggregated wellbeing. I also believe that equality counts as a separate moral aspect. My reasons for preferring this particular version of multidimensional consequentialism are explained in Peterson (2013).[27] The claim that each person counts as a separate moral aspect is, however, easy to misunderstand. It is therefore appropriate to say a bit more about what I mean by this. Strictly speaking, it is not the person itself that counts as a moral aspect, but rather the changes we bring about to each person's wellbeing.

The following example helps to explain what I have in mind: Imagine that you can either preserve the status quo in society, or increase the wellbeing of the worst-off person by one unit by reducing the wellbeing of the best-off person by one unit. On the multidimensional view I propose, there are three moral aspects to consider: (i) equality, (ii) changes to Mr. Worst-Off's wellbeing, and (iii) changes to Mr. Best-Off's wellbeing. With respect to two of these aspects, (i) and (ii), it is right to redistribute wellbeing from Mr. Best-Off to Mr. Worst-Off. That act would lead to more equality, and to more wellbeing for Mr. Worst-Off. However, with respect to aspect (iii), Mr. Best-Off, the act is wrong. This is because Mr. Best-Off will end up with less wellbeing. Needless to say, it is possible that this redistribution of wellbeing is right to a much higher degree than it is wrong, but it is not entirely right.[28] Acts that lead to suboptimal outcomes for some agents are always wrong to some (perhaps very low) degree. Only acts that are optimal for everyone are entirely right.[29]

Let us apply this version of multidimensional consequentialism to the Repugnant Conclusion. According to the view I propose, it is right to some degree to bring about the A-

world instead of the Z-world, but doing so is also wrong to some degree. All things considered, it is neither entirely right nor entirely wrong to bring about the A-world instead of the Z-world. If each person counts as a separate moral aspect, then there will be a large number of people in the Z-world who benefit if we bring about that world. It is better to be alive and lead a life worth living than to not exist at all (with respect to that moral aspect) even if one's wellbeing is low. However, there is also a small group of people who would have been better off in the A-world, but do not exist in the Z-world. If we believe that each individual is a separate moral aspect, then we should count the wellbeing of each person not just in the possible worlds in which she is alive, but also in the worlds in which she *could* have existed. This does not amount to saying that being alive with a wellbeing of zero units (or some other neutral level, depending on how the scale is constructed) is morally equivalent to not existing at all. The claim at stake here is just the following: with respect to one moral aspect, individual *i*, it is morally right to see to it that *i* exists and has a life worth living, compared to not existing at all. This is a claim about rightness with respect to a single moral aspect, not with respect to rightness all-things-considered. Some traditional one-dimensional theories entail that the large quantity of aggregated wellbeing in the Z-world outweighs the much smaller amount of aggregated wellbeing in the A-world, but multidimensional consequentialists can avoid this Repugnant Conclusion by insisting that each person counts as a separate moral aspect.

The multidimensional analysis also has consequences for the Mere Addition Paradox. This holds true no matter whether we believe each person is a separate moral aspect, or insist that it is the total aggregated wellbeing that counts. Imagine, to start with, that equality and total wellbeing are the only moral aspects. To some degree, it will then be right to choose the A+-world instead of the A-world because there is more wellbeing in A+. However, it will also be wrong to some degree to choose the A-world instead of the A+-world, because there

is some inequality in A+. Even if neither of the groups is aware that some other group is not equally well off, the mere fact that not everyone is equally well off makes it wrong to some degree with respect to equality to prefer A+ over A. It can be easily verified that the conclusion will be the same if we believe that each and every person counts as a separate moral aspect.

Parfit explicitly denies that the inequality in A+ makes that world worse than A. He writes: "We cannot plausibly claim that the extra people should never have existed, *merely because, unknown to them, there are other people who are even better off*"[30] Multidimensional consequentialists will, of course, agree with this. However, the multidimensional consequentialist believes that this type of one-dimensional analysis is too blunt. Rather than just concluding that it would not have been better (or right) *all things considered* if the extra people in A+ had not existed, multidimensional consequentialists offer a more nuanced analysis. On their view, it would be right to bring about the A+-world instead of the A-world because there is more (individual) wellbeing in A+. But it would also be wrong to some degree to bring about the A-world instead of the A+-world, because there is some inequality in A+. The all-things-considered verdict is nonbinary. This is a possibility that is overlooked in Parfit's analysis.

<1> **3.5. Arrhenius' Sixth Impossibility Theorem**

We are now in a position to explain how multidimensional consequentialists analyze Arrhenius' impossibility results. Because multidimensional consequentialism is a claim about the rightness and wrongness of acts, and not about the goodness or badness of possible worlds, I will focus on Arrhenius' (2000) sixth impossibility theorem.[31] In my view, this is the most general and interesting formal result in the literature on population ethics.

Arrhenius' sixth impossibility theorem is derived from six normative conditions. The theorem simply says that no moral theory can satisfy all of them. The challenge for the multidimensional consequentialist is, thus, to show that we have reason to reject at least one of the six conditions.

The first condition is the Condition of Separate Satisfiability:

> For any agent and any situation, it is logically possible for her not to act in a morally wrong way.[32]

This condition excludes the existence of moral dilemmas of the traditional type in which all alternatives are entirely wrong. It is thus a rather strong condition that not all one-dimensional consequentialists accept; for instance, authors who believe that cyclical orderings are possible reject this condition.[33]

However, from a multidimensional point of view this condition is less controversial. Arrhenius distinguishes between "wrong" and "not wrong" (and never uses the term "right" in his formal treatment). So multidimensional consequentialists can read "wrong" as "entirely wrong" but still believe that acts that are not entirely wrong may be wrong to a very high degree.[34] Hence, multidimensional consequentialists can account for intuitions about moral dilemmas (of the broad type in which we face a dilemma whenever every alternative is somewhat wrong) without rejecting the Condition of Separate Satisfiability. Multidimensional consequentialists thus have no reason to reject Arrhenius' first condition. There seems to be no minimally plausible version of the theory according to which all alternatives in some situation are entirely wrong.

Arrhenius' second condition is the Normative Weak Quality Addition Condition.

> For any population X, there is at least one perfectly equal population with very high welfare such that if it is wrong in a certain situation to add this population to X, then it is also wrong in the same situation to add any population with very low positive welfare to X, other things being equal.[35]

This condition entails that it could be wrong to add a population with very low wellbeing even if this would bring about more equality. Multidimensional consequentialists, therefore, reject the Normative Weak Quality Addition Condition. To see why, imagine a case in which the mere addition of a population with very low wellbeing brings about more equality. In such a case, it would be right *with respect to equality* to add the population with very low positive wellbeing, but wrong *with respect to equality* to add the perfectly equal population with very high wellbeing mentioned in Arrhenius' formulation of the condition.[36]

Arrhenius' third condition is the Normative Inequality Aversion Condition:

> For any triplet of welfare levels ***A***, ***B***, and ***C***, ***A*** higher than ***B*** and ***B*** higher than ***C***, and for any population A with welfare ***A*** [where ***A*** is a real number], there is a larger population C with welfare ***C*** such that if it is wrong in a certain situation to choose a perfectly equal population B of the same size as A∪C and with welfare ***B***, then it is also wrong in the same situation to choose A∪C, other things being equal.[37]

This condition resembles, but is not identical to, the Pigou-Dalton condition, which is frequently discussed in the literature on economic inequality. The Pigou-Dalton condition holds, roughly, that if we make the income distribution in the world more equal by redistributing income from the rich to the poor, without reducing the overall amount of economic resources, then the new world is at least as good as the initial one.

The Normative Inequality Aversion Condition can be plausibly rejected by the version of multidimensional consequentialism that holds that every person counts as a separate moral aspect. On this view, any extra wellbeing we give to a person, or any extra life we create that is worth living, contributes towards making the act right to a higher degree.[38] So even if it is entirely wrong in some situation to choose a perfectly equal population B, it does not follow that it is entirely wrong in the same situation to choose A∪C, other things being equal. This is because each person in A is better off than the corresponding (but non-identical) people in B. Therefore, A∪C scores better with respect to some moral aspects: the members of A. As explained earlier, it is better *with respect to the A-aspects* to be a live and have a life worth living than to not exist at all. Choosing A∪C may thus be right to some degree at the same time as choosing B is entirely wrong.

Arrhenius' fourth condition is the Normative Non-Extreme Priority Condition:

> There is a number *n* of lives such that for any population X, if it is wrong in a certain situation to choose a population consisting of the X-lives, *n* lives with very high welfare, and a single life with slightly negative welfare, then it is also wrong in the same situation to choose a population consisting of the X-lives and *n*+1 lives with very low positive welfare, other things being equal.[39]

This condition can also be rejected by the sophisticated multidimensional theory mentioned above, according to which the wellbeing experienced by each individual counts as a separate moral aspect. This holds true no matter how large *n* is. In order to see this, note that if we choose the population consisting of the X-lives and *n*+1 lives with very low positive welfare we change the situation to the better for one person: the "single life with

slightly negative welfare". Therefore, this act is not entirely wrong. To some positive degree, which is likely to be tiny if *n* is large enough, the act is right.

From a multidimensional perspective, an act is entirely wrong only if it is suboptimal with respect to every aspect, so no act that leads to an improvement for at least one person will be entirely wrong according to the sophisticated multidimensional theorist who holds that every person counts as a separate aspect.

The upshot of all this is that multidimensional consequentialists can reject at least three of Arrhenius' conditions. The remaining two conditions will not be discussed here. I see no reason to reject them, but I invite the reader to consider them carefully.[40]

To avoid any misunderstanding, I am not claiming that *every* version of multidimensional consequentialism is incompatible with the three conditions discussed above. All I believe to have shown here is that some plausible versions of multidimensional consequentialism are incompatible with some of Arrhenius' conditions. This invites the moral theorist to choose between two options. The first is to accept some version of multidimensional consequentialism and reject Arrhenius' impossibility theorem as irrelevant. The second is to insist that all of Arrhenius' conditions are true and thus accept that no theory of population ethics is satisfactory. I leave it to the reader to adjudicate which option is best.

<1> **References**

Arrhenius, Gustaf. *Future Generations: A Challenge for Moral Theory*, Diss. Uppsala: University Printers, 2000.

Arrhenius, Gustaf. "Population Ethics and Different-Number-Based Imprecision." *Theoria*, 82 (2016), 166-81.

Bader, Ralf M. “Review of Martin Peterson’s The Dimensions of Consequentialism,” *Dialectica*, 68 (2014), 620-25.

Crisp, Roger. “Rightness, Parsimony, and Consequentialism: A Response to Peterson”, *Ethical Theory and Moral Practice* (2016), in press.

McLaughlin, Brian and Bennett, Karen. “Supervenience.” In: *The Stanford Encyclopedia of* Philosophy, edited by Edward N. Zalta, 2014.

Ng, Yew-Kwang. “What Should We Do About Future Generations? The impossibility of Parfit’s Theory X.” *Economics and Philosophy*, 5 (1989), 235-253.

Norcross, Alastair. “Good and bad actions.” *Philosophical Review* 106 (1997):1-34.

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

Peterson, Martin. “Multi-Dimensional Consequentialism”, *Ratio*, 25 (2012), 177-194.

Peterson, Martin. *The Dimensions of Consequentialism*, Cambridge: Cambridge University Press, 2013.

Peterson, Martin. *Reply to Ralf Bader*, *Dialectica*, 68 (2015), 625-29.

Portmore, Douglas W. “Perform Your Best Option. *Journal of Philosophy* 110 (2013), 436-459.

Portmore, Douglas W. “Review of Martin Peterson’s The Dimensions of Consequentialism”, *Journal of Moral Philosophy*, forthcoming

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

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

2 Yew-Kwang Ng, “What should we do about Future Generations? The impossibility of Parfit's Theory X”, Economics and Philosophy 5 (1989), 235-253.

3 Gustaf Arrhenius, *Future Generations: A Challenge for Moral Theory* (Diss. Uppsala: University Printers, 2000) and Gustaf Arrhenius, "Population Ethics and Different-Number-Based Imprecision'" Theoria 82 (2016).

4 See Martin Peterson, *The Dimensions of Consequentialism* (Cambridge: Cambridge University Press, 2013) for a defense of multidimensional consequentialism.

5 By "deontic status" I mean an act's (degree of) rightness or wrongness. The analogy with geometry is discussed in Chapter 1 in Peterson, *Dimensions*. See also Martin Peterson, "Reply to Ralf Bader", *Dialectica* 68 (2015), 625-629.

6 Average utilitarianism is a well-known exception. However, average utilitarianism entails that a world inhabited by a single very happy person is better than the world we are currently living in. This is absurd.

7 Parfit, *Reasons and Persons* , 388.

8 See e.g. Larry S. Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford: Oxford University Press, 2012).

9 See Gustaf Arrhenius, "Population Ethics" for a discussion of this proposal.

10 See Gustaf Arrhenius, *Future Generations.*

11 Only positive numbers are permitted by the root function, so this version of prioritarianism is not compatible with all measures of wellbeing. The log function is a possible alternative that permits for negative levels of wellbeing.

12 Alastair Norcross defends a somewhat similar view he calls scalar consequentialism, according to which acts are better or worse than others, but never right or wrong. That is, goodness and badness vary in degrees, but scalar consequentialists reject the notions of rightness and wrongness altogether. For a comparison of scalar and multidimensional consequentialism, see section 2.5 of Peterson, *Dimensions*, and see Norcross for a general discussion of scalar consequentialism and some related ideas.

13 For a discussion of this point, see section 2.5 in Peterson, *Dimensions* and Roger Crisp, "Rightness, Parsimony, and Consequentialism: A Response to Peterson", *Ethical Theory and Moral Practice*, in press (2016).

14 For a detailed discussion of three different versions of multidimensional consequentialism, see the appendix to Peterson *Dimensions*. The version of the theory summarized here is equivalent to one of the three versions, but note that I remain neutral about which of the three versions is best.

15 This measure clearly presupposes that wellbeing can be measured on a ratio scale. If only interval measurement is possible some other measure has to be chosen.

16 By letting $\bar{w}$ be the mean of $w_1,...,w_n$ the Gini index *g* can be written as

$$g\left(\langle w_1,...,w_n\rangle\right)=\frac{1}{2n^2\bar{w}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|w_i-w_j\right|$$

17 See Peterson, *Dimensions*.

18 See Chapter 3 in Peterson, *Dimensions*.

19 This definition is from Peterson, *Dimensions*, 3. In a footnote on the same page I also point out that the identity function has to be excluded, because an act's deontic status is not an aspect that determines its deontic status.

20 For an overview, see Brian McLaughlin and Karen Bennett, "Supervenience", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (2014).

21 Ralf M. Bader, "Review of Martin Peterson's The Dimensions of Consequentialism", *Dialectica* 68 (2014), 620-625, 620.

22 Douglas W. Portmore, "Review of Martin Peterson's The Dimensions of Consequentialism", *Journal of Moral Philosophy* (forthcoming).

23 Here I focus on the version of multidimensional consequentialism in which equality and total wellbeing and the relevant moral aspects. However, everything I say here holds true *mutatis mutandis* for the version of the theory (discussed below) in which everyone's individual wellbeing-levels count as a separate aspects.

24 Douglas W. Portmore, "Perform Your Best Option", *Journal of Philosophy* 110 (2013), 436-459, 437.

25 Douglas W. Portmore, "Perform Your Best Option", *Journal of Philosophy* 110 (2013), 436.

26 If it is deemed implausible that acts that are right to a very low degree are permitted to be performed with some non-zero probability, then one could add that only acts that are right to a degree above some threshold should be performed with some nonzero probability. In Peterson, *Dimensions*, I did not discuss this possibility, but I now think this might be an attractive response to some potential counterexamples.

27 See Chapter 3 in Peterson, *Dimensions*, for a detailed defense of the claim that persons count as separate aspects.

28 For a discussion of the implications of this view for the Pareto principle, see pp 24-25 in Peterson, *Dimensions* .

29 A possible objection to the suggestion that equality as well as individual wellbeing count as separate moral aspects, proposed by Bader, "Review", is that every plausible measure of equality (such as the Gini index) seems to depend on individual wellbeing levels. It thus seems difficult to hold all wellbeing levels fixed in the population but vary its equality. However, as explained in *The Dimensions of Consequentialism* (on pp. 56–65) the claim that it is individual wellbeing levels that matter is, strictly speaking, more complex. My view is that the best measure of how well things go for an individual at time *t* should be based on what I call "individual moral value". This, in turn, is a measure that depends on two variables: (i) on how high or low the individual's wellbeing is at *t*, and (ii) on whether the individual's wellbeing was higher or lower in the past. The flexibility built into this measure enables us to hold all individual moral values fixed at the same time as we vary the equality in the population.

30 Parfit, *Reasons and Persons*, 425. Italics in original.

31 Note that the multidimensional response could be applied, *mutatis mutandis*, to all of Arrhenius' theorems.

32 Arrhenius, *Future Generations*, 191)

33 Temkin, *Rethinking the Good*, defends a conditional version of this claim.

34 For a multidimensional consequentialist, a moral dilemma is a situation in which no available alternative is entirely right, meaning that every alternative is somewhat (but perhaps not entirely) wrong. A consequence of this account of moral dilemmas is that the total number of dilemmas will be much larger than previously thought. Readers who find this worrisome can think of moral dilemmas in a *narrow sense* as cases in which every alternative is *entirely* wrong.

35 Arrhenius, *Future Generations*, 194.

36 In this discussion I do not wish to take a stand on which aggregation mechanism for all-things-considered rightness we have most reason to accept. Note, however, that if we opt for the maximin account discussed in section 4.3., the counterexample to Arrhenius' condition has to be constructed such that the change in equality affects the minimum level of rightness that determines the acts all-things-considered degree of rightness.

37 Arrhenius, *Future Generations*, 192.

38 As noted earlier, I do not take any stand in this essay on how moral rightness is to be aggregated across different aspects. It is, of course, true that one can construct aggregation principles that preserve the conditions of Arrhenius' theorems, but such maneuvers seem to be *ad hoc* and overly complex.

39 Arrhenius, *Future Generations,* 192.

40 Arrhenius' fifth condition is the Normative Non-Sadism Condition: "If it is wrong in a certain situation to add any number of lives with positive welfare, then it is also wrong in the same situation to add any number of lives with negative welfare, other things being equal." (Arrhenius, *Future Generations,* 194) The sixth of Arrhenius' conditions is the Normative Egalitarian Dominance Condition, which is also a very reasonable condition that all minimally plausible one- and multi-dimensional theories satisfy: "If population A is a perfectly equal population of the same size as population B, and every person in A has higher welfare than every person in B, then, in any situation involving a choice between A and B, it is wrong to choose B, other things being equal." (Arrhenius, *Future Generations,* 191).