GUSTAF ARRHENIUS

THE PARADOXES OF FUTURE GENERATIONS AND NORMATIVE THEORY

1. INTRODUCTION

As the title of this paper indicates, I'm going to discuss what we ought to do in situations where our actions affect future generations. More specifically, I shall focus on the moral problems raised by cases where our actions affect who's going to live, their number and their-well being. I'll start, however, with population axiology. Most discussion in population ethics has concentrated on how to evaluate populations in regard to their goodness, that is, how to order populations by the relations "is better than" and "is as good as". This field has been riddled with "paradoxes" which purport to show that our considered beliefs are inconsistent in cases where the number of people and their welfare varies. Derek Parfit's Mere Addition Paradox is a case in point. The main question of my paper concerns the implication of such axiological paradoxes for normative theories. Do the axiological paradoxes translate into paradoxes for normative theories or will they, as some believe, disappear if we switch to a normative framework?

2. AN AXIOLOGICAL POPULATION PARADOX

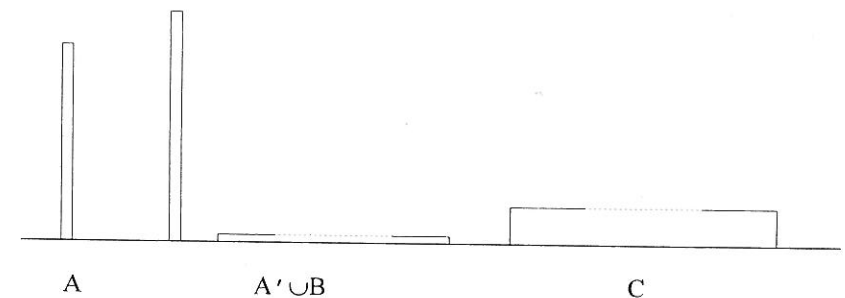Here's a version of Derek Parfit's well-known axiological population paradox, the Mere Addition Paradox: [1]

Diagram 1

201

The diagram above shows four populations: A, A', B, and C. The width of each block represents the number of people, and the height represents their lifetime welfare. All the lives in the above diagram have positive welfare, or, as we also could put it, have lives worth living.[2] A is a population of people with very high welfare, A' is a population of the same size as A but with even higher welfare, B is a much larger population than A and A' but consisting of people with very low positive welfare. C is a population of the same size as A'∪B (that is, the population consisting of all the lives in both A' and B). Everybody in C has very low positive welfare but they are all better off than the people in B. Moreover, there is perfect equality in C and the total and average welfare in C is higher than in A'∪B.

How should we rank these populations? Consider first population A and A'∪B. Since the B-people have lives worth living and the A'-people have even higher welfare than the A-people, many would agree that A'∪B is better than A. Here's a principle that expresses this view:

> *The Dominance Addition Condition*: If population A and B are of the same size and everyone in A has lower welfare than everyone in B, then A is worse than a population consisting of the B-lives and any number of lives with positive welfare, other things being equal.

What about A'∪B and C? Since there is perfect equality in C and higher total and average utility in C as compared to A'∪B, it seems reasonable to claim that C is better than A'∪B. Perhaps the following principle captures our intuition:

> *The Non-Anti Egalitarianism Condition*: A population with perfect equality is better than a population with the same number of people, inequality, and lower average (and thus total) welfare, other things being equal.[3]

Lastly, how should we rank A and C? Derek Parfit has formulated a conclusion that seems to express most people's intuition about the relative value of A and C:

> *The Repugnant Conclusion*: For any perfectly equal population with very high positive welfare, there is a population with very low positive welfare which is better, other things being equal.[4]

As the name indicates, Parfit finds this conclusion unacceptable. To avoid the Repugnant Conclusion, we could claim that A is better than C, a belief expressed by the following principle:

> *The Quality Condition*: There is at least one perfectly equal population with very high positive welfare which is better than any population with very low positive welfare, other things being equal.

By now, we have contradicted ourselves. If C is better than A'∪B, and A'∪B is better than A, then by transitivity of "better than", it follows that C is better than A. But we said that A is better than C, that is, C is not better than A. Hence, these valuations imply a contradiction: C is better than A and C is not better than A.

When faced with an impossibility result like the one just described, a sensible response is to question the principles involved and try to find reasons to reject one of them. One could for example reject the Dominance Addition Condition by claiming that A'∪B is worse than A since there is inequality in the former population but not in the latter, or that although the B-people enjoy positive welfare, their presence makes the population worse since people deserve higher welfare than the B-people enjoy, and so forth. I shall not discuss that route here since I have discussed it at length elsewhere and showed that one can construct impossibility results with logically much weaker conditions that are very hard to reject.[5] For example, the Dominance Addition Condition could be replaced by a condition according to which, roughly, there is at least some number of horrible lives whose addition makes a population worse than an addition of lives with positive welfare. Likewise for the other conditions used above. I'm using the above paradox not because it involves the logically weakest and most compelling condition but because it is easy to present and quite well-known. The paradoxes that involve weaker conditions are much more complicated and would unnecessarily complicate my presentation. The simpler Mere Addition Paradox makes my argument take a clearer form. Moreover, if one can show that it is true, or not true, that the simpler axiological paradoxes translate into normative ones, then it is most probable that this will also hold for the more complicated paradoxes. I will say more about this in the final section.

3. DO THE AXIOLOGICAL IMPOSSIBILITY RESULTS DIRECTLY TRANSLATE INTO NORMATIVE ONES?

It is natural to think that the axiological impossibility results directly translates into impossibility results for normative theories since this part of our morality – our theory of beneficence – is consequentialist in nature and thus must be based on an ordering of outcomes in regard to their "welfarist" goodness. As a matter of fact, most of the population theories presented in the literature explicitly or implicitly include some form of consequentialism as a bridging principle from the axiological level to the normative level. The most common form of consequentialism is act-consequentialism according to which, roughly, an action is right if and only if it maximises the good. More exactly, we shall define act-consequentialism as follows:

> *Act-Consequentialism*: An action is right (obligatory) if and only if its outcome is at least as good as (better than) that of every alternative. An action is wrong if and only if it is not right.[6]

Populations can be seen as outcomes of actions, namely populations that consist of all the lives that are part of the outcomes. Which lives are included in the outcome of an action depends, of course, on what we consider the morally relevant outcome of an action. The three most common answers are the possible world that would be the case if the action were performed, the total future state of the world that would be the case if the action were performed, and the causal consequences of the action.[7] These three views correspond to three types of populations, namely populations that consist of all the past, present and future lives, or all the present and future lives, or all the lives that are causally affected or consequences of the action. Now, there is nothing in the Mere Addition Paradox that rules out that the involved populations are of these types. Moreover, inconsistent evaluations of outcomes are, of course, a devastating problem from a consequentialist perspective, since it will not get off the ground in the first place.

Inconsistent evaluations are not only a problem for pure consequentialists. Such consequentialists assume that all morally relevant factors can be taken into account in the value of outcomes. One might think that certain moral relevant factors cannot be taken into account in such a manner but should be incorporated on the deontic level in terms of actions that are right or wrong by virtue of being of a certain type. Examples are rights, promises, and actions that involve great personal sacrifice for the agent. One may judge actions that involve violations of people's rights or the breaking of promises as wrong, and actions that involve great personal sacrifice as supererogatory, irrespective of how good the consequences of those actions would be. It is not clear, however, that such theories cannot be formulated as extensionally equivalent consequentialist theories since it is possible to incorporate a wide range of non-welfarist values in a consequentialist theory.[8] Fred Feldman's desert adjusted utilitarianism is a case in point.[9] At any rate, some of those critics of consequentialism that take this line do take the consequences of actions into account but they think that there are deontic "constraints" that exclude actions of certain types, or deontic "options" that make certain types of actions permissible. The remaining alternatives are, however, evaluated in a consequentialist manner. They accept what we could call *Ceteris Paribus* Act-Consequentialism:

> *Ceteris Paribus Act-Consequentialism*: Other things being equal, an action is right (obligatory) if and only if its outcome is at least as good as (better than) that of every alternative. An action is wrong if and only if it is not right.

In other words, if a choice situation doesn't involve actions that are right or wrong by virtue of a certain deontic constraint or option, then the normative status of the actions are determined by the value of their respective outcomes. Assuming that the involved deontic constraints and options don't concern the number and the welfare of lives in populations that are outcomes of actions (which is a questionable assumption, however), this view clearly runs into the same problem as pure Act-Consequentialism in respect to the inconsistent evaluations of outcomes.

It is all too hasty to conclude from the above discussion, however, that the axiological impossibility results directly translate into normative ones for two reasons: one can reject the transitivity of "better than" and one can reject consequentialism.

One can be a consequentialist but reject the transitivity of "better than", the latter a move that has become increasingly popular. The axiological population paradoxes presuppose that the relation "better than" is transitive. Some theorists find this a matter of logic, claiming that it is part of the meaning of "better than".[10] One might think otherwise, and argue that the axiological impossibility results actually demonstrate that these relations are not transitive. Among others, Larry Temkin and Stuart Rachels have suggested this.[11] What is attractive with this move is that given non-transitivity of "better than", we can stick to our axiological evaluations without any contradiction. However, as many seem to fear (including Temkin), non-transitivity of "better than" might spell the end for a consequentialist morality and practical reason in general.[12] This needs to be shown, however.

How would non-transitivity of "better than" affect the possibility of a consequentialist population morality? The evaluations involved in the Mere Addition Paradox above and in the other results in the literature exhibit the following structure: Outcome A_1 is better than A_2, which is better than A_3 (, ..., which is better than A_n), which in turn is better than A_1. In such cases, is there an outcome which is at least as good as all alternative outcomes? No, since in cases involving such evaluations it is neither true of any outcome that it is at least as good as all the other outcomes, nor is it true of any outcome that it isn't worse than any other outcome. Consequently, in respect to such cases consequentialism implies, implausibly one might think, that all the available actions are wrong. Consequentialism seems to require some form of acyclicity of the ranking of outcomes in a choice situation.[13] In other words, it doesn't look like abandoning transitivity of "better than" will save consequentialism.

One could claim, however, that what we have here is a plausible interpretation of the Mere Addition Paradox and similar results. In a choice situation involving alternatives like these, we are facing a *moral dilemma*: whatever act we perform we are going to act wrongly.[14] We could claim that the existence of moral dilemmas is part of our moral phenomenology and that it is not surprising that we should face a moral dilemma in situations involving such awesome alternatives as are involved in the Mere Addition Paradox. We could stick to our axiological intuitions and use them to guide us in situations that don't involve intransitive evaluations. Moreover, we would still have something to say about cases like the Mere Addition Paradox: We should avoid putting ourselves in such choice situations since if we do, we will be bound to act wrongly. It other words, intransitivity of "better than" doesn't spell the end for consequentialism and practical reason as some have feared.

Although I find this a very interesting interpretation of the Mere Addition Paradox and the like, I don't find it satisfactory since I'm sceptical about this type of

moral dilemma. As Jan Österberg suggests, any plausible morality is separately satisfiable:

The Condition of Separate Satisfiability: For any agent and any situation, it is logically possible for her not to act morally wrong.[15]

It is reasonable to claim that it should at least be logically possible for a person not to do the wrong thing. Normative theories which imply that there are moral dilemmas in which all the available actions are wrong, imply that there are situations where it is not even a *logical possibility* for an agent to do what morality requires of her. This seems implausible.[16] Consequently, since an adequate morality should be separately satisfiable, the axiological paradoxes challenge the existence of an acceptable consequentialist morality.

We could, however, satisfy the Condition of Separate Satisfiability by slightly reformulating our statement of Consequentialism:

Act-Consequentialism 2: An action is right (obligatory) if and only if its outcome is at least as good as (better than) that of every alternative. An action is wrong if and only if it is not right and there is at least one action that is right.

According to this version of consequentialism, all the involved actions in the Mere Addition Paradox lack normative status, they are neither right nor wrong. This shows that a consequentialist that accept intransitivity of "better than" is not committed to endorsing moral dilemmas. She could perhaps motivate her position by saying that moral theory has nothing to say about cases that involve cyclical evaluations and that lacks a maximal alternative since these are beyond the scope of moral theory.[17] Compare with the theory of quantum mechanics in physics and the impossibility of deriving the next position and velocity of an electron from the measurement of its current position and velocity (often expressed by saying that it is impossible to determine both the position and velocity of an electron).

One might find this satisfactory in the case of the Mere Addition Paradox but consider a version of it where there is also an alternative D available in which everyone has very bad lives. D is clearly worse than A, A'∪B, and C. Hence, it would be clearly better if we choose one of those alternative rather than D and clearly wrong to choose D. Act-Consequentialism 2, however, again yields that all the actions lack normative status, which seems implausible.

We could go on and formulate other versions of consequentialism that avoids the drawbacks of the two version above. Since there is an infinite number of ways we can formulate consequentialist principles, one might hope that there is one that can handle non-transitive evaluations in a way that satisfy our normative intuitions. Actually, there have been a number of promising proposals in this direction – much more elaborated than the simple principles I have discussed above --- in connection with cyclical evaluations in rational choice theory.[18] I shall not discuss these suggestions here, however, since I shall later informally prove that all

of these theories have to be deficient in some respect. Moreover, even if the axiological paradoxes in the end could be shown to be the swansong of consequentialism, this wouldn't suffice to show that they directly translate into paradoxes for all normative theories since we can reject consequentialism. We could instead turn to theories that take welfare into account in a non-consequentialist manner, that is, theories that take welfare into account directly on the normative level instead of taking the route over an ordering of outcomes in regard to their "welfarist" goodness. For example, one could claim that it is always wrong to increase a population with lives not worth living when it is avoidable, or that in the choice between giving a small benefit to one person or a great benefit to many people, one ought to do the latter.[19] Since such theories don't rely on an axiological ordering, the axiological paradoxes don't show that these theories are in trouble. There are also prominent normative theories that take welfare into account in such a different way as compared to consequentialism that it is unclear whether the axiological paradoxes pose a problem for them. Examples are David Gauthier's mutual advantage contractarianism and Richard Arneson's theory of equality of opportunity for welfare.[20] Actually, David Boonin-Vail, who has proposed a non-consequentialist population morality, suggests that whereas there is no satisfying solution to the axiological Mere Addition Paradox (which he calls "the Goodness Paradox"), the normative version of this paradox (which he calls "the Oughtness Paradox") can be solved and this result deprives the former paradox of its moral significance.[21]

One might object that one cannot deprive the axiological Mere Addition Paradox of its relevance in this way since if we cannot claim that, say, A is better than C, then we cannot justify why we ought to choose A when A and C are the alternatives. Moreover, if we claim that A is better than C, and so forth, then we are going to be faced with the axiological Mere Addition Paradox again. Parfit seems to have an argument like this in mind:

All we have shown is that ... we can coherently believe that (4) choosing to produce [C] would be morally worse than choosing to produce A. This does not yet show that (4) is either true, or defensible. Only by defending (4) could we deprive the [axiological] Mere Addition Paradox of its force. ... The most straightforward [way of defending (4)] would be to appeal to (1) [C], as an outcome, would be worse than A. ... But ... this way of defending (4) cannot deprive my paradox of its force. If my arguments were sound, we could not appeal to (1), since this argument would show that [C] could not be worse than A.[22]

I find this objection question-begging since we could claim that A is better than C without committing us to the beliefs that generates the Mere Addition Paradox. Firstly, there are ways of understanding value-concepts such as "better than" in terms of normative ones such as using "A is better than C" as synonymous with "A ought to be chosen in a situation where A and C are the only alternatives", "A is more choice-worthy than C", and the like. From this understanding of "better than", no axiological Mere Addition Paradox follows since we have explained "better than" in terms of normative concepts and these claims are restricted to pairwise

comparisons and there is no plausible analogue to the transitivity of "better than" for these normative concepts (more on this below). Hence, one can claim that "A is better than C" without committing oneself to the beliefs that lead to the axiological Mere Addition Paradox.

Secondly, we could justify our belief that A ought to be chosen when A and C are the alternatives without any appeal to values. We could appeal directly to facts about A and C, for example, that all the people in A enjoy excellent lives whereas all the people in C have lives barely worth living. The belief that these facts give us reason to choose A rather than C when these are the only alternatives don't commit us to the claim that this facts are always decisive since the fact that other alternatives are available, and facts about these alternatives, might act as countervailing reasons.

So I agree with Boonin-Vail's point regarding the significance of a solution to the Normative Mere Addition Paradox. It is another question, however, whether it is easier to solve than the axiological one. This is what I shall discuss in the rest of the paper.

4. TRANSITIVITY PRINCIPLE FOR "OUGHT TO BE CHOSEN"?

Consider the Mere Addition Paradox again and assume that our normative evaluations are, as I think many would agree, as follows (assuming now that the populations in question are outcomes of actions): In the choice between population A and A'∪B, it is permissible to choose A'∪B; in the choice between A'∪B and C, we ought to choose C; in the choice between C and A, we ought to choose A; and in the choice among A, A'∪B, and C, we ought to choose A, and it would be wrong to choose A∪B or C. Have we contradicted ourselves? As a matter of fact, we haven't. As long as we don't add any more restrictions on our normative evaluations, there is no contradiction involved in the above evaluation. This suggests, as Boonin-Vail believes, that evaluations that are contradictory on the axiological level may not be so on the normative level, the reason being that there is no analogue to transitivity on the normative level.[23]

One might think otherwise, however. Gregory Kavka, for example, has suggested the following transitivity principle for moral permissibility: "If it would be permissible to do A if A and B were the alternatives, and would be permissible to do B if B and C were the alternatives, then it is permissible to do A if A, B, and C are [sic] the alternatives."[24] Given this requirement on normative judgements, the above evaluations are inconsistent. Since it is permissible to choose C in the choice between C and A'∪B (if an action is obligatory, it is of course permissible), and permissible to choose A'∪B in the choice between A'∪B and A, it follows from Kavka's principle that it is permissible to choose C in the choice among A, A'∪B, and C. But we said above that in the latter situation, we ought to choose A and it would be wrong to choose C. So we are back in trouble again.

I'm sceptical about Kavka's transitivity principle for moral permissibility, however. Its implication in regard to the Mere Addition Paradox could in itself be

used as an argument against it. Here's another counter-example suggested by Parfit. Suppose that a woman at some point faces the following options:

> P: Having a handicapped child.
> Q: Having no child.

As Parfit points out, "[i]f this child's handicap would not be severe, and we make certain other assumptions, we can plausibly believe that it would be permissible for the woman to choose either P or Q …".[25] Moreover, this evaluation is, arguably, still plausible if P is replaced by the following alternative:

> R: Having the same child, but in a way that would ensure that he wouldn't be handicapped.

Assume now that all of these three alternatives are available to the woman. According to Kavka's transitivity principle, since P is permissible in the choice between P and Q, and since Q is permissible in the choice between Q and R, it follows that P is permissible in the choice among P, Q, and R. But, as Parfit writes, "[w]e can plausibly believe that, if R were also possible, it would be wrong for this woman to choose P rather than R".[26]

One might think that the problem that Kavka's principle runs into demonstrates an important difference between axiological and normative evaluations. It is usually thought that the intrinsic goodness of an outcome doesn't depend on its relation to other outcomes. If an outcome A is good, or better than another outcome B, then we usually think that this holds irrespective of whether A and B are alternative outcomes in some choice situation, or whether there are other alternative outcomes available. As it is often put, the intrinsic value of a state of affairs is independent of its relation to other distinct states of affairs. The normative status of actions, however, depends on what other actions are available in a choice situation. For example, it is permissible and obligatory to inflict harm on somebody if the only other alternative is to inflict even more harm, but if harming is avoidable, then it is wrong.

I don't think, however, that this kind of context (in)dependence is the essential difference between axiological and normative concepts. Although axiological concepts such as "good" and "better than" are not context sensitive as ordinarily understood, there are other ones that are, such as "best" and "worst". So the defining difference between axiological and normative concepts cannot be that the latter but not the former are context sensitive.

Moreover, as I said above, there are ways of understanding value-concepts such as "better than" in terms of normative concepts such as using "A is better than B" as synonymous with "A is more choice-worthy than B" or "A ought to be chosen in a situation where A and B are the alternatives", and the like. If these normative concepts are context sensitive in a way that make them non-transitive, which seems probable, then "better than" will be context sensitive and non-transitive too.

Actually, according to an influential tradition in value analysis, to be valuable is to be a fitting object of a pro-attitude. Value is thus analysed in terms of the stance that one *ought* to take towards an object. As Franz Brentano writes in *The Origin of Our Knowledge of Right and Wrong*: "When we call one good 'better' than another, we mean that ... it is correct to prefer the one good, for its own sake, to the other." Or as A. C. Ewing, the foremost exponent of this reduction of axiological concepts to deontic ones, writes "... we define 'better' as meaning 'what ought to be preferred' ..."[27] Given this analysis, it should come as no surprise that "better than" will be context sensitive and that transitivity fails, since this will happen for the same reasons that we gave above in regard to the explicitly normative concepts. This possible understanding of value-concepts might explain why some theorists have been willing to abandon the transitivity of "better than".[28]

5. NEUTRALISING THE CONTEXT DEPENDENCE OF NORMATIVE STATUS

It could of course be the case that whereas only *some* axiological concepts are context sensitive, this holds for all normative concepts. Be this as it may, this would not suffice to show that Boonin-Vail's conjecture regarding the normative paradoxes is correct since it may still be some normative *relations* among alternative actions that holds irrespective of what other available actions there are in a choice situation. I think this is the case. We can, so to say, neutralise the context dependence of normative status when we formulate normative conditions by partly formulating them in terms of certain features of the choice situation.[29] Consider the following pattern for a normative condition:

> (i) If action h_1 is of type G and action h_2 is of type B, and both h_1 and h_2 are available in a certain choice situation, then h_2 is wrong in this choice situation.

The actions P and R in the example earlier, that is, having a handicapped child (P) or having the same child, but in a way that would ensure that he wouldn't be handicapped (R), fit this pattern, as the quote from Parfit suggests. If P and R are both available actions in some situation, then P is the wrong choice.[30] Loosely speaking, we can say that (i) is a normative analogue of "better than". We shall formulate the following condition along these lines:

> *The Normative Egalitarian Dominance Condition*: 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.

This condition is, I think, as plausible as its axiological counterpart. The *ceteris paribus* condition involved here is a natural extension of the *ceteris paribus* condition used in the discussion of different axiologies. There are neither any

constraints (for example, promise-keeping) nor options (for example, great personal sacrifice for the agent which is beyond the call of duty), nor any non-welfarist values in the outcomes (for example, cultural diversity) that give us a reason to (not) choose one or the other of the involved actions. The only reasons for choosing one or the other of the involved actions arise from the welfare of the lives in the involved populations.[31] Consider a situation where you could, at no cost to yourself (you might even be among the beneficiaries), and without violating any other duties or compromising any other values, choose an outcome in which everybody is equally well off, and better off as compared to another outcome involving the same number of people. Surely it would be wrong to choose the latter outcome in this situation.

We could formulate normative versions of all the axiological adequacy conditions involved in the Mere Addition Paradox using this pattern (i). For the result we shall prove, however, it suffices to use the following logically weaker construction which, loosely speaking, can be said to be a normative analogue of "at least as good as":

> (ii) If action h_1 is of type G and action h_2 is of type B, and both h_1 and h_2 are available in a certain choice situation, and h_1 is wrong in this choice situation, then h_2 is also wrong in this choice situation.

Assume that next Sunday you can help either John or Krister with their gardening, and that they both need your help equally as much. You haven't promised either one of them your help, it is neither John's or Krister's birthday, nor any other circumstances that speaks in favour of helping one of them rather than the other. Now, it is reasonable to claim that in a situation involving these two alternatives, if it would be wrong of you to help Krister, then it would also be wrong of you to help John. It could be wrong of you to help Krister if you have promised your elderly aunt to help her next Sunday with the much needed gardening at her house (assuming that the involved acts are mutually exclusive). If that were the case, however, then it would also be wrong of you to help John.

Here's a normative version of the Non-Anti Egalitarianism condition formulated in this way:

> *The Normative Non-Anti Egalitarianism Condition*: For any perfectly equal population A, and any population B of the same size as A but with inequality and lower average (and thus total) welfare, if it is wrong in a certain situation to choose A, then it is also wrong to choose B, other things being equal.

Apart from the Normative Egalitarian Dominance Condition, we shall formulate all the normative adequacy conditions used in the normative version of the Mere Addition Paradox below along the lines of pattern (ii). Here are the two remaining conditions, the Normative Dominance Addition Condition and the Normative Quality Condition:

The Normative Dominance Addition Condition: If population A and B are of the same size and everyone in A has lower welfare than in B, and if it is wrong in a certain situation to choose a population consisting of the B-lives and any number of people with positive welfare, then it is also wrong in this situation to choose population A, other things being equal.

The Normative Quality Condition: There is at least one perfectly equal population with very high positive welfare such that if it is wrong in a certain situation to choose that population, then it is wrong in the same situation to choose any population with very low positive welfare, other things being equal.

In addition, we shall require that a plausible population morality is separately satisfiable.

6. AN INFORMAL DEMONSTRATION OF A NORMATIVE MERE ADDITION PARADOX

We shall informally show that the following theorem is true:

The Impossibility Theorem: There is no separately satisfiable population morality which satisfies the normative version of the Quality Condition, the Non-Anti Egalitarianism Condition, the Egalitarian Dominance Condition, and the Dominance Addition Condition.
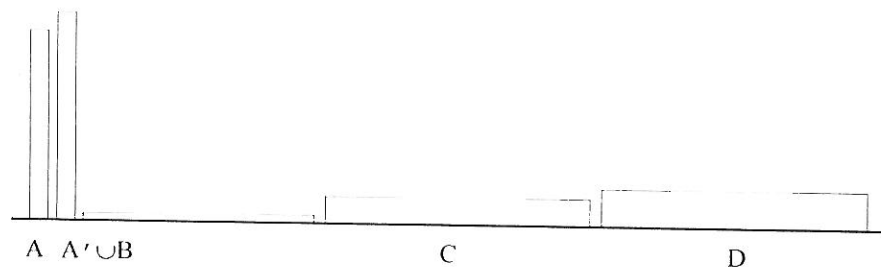
Consider the following situation:

Diagram 2

Here we have the same situation as in the axiological Mere Addition Paradox but for population D. Assume that these populations are possible outcomes of actions. Assume further that the only actions available to a certain individual or group of

individuals are the actions with either population A, A'∪B, C or D as outcome. So these are the only actions available in the choice situation. Again, these populations could consist of all the future lives, or all the lives in some part of the future, that would be the case if the respective action were performed.

In the diagram above, D is a perfectly equal population, and every person in D has higher welfare than every person in C. Consequently, it follows from the Normative Egalitarian Dominance Condition that all the actions that have population C as outcome are wrong.

C is a perfectly equal population of the same size as A'∪B but with higher average welfare. Thus, since the actions with C as an outcome are wrong, the Normative Non-Anti-Egalitarianism Condition implies that all the actions with population A'∪B as outcome are wrong.

The A'-people are better off than the A-people, and the B-people have positive welfare. Thus, since the actions with population A'∪B as outcome are wrong, it follows that all the actions with population A as outcome are wrong.

According to the normative version of the Quality Condition, there is at least one perfectly equal population with very high positive welfare such that if it is wrong in a certain situation to choose that population, then it is wrong in the same situation to choose any population with very low positive welfare. We can assume that A is such a high welfare population. Since the actions with A as outcome are wrong, and since D is a population with very low welfare, it follows that all the actions with population D as outcome are wrong.

Since the actions with either population A, A'∪B, C or D as outcome are all wrong, and these are all of the available actions in the situation, it follows that all the available actions in this situation are wrong. However, according to the Condition of Separate Satisfiability, this cannot be the case. Hence, the assumption that there is a morality which satisfies all the adequacy conditions entails a contradiction. Thus, the impossibility theorem must be true. In other words, we have showed that a version of the Mere Addition paradox can be reproduced on the normative level.

7. DISCUSSION

One might object to my derivation of a normative Mere Addition Paradox in the following way: "Clearly, one of the normative conditions must be false since we don't believe that all actions are wrong in the Mere Addition Paradox. Moreover, the Dominance Addition Condition is a good candidate for falsehood. So there is no paradox here, just implausible conditions that are not jointly satisfiable."

Of course, one of the conditions must be false since they cannot all be true. However, this objection misses the point of my argument. The point of the above demonstrations is that we now know that we have to reject one of the evaluative or normative conditions or accept the existence of moral dilemmas. We cannot get rid of the paradoxes of population ethics by giving up a formal condition like the transitivity of "better than" or by rejecting consequentialism and switching to a

normative framework. Again, I used the Mere Addition Paradox as an example not because it involves the logically weakest and most compelling condition but because it is easy to present and quite well-known.

It is true that the Dominance Addition Condition is a questionable condition but it was the weakest link already in the axiological case. The reasons against the axiological version of this condition work as well as against it in its normative guise. It might be that we have some further reasons against the normative version of he Dominance Addition Condition that has to do with it not being enough context sensitive. Be this as it may, the point is that we now have a structure for employing the same strategy as we have done in the axiological case: When in doubt over the truth of a condition, try to replace it with a more compelling condition. Now, we can replace the Dominance Addition Condition with two conditions which are, I surmise, as hard to deny as the Normative Egalitarian Dominance Condition. Here are rough formulations of these conditions:[12]

> *The Normative Weak Non-Sadism Condition (roughly)*: There is a number of horrible lives such that if we could instead add some number of people with positive welfare, then it would be wrong to add the horrible lives, other things being equal.

> *The Normative Non-Extreme Priority Condition: (roughly)*: There is some very large benefit such that in a choice between giving this benefit to a vast number of people and giving a very small benefit to only one person, it would be wrong to do the latter, other things being equal.

As with the Normative Egalitarian Dominance Condition, it is hard to imagine any alternatives in a choice situation that would make us doubt the truth of these conditions. Moreover, one can show that there is no separately satisfiable moral theory that jointly satisfies these conditions, the Normative Egalitarian Dominance Condition, a version of the Quality Condition, and a very compelling egalitarian condition (logically weaker than the Non-Anti Egalitarianism Condition).[33]

So what do these results show? We are back to square one, or we are even worse off. We cannot exorcise the paradoxes of population ethics by giving up some formal condition like the transitivity of "better than" or by rejecting consequentialism and switching to a normative framework. The paradoxes are a problem for any moral theory, consequentialist and non-consequentialist alike. We have to find a reason to abandon one of the moral adequacy conditions or accept all things considered moral dilemmas. Or become moral sceptics and accept that our considered moral beliefs are not justified.

NOTES

[1] For a formal proof with slightly weaker conditions, see Arrhenius (2000), section 10.6. It should be stressed that the above paradox is not identical to Parfit's (1984), pp. 419ff, Mere Addition Paradox since it involves stronger assumptions. This version is similar to the one presented in Ng (1989), p. 240. A formal proof with slightly stronger assumptions than Ng's can be found in Blackorby and Donaldson (1991).

[2] We shall say that a life has neutral welfare if and only if it has the same welfare as a life without any good or bad welfare components, and that a life has positive (negative) welfare if and only if it has higher (lower) welfare than a life with neutral welfare. A hedonist, for example, would typically say that pain is bad and pleasure is good for a person, and that a life without any pain and pleasure has neutral welfare. This definition can be combined with other welfarist axiologies, such as desire and objective list theories. A number of alternative definitions of a life with positive (negative, neutral) welfare figures in the literature. For a discussion of these, see Arrhenius (2000) and Broome (1993). For a discussion of this issue in connection to the Repugnant Conclusion, see Fehige (1998), Tännsjö (1998) and Arrhenius (2000). Also Cf. Parfit (1984), p. 358.

[3] See Ng (1989), p. 238 for a similar principle. It seems to be unanimously agreed in the literature that inequality aversion of some kind is a prerequisite for an acceptable population axiology. Ng (p. 239, fn. 4), states that "Non-Antiegalitarianism is extremely compelling" and Carlson (1998), p. 288; claims that "[r]ejecting NAE [the Non-Anti Egalitarianism Principle] is ... a very unattractive option". Blackorby, Bossert and Donaldson (1997), p. 13, hold that "weak inequality aversion is satisfied by all ethically attractive ... principles". Fehige (1998), p. 12, rhetorically asks "...if one world has more utility than the other and distributes it equally, whereas the other doesn't, then how can it fail to be better?" See also Sider (1991), p. 270, fn. 10. There are, however, reasons for rejecting the Non-Anti Egalitarianism principle. See Arrhenius (2000), section 6.2.

[4] See Parfit (1984), p. 388. My formulation is more general than Parfit's and he doesn't demand that the people with very high welfare are equally well off.

[5] Arrhenius (2000).

[6] Cf. Carlson (1995), p. 13. My formulation differs slightly from Carlson's (the brief discussion of consequentialism below draws heavily on Carlson's work). The definition of consequentialism that we have suggested has counter-intuitive implications in cases involving outcomes that are incommensurable in value. I don't think, however, that consequentialism necessarily presupposes a complete ordering of the outcomes in a choice situation. See Carlson (1995), p. 25, fn. 48, for some suggested revisions of the definition of consequentialism that can handle incommensurable outcomes. There are other versions of consequentialism apart from act-consequentialism, such as rule-consequentialism according to which an act is right if and only if it can be subsumed under a rule whose general acceptance (or general implementation) would give the best result, which I'm not going to discuss. It should be clear, however, that the discussion below applies, *mutatis mutandis*, equally well to rule-consequentialism.

[7] See Carlson (1995), pp. 10-12, and ch. 4, for an extensive discussion of the morally relevant outcome of an action.

[8] For a discussion of whether any morality can be formulated as a consequentialist morality, see Carlson (1995), Danielsson (1988), Vallentyne (1988), and Bykvist (1996).

[9] Feldman (1997).

[10] Broome (1991), p. 11.

[11] Rachels (1998, 2001), Temkin (1987, 1996).

[12] Temkin suggests that arguments to the effect that "better than" is intransitive "are [perhaps] best interpreted as a frontal assault on the intelligibility of consequentialist reasoning about morality and rationality. Such reasoning may need to be severely limited, if not jettisoned altogether." (Temkin 1987, p. 186, fn. 49). Elsewhere, he considers that intransitivity "...opens the possibility that there would be no rational basis for choosing between virtually any alternatives" (Temkin 1996, p. 209).

[13] The acyclicity I have in mind is that if A_1 is better than A_2, A_2 is better than A_3, ..., A_{n-1} is better than A_n, then A_1 is not worse than A_n. Cf. Sen (1970), p. 47. For a discussion of cyclical evaluations, see Danielsson (1996), Carlson (1996), and Rabinowicz (2000).

[14] Following Vallentyne (1988), we could call a dilemma of the above mentioned type a "prohibition

dilemma". There are also "obligation dilemmas", that is, situations where more than one action is obligatory.

[15] See Österberg (1988), pp. 127, 145-6. My formulation is weaker than Österberg's which is formulated in terms of the possibility of an agent to act morally right.

[16] Österberg (1988), p. 146, suggests an interesting argument to the effect that his version of the Condition of Separate Satisfiability is entailed by the common idea that "ought" implies "can". His argument rests, however, on the questionable principle that if an agent ought to do each conjunct, he ought to do the conjunction. Moreover, from his argument it only follows that at least one action is not wrong, not, as Österberg suggests, that at least one action is right (it could be that some actions lack normative status).

[17] Notice that there are cases involving cyclical evaluations where there still will be right actions according to Act-Consequentialism 2. Consider a case where B is better than C, C is better than D, D is better than B, and A is better than B, C and D. In this case, the action with A as outcome is the right action according to Act-Consequentialism 2.

[18] Carlson (1996).

[19] Jan Narveson's theory is an example of a more developed effort in this direction. Narveson suggest the following principle: (1) New additions to population ought not to be made at the expense of those who otherwise exist, even if there would be a net increment in total utility considered in person-independent terms. But (2) new additions ought to be made if the benefit to all, *excluding* the newcomer, would exceed the cost to all, *including* him or her, as compared with the net benefit of any alternatives which don't add to population [i.e., if the benefit minus the cost would exceed the net benefit of any alternative]. Finally, (3) within those limits, the decision whether to add to population is up to the individuals involved in its production, provided that if they have a choice of which child to produce they produce the happier one, other things being equal. Narveson (1978), pp. 55-6. Cf. Narveson (1967, 1973).

[20] See Gauthier (1986), p. 299 and Heath (1997), for the former kind of theory, and Arneson (1989) for the latter. I discuss Gauthier's and Heath's suggestion at length in Arrhenius (1999).

[21] Boonin-Vail (1996), pp. 279-80, 307.

[22] Parfit (1996), p. 312.

[23] Cf. Boonin-Vail (1996), p. 285.

[24] Kavka (1982), p. 100, fn. 16. Boonin-Vail (1996), p. 283, discusses a similar principle.

[25] Parfit (1996), p. 311.

[26] Parfit (1996), p. 311.

[27] Brentano (1969 [1889]), p. 26; Ewing (1959), p. 85. Ewing writes further "Bad, I should say, is just what ought to be the object of an unfavourable attitude, as good is what ought to be the object of a favourable." Cf. Scanlon (1998), p. 97. An extensive discussion of this subject can be found in Rabinowicz and Ronnow-Rasmussen (2002).

[28] For a discussion of some other differences between these kinds of concepts, see von Wright (1963), ch. 1, sect. 4, and Danielsson (1999).

[29] We are drawing on a suggestion made by Sen (1995), p. 5, in response to certain criticisms of Arrow's impossibility theorem.

[30] Notice that we could formulate an analogous condition for "best": If both P and R are available, then P cannot be the best action in this choice situation.

[31] One way in which other things cannot be equal is that at least one of the actions in a choice situation will be an omission and we might think that this is of relevance for an action's moral status. Even if this is sometimes true, there are clear cases where the fact that an action is an omission doesn't affect its deontic status such as when the consequence of the omission is much worse than that of the other alternatives. Consequently, we could restrict the conditions presented above to only concern comparisons between "active" actions and not to cover omissions and then include in the cases we consider a very bad "omission alternative" that is forbidden anyway.

[32] For exact formulations of these conditions, see Arrhenius (2000), section 11.5.

[33] See Arrhenius (2000), section 11.5-11.6.

REFERENCES

Arneson, R. "Equality of Opportunity for Welfare". *Philosophical Studies* 5 (1989): 77-93.

Arrhenius G. *Future Generations: A Challenge for Moral Theory*, F.D.Diss.,Uppsala: University Printers, 2000a.

Arrhenius, G. "Mutual Advantage Contractarianism and Future Generations", *Theoria*, Vol. LXV, part 1, (1999).

Blackorby, C., and Donaldson, D. "Normative Population Theory: A Comment". *Social Choice and Welfare* 8 (1991): 261-7.

Blackorby, C., Bossert, W., and Donaldson, D. "Critical-Level Utilitarianism and the Population-Ethics Dilemma". *Economics and Philosophy*, 13 (1997): 197-230.

Boonin-Vail, D. "Don't Stop Thinking About Tomorrow: Two Paradoxes About Duties to Future Generations". *Philosophy and Public Affairs* 25 (No. 4) (1996): 267-307.

Brentano, Franz, *The Origin of Our Knowledge of Right and Wrong*, London and Henley: Routledge & Kegan Paul, 1969 [1889].

Broome, J. *Weighing Goods*, Oxford: Basil Blackwell, 1991.

Broome, J. *Ethics out of Economics*, Cambridge: Cambridge University Press, 1999.

Bykvist, K. "Utilitarian Deontologies?", pp. 1-16 in W. Rabinowicz (ed.), *Preference and Value: Preferentialism in Ethics*, Studies in Philosophy, Dept. of Philosophy, Lund University, Vol. 1 (1996).

Carlson, E. *Consequentialism Reconsidered*. Dordrecht: Kluwer Academic Publishers, 1995.

Carlson, E. "Mere Addition and Two Trilemmas of Population Ethics". *Economics and Philosophy* 14 (1998): 283-306.

Carlson, E. "Cyclical Preferences and Rational Choice", *Theoria* 62 (1996): 144-160.

Danielsson, S. "Konsekvensetikens gränser" in *Filosofiska Utredningar*, Stockholm: Thales, 1988.

Danielson, S. "The Refutation of Cyclic Evaluations". *Theoria* 62 (1996): 161-168.

Danielson, S. "The Norm-Value Distinction in 1963" pp. 325-330 in G. Meggle (ed.) *Actions, Norms, Values: Discussions with Georg Henrik von Wright*, Berlin: de Gruyter, 1999.

Ewing, A. C. *Second Thoughts in Moral Philosophy*, London: Routledge and Kegan Paul, 1959.

Fehige, C. "A Pareto Principle for Possible People", pp. 509-43 in C. Fehige and U. Wessels (eds.), *Preferences*. Berlin: de Gruyter, 1998.

Feldman, F. *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*. Cambridge: Cambridge University Press, 1997.

Gauthier, D. *Morals By Agreement*. Oxford: Clarendon Press, 1986.

Heath, J. "Intergenerational Cooperation and Distributive Justice". *Canadian Journal of Philosophy* 27 (No. 3) (1997): 361-76.

Kavka, G. S. "The Paradox of Future Individuals". *Philosophy and Public Affairs* 11 (No. 2) (1982): 93-112.

Narveson, J. "Utilitarianism and New Generations". *Mind* 76 (1967): 62-72.

Narveson, J. "Moral Problems of Population". *The Monist* 57 (1973): 62-86.

Narveson, J. "Future People and Us", pp. 38-60 in R. I. Sikora and B. Barry (eds.), *Obligations to Future Generations*, Philadelphia: Temple University Press, 1978.

Ng, Y-K. "What Should We Do About Future Generations? Impossibility of Parfit's Theory X", *Economics and Philosophy* 5 (2) (1989): 235-253.

Österberg, J. *Self and Others*. Dordrecht: Kluwer Academic Publishers, 1988.

Parfit, D. *Reasons and Persons*. Oxford: Clarendon Press, [1991] 1984.

Parfit, D. "Acts and Outcomes: A Response to Boonin-Vail". *Philosophy and Public Affairs* 25 (No. 4) (1996): 308-317.

Rabinowicz, W. "Money Pump with Foresight", pp. 201-234 in W. Rabinowicz (ed.) *Value and Choice: Some Common Themes in Decision Theory and Moral Philosophy*, Lund: Lund Philosophy Reports, Vol. 1 (2000).

Rabinowicz W., and Ronnow-Rasmussen T. "The Strike of the Demon: On Fitting Pro-attitudes and Value", mimeo, Dept. of Philosophy, Lund University, 2002.

Rachels, S. "Counterexamples to the Transitivity of Better Than". *Australasian Journal of Philosophy* 75 (no.1) (1998): 71-83, March 1998.

Rachels, S. "A Set of Solutions to Parfit's Problems". *Noûs* 35 (no.2) (2001).

Scanlon, T. M. *What We Owe to Each Other*. Cambridge, Mass.: Harvard UP, 1998.

Sen, A. *Collective Choice and Social Welfare*, Mathematical Economics Texts 5, 1970.

Sen, A. "Rationality and Social Choice". *American Economic Review* 85 No. 1 (1995): 1-25.

Sider, T. R. "Might Theory X Be a Theory of Diminishing Marginal Value?". *Analysis* 51 (no.4) (1991): 265-271.

Tännsjö, T. *Hedonistic Utilitarianism*. Edinburgh: Edinburgh University Press,

Temkin, L. S. "Intransitivity and the Mere Addition Paradox". *Philosophy and Public Affairs* Vol. 16 (No. 2) (1987): 138-87.

Temkin, L. S. "A Continuum Argument for Intransitivity". *Philosophy and Public Affairs* 25 (No. 3) (1996): 175-210.

Vallentyne, P. "Gimmicky Representations of Moral Theories". *Metaphilosphy*, Vol. 1 (no.34) (1998).

von Wright, G. H. *The Varieties of Goodness*. Bristol: Thoemmes Press [1993], 1963.