# Population Paradoxes without Transitivity

Gustaf Arrhenius

The Oxford Handbook of Population Ethics
*Edited by Gustaf Arrhenius, Krister Bykvist, Tim Campbell, and Elizabeth Finneron-Burns*

Print Publication Date: Mar 2022 Subject: Philosophy, Moral Philosophy
Online Publication Date: Apr 2022 DOI: 10.1093/oxfordhb/9780190907686.013.11

## Abstract and Keywords

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” and impossibility theorems which purport to show that our considered beliefs are inconsistent in cases where the number of people and their welfare varies. 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. However, it is all too hasty to conclude 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. The latter idea is that we should turn to theories that take welfare into account in a nonconsequentialist 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 goodness. Since transitivity of “better than” plays an important role in the axiological paradoxes, and since there is no convincing analog to transitivity on the normative level, the paradoxes will not reappear on the normative level, or so it is claimed. This chapter shows that this claim is unfortunately false. As a corollary, it shows how we can prove the axiological impossibility theorems without an appeal to the transitivity of “better than.”

Keywords: population axiology, population paradoxes, transitivity, normative population theory, impossibility theorems, Mere Addition Paradox, consequentialism

# 1. Introduction

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 and impossibility theorems which show that our considered beliefs are inconsistent in cases where the number of people and their welfare varies. It is natural to think that the axiological im-

possibility results directly translate 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. However, it is all too hasty to conclude 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.

The former idea is that since transitivity of "better than" is presupposed in the axiological impossibility theorems, we can avoid these results by simply giving up on transitivity. What is attractive with this move is that given nontransitivity of "better than" (and thus "at least as good as"), we can stick to our axiological evaluations without any contradiction. We can still, it is hoped, get enough structure in the value ordering of populations to guide our choices via some form of consequentialism.

The latter idea is that we should turn to theories that take welfare into account in a nonconsequentialist 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 goodness. Since transitivity plays an important role in the axiological theorems, and since there is no convincing analog to transitivity on the normative (p. 182) level, the theorems will not reappear on the normative level, or so it has been claimed, as we shall see below.

In this chapter, I'll show that this latter claim is unfortunately false. As a corollary, I'll show how we can prove the axiological impossibility theorems without an appeal to the transitivity of "better than." Hence, the idea that we can avoid impossibility theorems in population axiology by just giving up on transitivity of "better than" is unfortunately also false.[1]

# 2. An Axiological Population Paradox

Figure 8.1 illustrates a version of Derek Parfit's well-known axiological population paradox, the Mere Addition Paradox.[2] The diagram 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 people in the above diagram enjoy positive welfare, or, as we also could put it, have lives worth living.[3] 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 (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.

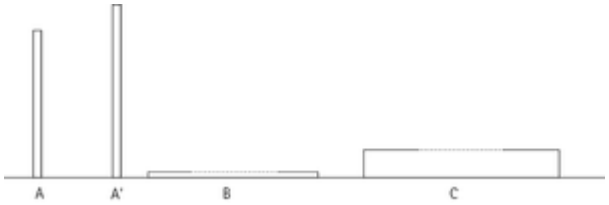


*Figure 8.1* Parfit's Mere Addition Paradox

How should we rank these populations? Consider first populations 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:

> *Dominance Addition*: If populations 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.[4]

(p. 183) What about A′∪B and C? Since there is perfect equality in C and higher total and average welfare 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:

> *Non-Anti Egalitarianism*: 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.[5]

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.[6]

As its 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:

> *Quality*: 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.[7]

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 Dominance Addition 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.[8] For example, Dominance Addition 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 impossibility result not because it involves the logically weakest and most compelling conditions but because it is easy to present and quite well known. The theorems that involve weaker conditions are much more complicated and would unnecessarily complicate the presentation. The simpler Mere Addition Paradox makes my argument take a clearer form. (p. 184) Moreover, if one can show that it is true, or not true, that the simpler axiological impossibility results translate into normative ones, or can be reformulated without transitivity, then this will also hold for the more complicated ones. 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 translate into impossibility results for normative theories since, one might argue, 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 maximizes 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.[9]

Populations can be 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 obtain if the action were performed, the total future state of the world that would obtain if the action were performed, and the causal consequences of the action.[10] 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 causal 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 morally 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 (p. 185) violations of 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.[11] Some of those critics of consequentialism that take this line do take the consequences of actions into account, however, but they think that there are deontic "constraints" that exclude actions of certain types, or deontic "options" that make certain types of actions permissible, irrespective of their consequences. 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. Other things being equal, 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 is 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 in the impossibility cases (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 this 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," a move that has become increasingly popular. The axiological population paradoxes in the literature have presupposed 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."[12] 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 sug-

gested this.[13] What is attractive about this move is that given nontransitivity of "better than," we can stick to our axiological evaluations without any contradiction. The evaluations involved in the Mere Addition Paradox above and in the other results in the literature exhibit the following structure (substituting outcomes for populations): 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 $A_1$.[14] With transitivity, we can derive a contradiction from these evaluations, such as that $A_1$ is both better and worse than $A_n$, but without it, we cannot. So perhaps there is a possibility for a consequentialist to eschew the impossibility theorems by going nontransitive in the value ordering.[15]

However, a problem with this move is that in cases such as the above involving cyclical evaluations, there is no outcome which is at least as good as all alternative outcomes: it is neither true of any outcome that it is at least as good as all the other outcomes, nor is it (p. 186) true of any outcome that it isn't worse than any other outcome. Consequently, in respect to such cases consequentialism implies, implausibly, that all the available actions are wrong (since an action is right if and only if its outcome is at least as good as that of every alternative, and otherwise wrong). Consequentialism requires some form of acyclicity of the ranking of outcomes in a choice situation to avoid such results.[16] In other words, it doesn't look like abandoning transitivity of "better than" is sufficient to save a consequentialist population morality.

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.[17] 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 grand 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 nontransitive 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. In other words, nontransitivity (and non-acyclicity) of "better than" doesn't need to 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 skeptical about this type of moral dilemma. As Jan Österberg suggests, any plausible morality is separately satisfiable:

> *Separate Satisfiability*: For any agent and any situation, there is an action such that if the agent were to perform this action, then her action would not be morally wrong.[18]

It is reasonable to claim, I think, that it should at least be logically possible for a person not to do the wrong thing. Normative theories which violate Separate Satisfiability by implying that there are situations in which all the available actions are wrong, or would be wrong were they performed, imply that there are situations where it is not even a *logical*

*possibility* for an agent to do what the theory requires of her.[19] This seems implausible.[20] Consequently, since an adequate morality should be separately satisfiable, the axiological impossibility theorems challenge the existence of an acceptable consequentialist morality.

We could, however, satisfy Separate Satisfiability by slightly reformulating our statement of Act-Consequentialism:

> *Incomplete 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 and there is at least one action that is right.

(p. 187) 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 who accepts nontransitivity of "better than" is not committed to endorsing moral dilemmas. Instead, there is a gap in the normative ordering of actions. She could perhaps motivate her position by saying that moral theory has nothing to say about cases that involve cyclical evaluations and that lack a maximal alternative, and that such cases are beyond the scope of moral theory.[21]

One might reasonably consider a theory that does not provide any guidance in situations such as the Mere Addition Paradox unsatisfactory. Moreover, even if one finds this acceptable in the case of the Mere Addition Paradox, there are other similar situations in which this is clearly unsatisfactory. Consider a version of this paradox in which 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 were to choose one of those alternatives rather than D and clearly wrong to choose D. Incomplete Act-Consequentialism, however, again yields that all the actions lack normative status, which is clearly implausible.

We could go on and formulate other versions of consequentialism that avoid the drawbacks of the two versions above. Since there are so many possible ways we can formulate consequentialist principles, one might hope that there is one that can handle nontransitive evaluations in a way that satisfies 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.[22] 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 impossibility results 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 nonconsequentialist 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.[23] 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.[24]

(p. 188) Actually, David Boonin, who has proposed a nonconsequentialist 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.[25]

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.[26]

I find this objection question-begging since we could claim that A is better than C without committing us to the beliefs that generate the Mere Addition Paradox.

First, 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 choiceworthy 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 analog 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.

Second, 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 these facts are always decisive since the fact that other alternatives are available, and facts about these alternatives, might provide countervailing reasons.

I therefore agree with Boonin'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, an issue to which we shall now turn.

## (p. 189) 4. A 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 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 analog to transitivity on the normative level.[27]

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."[28] 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 skeptical 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 is 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."[29] 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 she wouldn't be handicapped.

(p. 190) 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."[30]

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 intrinsically good, or intrinsically 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." Thus 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 choiceworthy 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 nontransitive, which seems probable, then "better than" will be context-sensitive and nontransitive 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 analyzed 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 fore-

most exponent of this reduction of axiological concepts to deontic ones, writes: "we define 'better' as meaning 'what ought to be preferred'."[31] 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."[32]

## (p. 191) 5. Neutralizing 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's conjecture regarding the normative paradoxes is correct, since it may still be some normative *relations* among alternative actions that hold irrespective of what other available actions there are in a choice situation. I think this is the case. We can, so to say, neutralize the context dependence of normative status when we formulate normative conditions by partly formulating them in terms of certain features of the choice situation.[33] 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 she 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.[34] Loosely speaking, we can say that (i) is a normative analog of "better than." We shall formulate the following condition along these lines:

> *Normative Egalitarian Dominance*: 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 (and possibly other alternatives), 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, desert) 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.[35] 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 equal-

ly 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.[36]

(p. 192) 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 analog 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 George or Tony with their gardening, and that they both need your help equally as much and would derive equally as much satisfaction from getting your help. You haven't promised either one of them your help, it is neither George's nor Tony's birthday, nor are there any other circumstances, such as side-effects on other people's well-being, that speak in favor of helping one of them rather than the other.[37] Now, it is reasonable to claim that in a situation involving these two alternatives, if it would be wrong of you to help Tony, then it would also be wrong of you to help George. It could be wrong of you to help Tony 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 George.

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

> *Normative Non-Anti Egalitarianism*: 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 Normative Egalitarian Dominance, 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, Normative Dominance Addition and Normative Quality:

> *Normative Dominance Addition*: If populations A and B are of the same size and everyone in A has lower welfare than everyone 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.

> *Normative Quality*: 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.

(p. 193) For the purpose of informally proving the theorem, it will be useful to state some terminology more exactly. We shall say that a *population morality* at least assigns the normative status wrong to some actions in some possible choice situations. When we say that an action is "wrong" in a choice situation, then that is short for "wrong or would be wrong if performed."[38] A *choice situation* is a set of at least two mutually incompatible actions available to a certain individual or group of individuals at a certain time.[39]

# 6. An Informal Demonstration of a Normative Mere Addition Paradox

We shall informally show that the following theorem is true:

> *The Normative Impossibility Theorem*: There is no separately satisfiable population morality which satisfies the normative versions of Quality, Non-Anti Egalitarianism, Egalitarian Dominance, and Dominance Addition.

We shall show that the contrary assumption leads to a contradiction. Consider the situation shown in Figure 8.2. 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. These are therefore 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 exist if the respective action were performed.

In Figure 8.2, D is a perfectly equal population of the same size as C, and every person in D has higher welfare than every person in C. Consequently, it follows from Normative Egalitarian Dominance that all the actions that have population C as outcome are wrong.

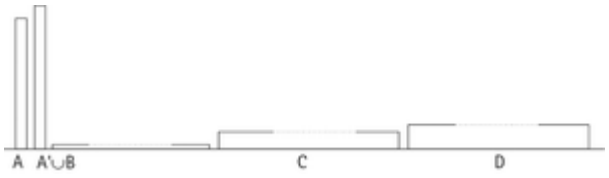


*Figure 8.2*

(p. 194) 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, Normative Non-Anti Egalitarianism 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 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 Quality, 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 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. An Axiological Theorem without Transitivity?

One might now reasonably ask: If you can prove the normative theorem without transitivity, why can you not prove the axiological theorem without transitivity?[40] And actually, we can, by using an analogous strategy to the one used in the normative case. Let me first introduce some new terminology. Earlier, we had assumed that a population axiology A is an “at least as good as” quasi-ordering of all possible populations, that is, a reflexive, transitive, but not necessarily complete ordering of all possible populations in regard to their goodness. Since we are going to drop the assumption of transitivity, we shall weaken this definition and define it on sets of populations: a *weak population axiology* is a reflexive “at least as good as” ordering of populations in any finite set of populations. Notice that this leaves open the possibility that the same pair of populations can be ordered differently in different sets.

As we noticed above, the adequacy conditions in the impossibility theorems give rise to cyclical orderings so one could replace transitivity with acyclicity to prove the impossibility theorems. However, we shall make use of the following more intuitive but logically closely related condition:

> *Maximality*: For any set of populations, there is at least one population that isn’t worse than any other population in the set.[41]

(p. 195) Maximality can be seen as an axiological analog to Separate Satisfiability. It is reasonable that, in a set of alternative populations, there should be at least one population that isn’t worse than another. If not, then for any population one picks, one could have picked a better one. Notice that Maximality is compatible with some or all popula-

tions being incommensurable, and cyclical orderings of some of the involved populations (as long as it isn't a top-cycle).

Given Maximality, it is easy to show that the following theorem is true:

> *The Axiological Impossibility Theorem*: There is no weak population axiology which satisfies Quality, Non-Anti Egalitarianism, Dominance Addition, and Maximality.

Assume the opposite and consider again the Mere Addition Paradox depicted in Figure 8.1. By invoking Quality, Non-Anti Egalitarianism, and Dominance Addition, we concluded that C is better than A′∪B, A′∪B is better than A, and A is better than C. Hence, it follows that for each population in the set, it is worse than at least one other population in the set, which is a violation of Maximality. Thus, the impossibility theorem must be true. And we have shown this without any appeal to the transitivity of "better than."[42]

# 8. Beyond the Mere Addition Paradox

Someone might object to the two impossibility theorems in the following way: "Clearly, one of the conditions must be false since we don't believe that all actions are wrong in the Mere Addition Paradox, or that all involved populations are worse than some other population in that case. Moreover, Dominance Addition 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 or situations in which whatever population one picks, one could have picked a clearly better one.[43] We cannot get rid of the paradoxes of population ethics just by giving up 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 Dominance Addition is a questionable condition, but it was the weakest link already in the transitive axiological case. The reasons against the axiological version of this condition work as well as against it in its normative guise. It might also be that we have some further reasons against the normative version of Dominance Addition that have to do with it not being context-sensitive enough.[44] 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 Dominance Addition with two conditions **(p. 196)** which are, I surmise, as hard to deny as Normative Egalitarian Dominance. Here are rough formulations of these conditions:[45]

*Normative Weak Non-Sadism (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.

*Normative Non-Extreme Priority (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 Normative Egalitarian Dominance, 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, Normative Egalitarian Dominance, a version of Quality, and a very compelling egalitarian condition (logically weaker than Non-Anti Egalitarianism).[46]

These results show that the impossibility theorems are a problem for consequentialist and nonconsequentialist alike, transitive and nontransitive value orderings alike. On a more positive note, we now know more clearly where we should and should not look for a possible way, if any, to eschew the impossibility theorems. We have to find a reason to reject one of the evaluative or normative conditions, or a reason to accept the existence of all things considered moral dilemmas or situations in which whatever population one picks, one could have picked a clearly better one.

# Acknowledgments

I would like to thank Vuko Andric, Andrea Asker, Krister Bykvist, John Broome, Tim Campbell, Erik Carlson, Sven Danielsson, Björn Eriksson, Kaj Børge Hansen, Tom Hurka, Mozaffar Qizilbash, Wlodek Rabinowicz, Daniel Ramöller, Don Ross, Joe Roussos, Howard Sobel, Orri Stefánsson, Robert Sugden, Wayne Sumner, and Michael Zimmerman for fruitful discussions and criticism of different ancestors of this paper. An earlier version of this paper was presented at the IFFS's PPE-seminar, January 2019—I would like to thank the participants at this occasion for their useful comments. Financial support from the Swedish Research Council (grant 421-2014-1037) and from Riksbankens Jubileumsfond (grant M17-0372:1) is gratefully acknowledged.

## References

Arneson, Richard J. 1989. "Equality and Equal Opportunity for Welfare." *Philosophical Studies* 56, no. 1: 77–93.

Arrhenius, Gustaf. 1999. "Mutual Advantage Contractarianism and Future Generations." *Theoria* 65, no. 1: 25–35.

Arrhenius, Gustaf. 2000a. "An Impossibility Theorem for Welfarist Axiologies." *Economics and Philosophy* 16, no. 2: 247–266.

Arrhenius, Gustaf. 2000b. *Future Generations: A Challenge for Moral Theory*, **http://www.diva-portal.org/smash/record.jsf?pid=diva2:170236**.

Arrhenius, Gustaf. 2004. “The Paradoxes of Future Generations and Normative Theory.” In *The Repugnant Conclusion: Essays on Population Ethics*, edited by Jesper Ryberg and Torbjörn Tännsjö, 201–218. Dordrecht: Kluwer Academic Publishers.

Arrhenius, Gustaf. 2011. “The Impossibility of a Satisfactory Population Ethics.” In *Descriptive and Normative Approaches to Human Behavior, Advanced Series on Mathematical Psychology*, edited by Hans Colonius and Ehtibar N. Dzhafarov, 1–26. Singapore: World Scientific.

Arrhenius, Gustaf. Forthcoming. *Population Ethics: The Challenge of Future Generations*. Oxford: Oxford University Press.

Blackorby, Charles and David Donaldson. 1991. “Normative Population Theory.” *Social Choice and Welfare* 8, no. 3: 261–267.

Boonin-Vail, David. 1996. “Don’t Stop Thinking About Tomorrow: Two Paradoxes About Duties to Future Generations.” *Philosophy & Public Affairs* 25, no. 4: 267–307.

(p. 202) Brentano, Franz. 1969. *The Origin of Our Knowledge of Right and Wrong*. Edited by Oskar Kraus. London: Routledge.

Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

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

Brown, Campbell. 2011. “Consequentialize This.” *Ethics* 121, no. 4: 749–771.

Bykvist, Krister. 1996. Utilitarian Deontologies? In *Preference and Value: Preferentialism in Ethics*, Wlodek Rabinowicz, 1–16. Lund: Dept. of Philosophy, Lund University.

Bykvist, Krister. 2007. “Violations of Normative Invariance: Some Thoughts on Shifty Oughts.” *Theoria* 73, no. 2: 98–120.

Bykvist, Krister. 2009. “No Good Fit: Why the Fitting Attitude Analysis of Value Fails.” *Mind* 118, no. 469: 1–30.

Carlson, Erik. 1995. *Consequentialism Reconsidered*. Dordrecht: Kluwer Academic.

Carlson, Erik. 1996. “Cyclical Preferences and Rational Choice.” *Theoria* 62, nos. 1–2: 144–160.

Carlson, Erik. 1998. “Two Short Papers on Harm and Value Aggregation.” *Uppsala Prints and Reprints in Philosophy* 4: 13–17.

Danielsson, Sven. 1988. *Konsekvensetikens gränser*. Stockholm: Thales.

Danielsson, Sven. 1996. “The Refutation of Cyclic Evaluations.” *Theoria* 62, nos. 1–2: 161–168.

Danielsson, Sven. 1999. “The Norm-Value Distinction in 1963.” In *Actions, Norms, Values: Discussions with Georg Henrik von Wright*, edited by Georg Meggle, 325–330. Berlin: De Gruyter.

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

Fehige, Christoph. 1998. A Pareto Principle for Possible People. In *Preferences*, edited by Christoph Fehige and Ulla Wessels, 508–543. Berlin and New York: De Gruyter.

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

Gauthier, David. 1986. *Morals by Agreement*. Oxford: Clarendon.

Heath, Joseph. 1997. “Intergenerational Cooperation and Distributive Justice.” *Canadian Journal of Philosophy* 27, no. 3: 361–376.

Heyd, David. 2019. Supererogation. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 1–23. Stanford: Stanford University Press. Retrieved from URL **https://plato.stanford.edu/archives/win2019/entries/supererogation/**.

Hooker, Brad. 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Oxford University Press.

Kavka, Gregory S. 1982. “The Paradox of Future Individuals.” *Philosophy & Public Affairs* 11, no. 2: 93–112.

Narveson, Jan. 1967. “Utilitarianism and New Generations.” *Mind* 76, no. 301: 62–72.

Narveson, Jan. 1973. “Moral Problems of Population.” *The Monist* 57, no. 1: 62–86.

Narveson, Jan. 1978. “Future People and Us.” In *Obligations to Future Generations*, edited by R. I. Sikora and Brian Barry, 38–60. Cambridge: White Horse Press.

Ng, Yew-Kwang. 1989. “What Should We Do About Future Generations?” *Economics and Philosophy* 5, no. 2: 235–253.

Österberg, Jan. 1988. *Self and Others: A Study of Ethical Egoism*. Dordrecht: Kluwer Academic.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon.

Parfit, Derek. 1996. “Acts and Outcomes: A Reply to Boonin-Vail.” *Philosophy & Public Affairs* 25, no. 4: 308–317.

(p. 203) Rabinowicz, Wlodek. 2000. “Money Pump with Foresight.” In *Value and Choice: Some Common Themes in Decision Theory and Moral Philosophy*, 1: 201–234. Boston, MA: Springer.

Rabinowicz, Wlodek and T. Rønnow-Rasmussen. 2004. “The Strike of the Demon: On Fitting Pro-Attitudes and Value.” *Ethics* 114, no. 3: 391–423.

Rachels, Stuart. 1998. “Counterexamples to the Transitivity of Better Than.” *Australasian Journal of Philosophy* 76, no. 1: 71–83.

Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, MA: Belknap.

Sen, Amartya. 1970. *Collective Choice and Social Welfare*. San Francisco: Holden-Day.

Sen, Amartya. 1995. “Rationality and Social Choice.” *The American Economic Review* 85, no. 1: 1–24.

Sider, Theodore. 1991. “Might Theory X Be a Theory of Diminishing Marginal Value?” *Analysis* 51, no. 4: 265–271.

Tännsjö, Torbjörn. 1998. *Hedonistic Utilitarianism*. Edinburgh: Edinburgh University Press.

Temkin, Larry. 1987. “Intransitivity and the Mere Addition Paradox.” *Philosophy and Public Affairs* 16, no. 2: 138–187.

Temkin, Larry. 1996. “A Continuum Argument for Intransitivity.” *Philosophy & Public Affairs* 25, no. 3: 175–210.

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

Vallentyne, Peter. 1988. “Gimmicky Representations of Moral Theories.” *Metaphilosophy* 19, nos. 3–4: 253–263.

von Wright, Georg Henrik. 1993. *The Varieties of Goodness*. Bristol: Thoemmes.

## Notes:

(1.) This paper is an extended and extensively revised version of my Arrhenius (2004). See also Arrhenius (2000b; forthcoming).

(2.) For a formal proof with slightly weaker conditions, see Arrhenius (2000b, Section 10.6) and Arrhenius (forthcoming). It should be stressed that the above paradox is not identical to Parfit’s (1984, 419ff) Mere Addition Paradox since it involves slightly different assumptions (stronger in some respects, weaker in some other respects). This version is similar to the one presented in Ng (1989, 240). A formal proof with slightly stronger assumptions than Ng’s can be found in Blackorby and Donaldson (1991).

(3.) 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 (forthcoming), (2000b) and Broome (1999). For a discussion of this issue in connection to the Repugnant Conclusion, see Fehige (1998), Tännsjö (1998), and Arrhenius (2000b; forthcoming). See also Parfit (1984, 358).

(4.) The *ceteris paribus* clause means that the compared populations are roughly equal in all other putatively axiologically relevant aspects apart from individual welfare levels. Hence, other values and considerations are not decisive for the value comparison of involved populations. For a discussion, see Arrhenius (2000b; forthcoming).

(5.) See Ng (1989, 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. For example, Ng (1989, 239, fn. 4), states that "Non-Antiegalitarianism is extremely compelling" and Carlson (1998, 288); claims that "[r]ejecting NAE [the Non-Anti Egalitarianism Principle] is … a very unattractive option." Danielsson (1988, 210) holds that "weak inequality aversion is satisfied by all ethically attractive … principles." Fehige (1998, 532), rhetorically asks "if … one world contains more welfare than the other *and* distributes it equally, whereas the other doesn't, then how can it fail to be better?." See also Sider (1991, 270, fn. 10). There are, however, reasons for rejecting the Non-Anti Egalitarianism principle. See Arrhenius (2000b, Section 6.2; forthcoming).

(6.) See Parfit (1984, 388). This formulation is more general than Parfit's except that he doesn't demand that the people with very high welfare are equally well-off.

(7.) This condition was introduced in Arrhenius (2000b). For simplicity, we are here using a slightly stronger formulation in terms of "better than" instead of "is at least as good as." Likewise for some of the other axiological conditions below.

(8.) See, for example, Arrhenius (2000a; 2000b; 2011; forthcoming).

(9.) Carlson (1995, 13). My formulation differs slightly from Carlson's (the brief discussion of consequentialism below draws 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, 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 (see, e.g., Hooker [2000]) 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 am not going to discuss. It should be clear, however, that the discussion below applies, *mutatis mutandis*, equally well to rule-consequentialism. For example, if our evaluations of outcomes are inconsistent, then we don't have any "best result" to guide our choice of rules under rule-consequentialism.

(10.) See Carlson (1995, 10–12, and Ch. 4), for an extensive discussion of the morally relevant outcome of an action.

(11.) 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. Fred Feldman's (1997) desert adjusted utilitarianism is a case in point. For a discussion of whether any morality can be formulated as a consequentialist morality, see, for example, Brown (2011), Bykvist (1996), Carlson (1995), Danielsson (1988), and Vallentyne (1988).

(12.) Broome (1991, 11).

(13.) Rachels (1998), Temkin (1987).

(14.) The impossibility results in, for example, Arrhenius (2000b; 2011; forthcoming) involve slightly weaker conditions based on the relation "is at least as good as" instead of "better than," but otherwise exhibit the same structure as discussed above.

(15.) Temkin fears, however, that nontransitivity of "better than" might spell the end for a consequentialist morality and practical reason in general. He suggests that arguments to the effect that "better than" is nontransitive "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" (1987, 186, fn. 49). In more recent work, Temkin suggests that nontransitivity "opens the possibility that there would be no rational basis for choosing between virtually any alternatives" (1996, 209), and that therefore the frontal assault on consequentialism would in fact be a frontal assault on "virtually every plausible moral and practical theory" (2012, 513). For an extensive discussion of his views about the relationship between nontransitivity and moral and practical reasoning, see Temkin (2012, 508–520). Many thanks to Tim Campbell for useful discussions of Temkin's view.

(16.) 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, 47). For a discussion of cyclical evaluations, see Carlson (1996), Danielsson (1996), and Rabinowicz (2000). Satisfaction of a condition I shall introduce below, Maximality, would also be sufficient for Act-Consequentialism to avoid implying moral dilemmas.

(17.) 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.

(18.) See Österberg 1988, 127, 145–146). My formulation is weaker than Österberg's, which is formulated in terms of the possibility of an agent to act morally right. Bykvist (2007) formulates and defends a version of Österberg's condition similar to mine in terms of the possibility of an agent to conform to a normative theory.

(19.) Österberg (1988, 146) suggests an interesting argument to the effect that his version of Separate Satisfiability is entailed by the common idea that "ought" implies "can." His argument rests, however, on the controversial agglomeration 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 (as in our condition above), not, as Österberg suggests, that at least one action is right (it could be that some actions lack normative status).

(20.) Some might think that prohibition dilemmas are acceptable if the agent has performed a preceding wrong action. The idea might be that it should be possible to conform to a moral theory but if you have already acted wrongly, then you have deviated from the right path and might end up in a morally impossible situation. To accommodate this point, we can formulate a weaker version which we could call *Weak Separate Satisfiability*: For any agent and any situation such that is hasn't been preceded by a morally wrong action by the agent, there is an action such that if the agent were to perform this action, then her action would not be morally wrong. For the purpose of this paper, the principle could be further weakened in terms of no preceding wrong actions by anyone, that is, complete moral compliance by everyone, but intuitively this seems unnecessarily weak.

(21.) Notice that there are cases involving cyclical evaluations where there still will be right actions according to Incomplete Act-Consequentialism. 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 Incomplete Act-Consequentialism.

(22.) Carlson (1996).

(23.) Jan Narveson's theory is an example of a more developed effort in this direction. He suggests 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, 55–56). Cf. Narveson (1967; 1973).

(24.) See Gauthier (1986, 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).

(25.) Boonin-Vail (1996, 279–280, 307).

(26.) Parfit (1996, 312).

(27.) Cf. Boonin-Vail (1996, 285).

(28.) Kavka (1982, 100, fn. 16). Boonin-Vail (1996, 283), discusses a similar principle.

(29.) Parfit (1996, 311).

(30.) Ibid.

(31.) Brentano (1969, 26); Ewing (1959, 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, 97). Extensive discussion of this subject can be found in Rabinowicz and Rønnow-Rasmussen (2004) and Bykvist (2009).

(32.) For a discussion of some other differences between these kinds of concepts, see von Wright (1993, Ch. 1, Sect. 4), and Danielsson (1999).

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

(34.) 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.

(35.) 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.

(36.) One might object here that whether it is morally wrong to choose B depends on the welfare level of its members. If all the lives in this outcome have a very high welfare level, then perhaps choosing outcome A would be supererogatory. However, since these actions don't involve any kind of personal sacrifice for the agent, they don't fit the paradigm description of supererogatory actions, such as someone rushing into a burning house to save its residents at the risk of her own life (see Heyd [2019]). Moreover, it is hard to see any reason for why it should be optional to choose B in the cases that fall under the condition's domain. Hence, I think this objection misses its target. Moreover, for the theorems we shall prove, we could employ a weaker version of Normative Egalitarian Domi-

nance in which the B-people have at most very low positive welfare. With this revision, it is, if possible, even harder to find a reason for considering the choice of the outcome in which everyone is better off a supererogatory action. I'm grateful to Michael Zimmerman for pressing this point.

(37.) Thanks to John Broome for reminding me that the introduction of another available action can make it wrong to help George rather than Tony because helping George would then have negative side-effects on other people's well-being. Assume that you can also help Howard with his gardening, and Howard's garden is within sight of George's garden but far away from Tony's garden. If you help George rather than Tony, Howard would notice and feel aggrieved at having been neglected. Hence, it is wrong to help George but right to help Tony. Examples such as this show the importance of including the effects on all the involved people's well-being in the specification of the outcomes of the alternative actions, as we indeed do in the statements of the adequacy conditions.

(38.) The latter is necessary to capture moralities that violate Normative Invariance, such as Normative Actualism, see Arrhenius (forthcoming), section 10.3.

(39.) For the sake of simplicity, we are assuming that a unique choice situation corresponds to any set of agent- and time-identical actions. There are, of course, a number of problems regarding how to individuate actions, what it means for an action to be available to an agent, or a group of agents, and the like. These problems fall outside the scope of this essay, however.

(40.) This question was put to me by Wlodek Rabinowicz.

(41.) Maximality is logically closely related to acyclicity since with an unrestricted domain (or a sufficiently rich domain that guarantees the existence of a subset with a top-cycle), one can derive acyclicity from Maximality. Assume that the unrestricted domain satisfies Maximality but violates acyclicity. Given the latter, there is a cycle in the domain. Given the unrestricted domain, one can pick a subset with a cycle involving all the populations in the set. Then Maximality is violated on this subset. Hence, by reductio, Maximality implies acyclicity given unrestricted domain. Moreover, the adequacy conditions in the impossibility theorems induces a top-cycle on the domain so from a logical perspective, it doesn't make much difference to go for Maximality instead of acyclicity. As we said above, however, we find the former more intuitive than the latter. I'm grateful to Wlodek Rabinowicz and Orri Stefánsson for a useful discussion of this issue.

(42.) We could also weaken the involved condition to only yield orderings relative to certain choice situations (or certain sets of populations), analogously to how we did it in the normative case. For example, we could weaken Dominance Addition in the following way: if population A and B are of the same size and everyone in A has lower welfare than in B, and if a population consisting of the B-lives and any number of people with positive welfare is worse than some other population X in a certain situation, then A is worse than X

in this situation too, other things being equal. We could then prove an impossibility theorem using such conditions. We shall not pursue this issue further here, however.

(43.) I'm here setting aside the possibility that one could find a good reason to reject some other background assumption in the theorems, such as the assumption that there are lives with positive and negative welfare or that there are lives with very high and very low positive welfare. For a discussion of these, see Arrhenius (forthcoming).

(44.) Parfit's example discussed in Section 4 is a case in point. The availability of the option of having the same child without a handicap (R) makes it wrong to have the child with the handicap (P), but it doesn't make it wrong to have no child (Q). In other words, the presence of option R has a wrong-making effect on option P but not on option R. If we add to this the assumption that the parents would be better off if they had a child, then we have a counter-example to Normative Dominance Addition which seems to have to do with it not being context-sensitive enough.

(45.) For exact formulations of these conditions, see Arrhenius (2000b; 2011; forthcoming).

(46.) See Arrhenius (2000b; 2011; forthcoming).

**Gustaf Arrhenius**

Gustaf Arrhenius is the Director of the Institute for Futures Studies, Stockholm, and Professor of Practical Philosophy at Stockholm University.