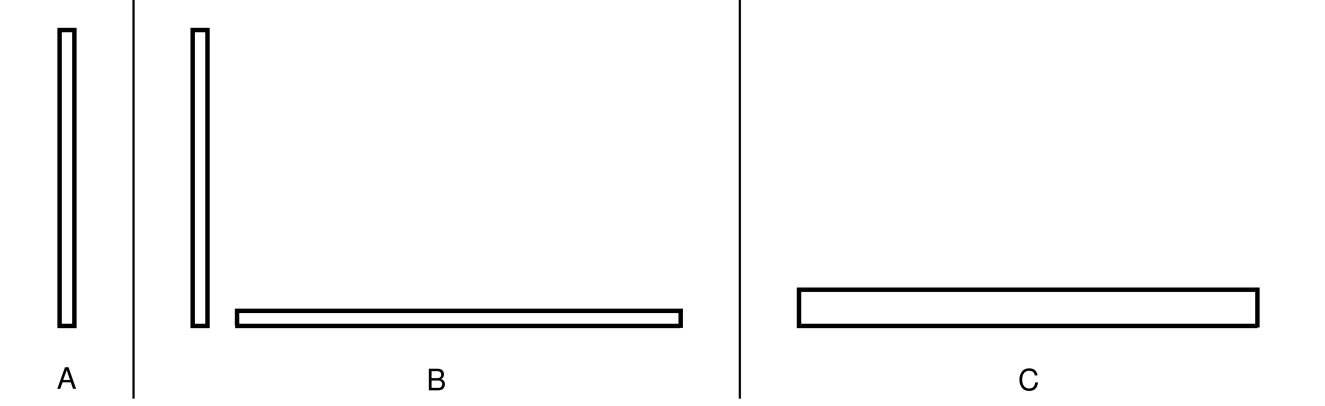
**Paradoxes of Future Fairness and the value of mere addition**

In this paper, I shall consider whether paradoxes involving mere addition in population axiology can be adequately explained by appealing to essentially comparative principles. The Mere-Addition Paradox takes the form of a set of three possible populations that we find it very hard to place into a coherent ordering according to their relative value. Traditionally such paradoxes have been explained by referring to the competing moral ideals that we try to apply in each case, such as perfectionism and egalitarianism or perfectionism and prioritarianism and the fact that different ideals are relevant when comparing different outcomes. I conclude that no such explanation is possible if we consider populations only in terms of 1) individual welfare levels and 2) the number of people in that population at each welfare level. First, I consider explanations involving perfectionism, the view that some lives have additional value because they are at a very high welfare level, and egalitarianism, the view that it would be better if welfare was distributed equally. I find that it is not plausible to explain the intransitivity of the ‘Mere Addition Paradox’ by appealing to these competing ideals. This is because the principle of equality, when modified to fit such an explanation, would face the ‘levelling down objection’ in a particularly troubling way. Next, I consider explanations involving perfectionism and prioritarianism, the view that the marginal value of a unit of welfare decreases as the welfare level of a life increases. However, perfectionism and prioritarianism not only represent competing ideals, but are directly inconsistent. They can only be plausibly combined if either or both of these principles are applied only to the value of more than individual welfare levels and the number of people in a population at each welfare level alone.

**The Mere Addition Paradox**

Consider the following three populations



Population A – a very large number of people all living at a very high welfare level (level α)

Population B – The same number of people all living at level α and a much larger number of people all living at a very low welfare level (level ω), equivalent to a life that is ‘barely worth living’

Population C – The same number of people as in population B, all at a welfare level slightly higher than level ω, but that represents a life that is still only barely worth living (level ψ), where C still has more welfare overall than B

It has been claimed that when we consider populations of this kind, we sometimes make intransitive judgements (Parfit 1984: 426, Rachels 2004…, Temkin 2011…). We judge population B to be all things considered better than population A. We also judge population C to be all things considered better than population B. Finally, however, we judge population C to be all things considered no better than population A. Such judgements form a paradox if the relationship 'X is all things considered better than Y' is transitive. This is because, if B is better than A and C is better than B then it follows, via the transitivity of betterness, that C must be better than A. However, C is not better than A. This result is known as the 'mere addition' paradox.

My aim in this paper will be neither to solve this paradox, nor to defend it’s paradoxicity. My aim is merely to examine why we appear to hold these apparently inconsistent judgements. I will do this by considering what kinds of moral principles we would need to accept in order to explain this intransitivity, and what this would imply for wider moral theory.

Any moral principles that could explain the apparent intransitivity of our moral judgements would need to do two things. Firstly, they would need to provide a reasonable explanation for our pairwise judgements about the value of each of the populations relative to each other. Secondly, they must ultimately produce an intransitive ordering of the three populations when considered together.

I find that it is not possible to provide a satisfactory explanation for the intransitivity of our judgements which does not have highly undesirable implications unless we assume further facts about these populations when we make judgements about them. Populations such as A, B and C are defined very narrowly in terms of only the welfare levels of their members. However, this information is insufficient to explain the judgements that we appear to make about them, so that we must in-fact be making assumptions about their other features as well. This appears to hold no matter how narrowly defined populations like A, B and C are or whether we try to make these judgements only on the basis of other things being equal. In order to incorporate these judgements into our axiology, we must therefore consider a wider range of features than simply individual welfare levels and the number of people in at each level.

**Utilitarianism**

Let us start by assuming that welfare itself is valuable. There are many numerous ways one can take account of this value.

Total Utilitarians take account of this value by seeking to maximise welfare ‘as a whole’ across all the lives of a population. Since population C has the most welfare overall, total utilitarians believe it to be best, and since population A has the least welfare overall, they believe it to be the worst. Even if total utilitarianism is only part of one’s moral view, additional welfare can always more than outweigh the loss of any other kind of value, such as the fact that there are no people in C at a very high welfare level (Greaves and Ord 2016…).

Average Utilitarians take account of the value of welfare by seeking to maximise the mean quantity of welfare across all the lives in a population. Since everyone in population A is at a much higher welfare level this will be the best population. However, even though nobody in population C is at a very high welfare level, the fact that C contains more welfare overall than B and is of the same size means that its average welfare level will be higher and therefore it will be preferred. Since the Average welfare level of a population does not increase with population size, this view is far easier to incorporate into a pluralist framework of values. However, most philosophers, whether pluralist or not, now reject Average Utilitarianism because of its unsavoury implications (Arrhenius forthcoming … Parfit 1984 …)

Finally, there are principles that lie somewhat in-between Total and Average Utilitarianism, including the view that would should not value either of these ideal directly, but rather score outcomes based on the extent to which they approach a particular ideal regarding each of them. (Parfit 1984 … Hurka … Temkin 2011 …). In order to explain our judgement that A was better than C, any such principle needs to severely limit the value we place on total welfare. However, in this case we should also believe that A is better than B, since B has less welfare overall and an even lower average quantity of welfare than A. Therefore, there is no way of combining total and average utilitarianism that is consistent with the view that Mere Addition is not bad.

If we are to explain our judgement that A is better than C there must, therefore, be something about population A that is at least as valuable as all the additional welfare in population C. Furthermore, It is claimed that our judgement that A is better than C cannot be reversed simply by making population C larger. This would imply that whatever it is that causes us to judge it to be better than C is in-fact more valuable than any additional quantity of welfare beyond that already present in A. Finally, if we are not to appeal to anything other than individual welfare levels and the number of people in a population at each welfare level then, this good feature must be directly implied by the existence of a sufficiently large number of people at a sufficiently high welfare level. I will consider the nature of this good below.

**Welfare perfectionism**

It is sometimes claimed that lives at a very high welfare level have additional value, because they involve enjoyment of ‘the best things in life’ and that “even if some change brings a great net benefit … it is a change for the worse if it involves the loss of one of the best things in life” (Parfit (1986) 18). If we accept this claim then we might value A more than C not for its welfare alone, but because the very high welfare level of people in A indicates that these people are enjoying the best things in life, where the value of this cannot be outweighed by the value of any additional quantity of welfare at a lower level (and hence from other sources) in population C. In population ethics, this view is called ‘perfectionism[[1]](#footnote-2).

If this view is to be applied to populations A, B and C as described above, it must equate the enjoyment of the best things in life with a sufficiently high welfare level. Lives at a high welfare level necessarily include enjoyment of the best things in life, whilst those at a lower welfare level necessarily do not – as I explain in a later section, other kinds of perfectionism are possible. Let us term this view ‘welfare perfectionism’. Only welfare perfectionism can apply to the populations A, B and C because they are defined in terms of individual welfare levels alone so that in order to apply perfectionism we need to be able to understand the perfectionist good as being equivalent to lives at certain welfare levels, and nothing more.

In order to explain our supposed judgements in this case Welfare Perfectionism must imply that C < B, since B has the same number of very high welfare lives as A, and also be consistent with the conclusion that A ≤ B, for the same reason. This produces the transitive ordering C < A < B. Hence, it is incompatible with our judgement that B < C. Therefore, we must appeal to some third good other than welfare and welfare perfection in order to explain these intransitive judgements.

In what follows I shall therefor assume that we adopt a pluralist approach to valuing populations, and that this includes something like Total Utilitarianism, implying that B > A, and welfare perfectionism, implying that A > C. I shall therefore consider what other moral ideal we might appeal to if we are to justify our intransitive judgements regarding these three populations.

**Essentially Comparative Egalitarianism**

If we are to explain the intransitivity of our judgements regarding these three populations, it is therefore necessary that our axiology include some additional element. Whatever good we appeal to in order to explain why B < C it must fulfil the following conditions

1. It gives us no decisive reason to believe that A < C
2. It gives us sufficient reason to believe that B < C

Furthermore, if we assume that B has no less welfare perfection than A it follows that the value of these goods must exceed the total value of the individual welfare in C, since whatever makes A better than C does so regardless of the relative welfare of the two populations.

When originally formulating the Mere Addition Paradox Parfit states his belief that both principle of equality and beneficence (by which he meant something like prioritarianism) might give us such reasons (Parfit (1984) 426). However, I will now argue that he was mistaken in believing that the principle of equality could perform this role, either on its own or in conjunction with prioritarianism or some other principle of beneficence.

Parfit is not clear on exactly what role the principle of equality might play in the mere addition paradox. However, this view has subsequently been developed by Larry Temkin. Temkin argues that we may be egalitarians, but egalitarians who take an essentially comparative view of the value of equality. Such a view implies that we should not judge a population’s value with respect to equality in isolation, but must also take account of the other populations with which it is being compared. Temkin summarises his view as follows:

*― Typically, when we say one outcome’s inequality is worse than another’s, the same people exist in both outcomes and the worse-off fare worse in the one outcome than the other. This, we may agree, is bad. However, comparing [B] to [A], the choice isn’t between an outcome where the worse-off fare poorly relative to the better-off and one where they fare better; rather, it is between one where they exist – with lives worth living – and one where they don’t. Here, it may seem, the inequality is not morally regrettable.* (Temkin (2011) 368)

This view seems attractive at first glance; however, it appears to quivocate over the term ‘fare worse’. There are two distinct ways in which a person may be said to fare better or worse. Firstly, a person may fare worse relative to somebody else. Secondly, a person may fare worse relative to how they would have fared in some other possible outcome.

When Temkin writes that “the same people exist in both outcomes and the worse-off fare worse in the one outcome than the other”, this seems to imply that they are faring worse than they would have done otherwise, and we might agree that this is always bad. On the other hand, when Temkin says that comparing B and A does not involve “an outcome where the worse-off fare poorly relative to the better-off and one where they fare better” he is clearly talking about the worse-off faring worse than the better-off. However, the force of this argument stems from the implication that it is the same moral ideal that implies that the inequality in B might be ‘always bad’ under some circumstances, but ‘not morally regrettable’ under others, not that there are two distinct moral principles at play in these comparisons. If it were merely the case that it is always bad for individuals to be worse off than they might have been, but not always bad for them to be worse off than others, there would be no inconsistency here and hence no reason to assume an essentially comparative view. In order to support Temkin’s claim that the value of inequality depends upon the outcomes being compared we must therefore clarify which of the senses of ‘worse off’ apply to these two claims.

What if ‘worse off’ referred, in both cases, to a person being worse off than they might have been? As I have said, it is easy to agree with Temkin’s claim that this is always bad. However, interpreting ‘worse off’ in this way undermines Temkin argument.

This is because being worse off than one might have been is not an egalitarian value. It implies that it is just as bad if the best-off fare worse than they might have done than if the worse-off do. This view would therefore give us no special reason to believe that population C is at least as good as population B. In C, the better-off fare worse and the worse-off fare better, so we are simply comparing the value of these two effects. Given that the best-off may be losing some of ‘the best things in life’, whilst the worse-off gain relatively little, it is not at all clear why a person-affecting view should give us a special reason to believe that C is at least as good as B in this case.

In order to complete his argument, I therefore believe that we must interpret both the uses of ‘worse off’ in Temkin’s argument to mean worse off relative to other members of the same population. Temkin’s argument therefore rests on the following claims:

1 – It is always bad when the same people exist in two outcomes and the worse-off fare worse, relative to the better-off, in one outcome than the other.

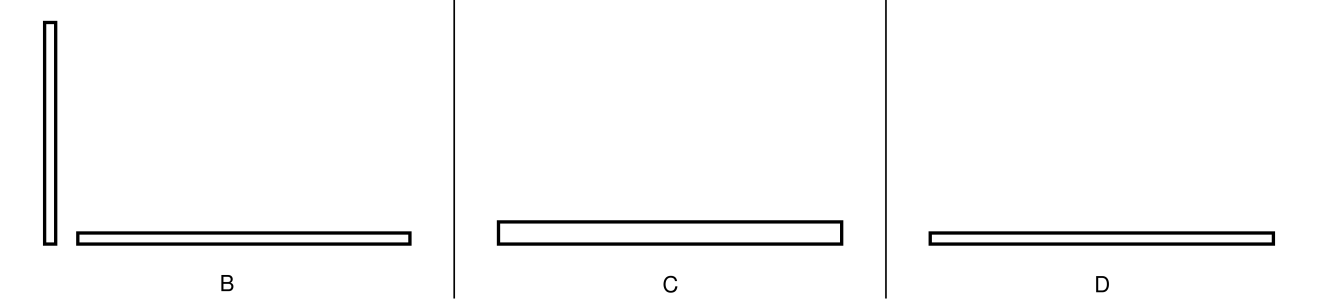
2 – When some people exist in one outcome – with lives worth living – and don’t exist in another outcome, any resulting inequality in the first outcome is not morally regrettable.

I reject this argument on the basis that it is hard to accept the first claim on its own, and that the arguments we might use to support the first claim will tend to undermine the second claim.

**The problem with Essentially Comparative Egalitarianism**

Let us consider a fourth population, D:

Population D – the same number of people as population C all living at a very low welfare level as the worst off in B (level ω)



The first claim in Temkin’s argument would imply that the inequality in B is bad when compared with D. This is because the same people exist in both populations, whilst the worse-off in D fare much better relative to the better-off than they do in B. This does not, of course, imply that D is better than B all things considered, since there are other respects in which B is better than D. However, many would not wish to say that D is better than B in any respect, since this is a case that involves ‘levelling down’. B and D are identical except for the single fact that some people in B are much better off than anyone in D.

Temkin does not accept this objection against egalitarianism in general. In his work on equality, Temkin has argued that there are in fact a number of problems with the Levelling Down Objection. However, his defence of equality carries certain implications about its value, and it is these that, I believe, ultimately undermine the second claim of his argument for the essential comparability of equalitarianism.

Although Temkin presents a number of arguments against the Levelling Down Objection in his work (Temkin (1998) 137–146), these are based on two principal claims about why we should value equality in the case of levelling down. The first of these is that the force of the Levelling Down Objection rests on the ‘Person-Affecting Condition’, which has a restricted scope (Temkin (1998) 137–8). The second is the claim that levelling down is a good thing in at least one respect, what Temkin terms ‘proportional justice’ (Temkin (1998) 138–146).

The Failure of the Person Affecting Condition

One of the key reasons for believing that the Levelling Down Objection is fatal to egalitarian theories is the view that an outcome can be better for a population only if it is good for at least one person in that population (the Person-Affecting Condition). Since, in the case of levelling down, some people are made worse off and nobody is made better off, it is hard to see how, if we accept the Person-Affecting Condition, levelling down could be in any way good. However, Temkin argues that the Person-Affecting Condition implies that we can only make judgements about cases in which at least some people are made better or worse off by our choices. However, in some cases nobody exists in more than one outcome. If we are to be able to say anything about the relative value of these outcomes then we must accept moral principles which imply that one outcome can be better than another even if there is no one for whom it is better. Therefore, even if we accept the Person-Affecting Condition for some cases, it does not follow that we should reject the egalitarianism for all cases.

However, the argument that Temkin is proposing here with regard to equality rests on something very similar to the Person-Affecting Condition. Temkin argues that the distribution of welfare in one population is more valuable than that in another population only if it is true that there are people for whom this distribution is more equal. This is precisely why, Temkin argues, the inequality in B does not make it worse than A. Since Temkin is claiming that the way in which we value equality depends on whether or not people are being made more equal, this value would not even apply in the kind of cases where the person affecting principle is problematic. The fact that in this case nobody in D is better off than anybody in B makes it much harder to claim that the equality involved in making some people worse off is a good thing. If, therefore, Temkin is to defend his view from the Levelling Down Objection, he must appeal to a value that applies in both person-affecting and non-person-affecting cases, i.e. he must defeat the Levelling Down Objection in all cases, and not merely limit its scope of application.

Levelling Down and Proportional Justice

Temkin’s remaining arguments against the Levelling Down Objection seek to motivate the claim that we have reasons to uphold principles of ‘proportional justice’ from which it follows that it is bad, in all cases, for some to be worse off than others through no fault or choice of their own. Proportional injustice, Temkin claims, is bad whether or not it is bad for the people concerned.

Temkin uses the term proportional justice in quite a vague sense, and it is not clear what exactly it is about proportional injustice that constitutes its badness. Proportional justice could merely reflect the badness of an unequal distribution of welfare, or it could represents the badness of some further fact about inequality, such as social justice or desert.

If proportional justice is a moral ideal concerning the distribution of welfare alone then it seems that it must count as a reason against accepting that A < B in relation to equality. This is because B contains a less equal distribution of welfare than A. However, Temkin sees nothing morally regrettable in the inequality in B when compared with A. This suggests that proportional justice must reflect some further fact about these cases beyond their distribution of welfare, such as the relationships that exist between the better off and worse off in B.

If, on the other hand, proportional justice is a moral ideal that applies only in certain cases, for instance those involving exploitation or social injustice, then these do not include the cases under consideration. This is because the inequality in B is the result of ‘mere addition’ and hence, by definition, is not the result of social injustice or unequal treatment of any kind. If, however, there is nothing bad about B, then the move from B to C cannot involve a reduction in the level of injustice or any other bad feature of populations usually associated with inequality, since none of these features is present in B.

To put this another way: if, on the one hand, proportional justice is the same thing as a certain pattern of lives at particular welfare levels, then it cannot make a difference to the level of proportional justice in an outcome whether it was the result of mere addition or not. On the other hand, if proportional justice stands for something else then it seems that in this case population B can contain no proportional injustice so that B is not better than C with respect to its level of proportional justice.

Therefore, if we invoke the existence of proportional justice as a justification for the claim that D is in any way better than C then this either gives us some reason to think that B might be no better than A or undermines our reason for thinking that C is worse than B. However, if we do not invoke proportional justice, then we have no further defence against the Levelling Down Objection in this case.

**Prioritarianism**

For this reason, I cannot accept that egalitarianism, even in an essentially comparative form, explains our inconsistent judgements in the Mere Addition Paradox. Instead, we must accept some principle that both a) values redistribution and b) does not face the levelling down objection. One logical alternative that meets both of these requirements is prioritarianism. The view that it matters how well off people are in absolute terms, rather than relative to one another.

Prioritarianism can be interpreted in many ways. However, all of these imply that the marginal value of individual welfare diminishes as a person's welfare level increases, so that it is always better to distribute any fixed unit of welfare to somebody who is worse off (in absolute terms) than to somebody who is better off.

We could plausibly explain the intransitivity of the mere addition paradox if we introduced essentially comparativist elements to prioritarianism. For instance if we believed, as Parfit sometimes suggests, that we should apply prioritarian weights to the value of individual welfare when comparing populations of the same size, but not when comparing populations of different sizes (Parfit (2012) 440) or that we should aggregate these two weightings differently in the two cases (Arrhenius forthcoming …). We might also explain it if we believe that we should only attach prioritarian weights to lives when we were making them ‘better’ and not when they were being introduced to a population. In this way prioritarianism might lead us to believe that C is at least as good as B without affecting our judgement that A is at least as good as B or that C is not at least as good as A. This may be an attractive view. However, there is a significant problem with using essentially comparative prioritarianism to explain the mere-addition paradox.

As I have already mentioned, there are many possible interpretations of the good I have termed welfare perfection. However, these all imply that at one or more points the marginal value of individual welfare is greatly increased by the additional value of the enjoyment of the best things in life, and that this is true at high welfare levels, but not low welfare levels. Furthermore, in order to make it the case that A is at least as good as C, no matter how large population C is, this must imply that there is at least some amount of welfare at the higher level (representing a significant amount of welfare perfection) such that no amount of priority weighted welfare at the lower level could be of greater value[[2]](#footnote-3).

Prioritarianism and perfectionism do not only express conflicting moral ideals however, they are directly inconsistent with one another – at least in their standard forms. This is because they make incompatible claims about the marginal value of individual welfare. Prioritarianism claims that this value decreases as a person’s welfare level increasing whilst welfare perfectionism claims that the marginal value of individual welfare at (some) very high welfare levels exceeds that of any amount of welfare at lower levels. In so far as we try to apply both of these principles to the value of people’s welfare, the two will simply cancel each other out. How should we respond to this inconsistency?

I see four ways of avoiding it, as follows:

1. Combine the two principals into a single complex view about the value of welfare. I shall call this view priofectarianism.
2. Only ever apply the two principles to the welfare of different people - for impersonal reasons. I shall call this the application of extreme prioritarianism and extreme perfectionism
3. Only ever apply these principles to the welfare of different people - for person affecting reasons. I shall call this the application of narrow person affecting prioritarianism and narrow person affecting perfectionism
4. Interpret either or both of these principles as concerned with goods other than welfare, I shall call such principles multiple goods prioritarianism / perfectionism

**Priofectarianism**

One possibility is that we are neither prioritarians nor perfectionists, but apply a multi-directional transformation to the marginal value of peoples welfare that incorporates both views. I call this priofectarianism. Priofectarianism represents a consistent position, so long as the transformations we equate with prioritarianism or perfectionism have different weights or different shapes, so that they do not entirely cancel each other out when combined. One form of priofectarianism would be the view that the marginal value of increasing people’s welfare decreases up to a certain welfare level and then increases again. Another would be the view that the marginal value of people’s welfare decreases along a numerical scale but crosses one or more points of discontinuity associated with an increase in the value of their total welfare.

|  |  |
| --- | --- |
|  |  |
| Two models of priofectarianism | |

Priofectarianism is not a view anybody actually purports to hold, and it seems hard to justify, except as a means of avoiding the inconsistency between two incompatible, but independently appealing, views. If it is to explain our intransitive judgements priofectarianism must include some very strange mathematical features. For instance, there must be points of discontinuity (on a scale that is constantly decreasing in the opposite direction) or points where the marginal value of welfare stops decreasing and starts increasing. As well as appearing ad hoc and arbitrary, such features are likely to produce further motivational challenges for any specific form of priofectarianism.

One such implication of priofectarianism is an apparently inconsistent view about the value of equality. For any version of priofectarianism there is a range of welfare levels β – ω (the prioritarian range) such that for any population whose members all have lives between level β and level ω a more equal distribution of welfare is better than a less equal distribution. However, there will be another range, α – ψ (the perfectionist range) such that for any population whose members lives lie solely in this range a less equal distribution of welfare is better than a more equal distribution of welfare, since this will produce more people at higher welfare levels. Furthermore, it is possible that a relatively small change in the welfare level of the lives in a population could entirely change our view about the value of an equal distribution of welfare for these lives.

**Extreme Prioritarianism and Extreme Perfectionism**

Another way to consistently apply both an increasing and decreasing transformation to the marginal value of people’s welfare is not to apply these two transformations to the welfare of the same people. One way to do this is to apply an increasing transformation only to the welfare of the best off people and a decreasing transformation only to the welfare of the worst off people.

For instance, we could interpret prioritarianism as a claim not about benefiting everyone, but only about benefiting those who are the worst off within a particular population. If implemented in a way that could potentially explain our intransitive judgements this might imply MaxiMin for individual welfare. A less extreme implementation would involve applying a weight to individual welfare that is determined by a combination of both how well or badly off a person is in an absolute sense and how well off they are relative to other people in a population.

In a parallel manner, we might reinterpret perfectionism not as a claim about the enjoyment of the best things in life in general, but about the extent to which they are enjoyed by at least one person in a population, i.e. the person who enjoys them to the greatest extent and who will therefore be the best off person in that population. On this claim it might be enough that one person was able to; for instance, listen to the music of Bach and it would not matter if others could not even listen to the music of Handel. This would turn perfectionism into some form of MaxiMax. I do not find either of these views compelling. However, in his work on population ethics and the mere addition paradox Temkin has proposed adopting both of these views as part of a ‘capped model’ for a pluralist population axiology (Temkin 2011…).

I shall not discuss Temkin’s capped model here, except to note that it has been shown to be very vulnerable to giving judgements that are clearly mistaken when cases are designed to manipulate its principles in certain ways. This is because, by moving people from one group, such as the worst off, into another, such as the best off, we can radically change the way in which we value their welfare. It therefore seems likely that this will be the case for any moral theory that includes elements such as extreme prioritarianism or extreme perfectionism, which make claims about the value of the welfare of certain specific individuals who have been separated from the wider population. For example, a view that accepted both MaxiMin and MaxiMax would be likely to claim that it is better to have one person at welfare level α and everyone else at level ψ than to have one person at welfare level ω and everyone else at welfare level β. MaxiMax and MaxiMin are also inconsistent with the spirit of prioritarianism or perfection. Since these views make claims about the overall goodness of an outcome, not merely its goodness for any one individual. However these two views imply that the number of people at a high or low welfare level does not matter, so it would be no better or worse in this respect to benefit everyone or just a single individual.

**Narrow Person Affecting Prioritarianism and Narrow Person Effecting Perfectionism**

A third way of avoiding the conflict between prioritarianism and perfectionism is to develop Temkin’s proposal for Essentially Comparative Values further than he suggests. Temkin himself tends to classify prioritarianism as a form of egalitarianism, even though, as I argued above, it does not allways value equality unless two populations have the same average welfare level. This suggests that, an essentially comparative prioritarianism may in fact have been the view Temkin was proposing as part of his explanation for the intransitivity of the Mere Addition Paradox.

However, it is not enough to avoid the inconsistency between prioritarianism and perfectionism simply to hold that either or both views are essentially comparative in the way Temkin describes. If there is any case in which both apply to the same person’s welfare then the inconsistency remains troubling. However, if we adopt an essentially comparativist view because we think one or both of these values are ‘person affecting’ then it may be that they apply in the same cases, but never to the same people.

For instance, we might value high welfare lives more in cases involving distinct populations, such as the Repugnant Conclusion. However, when some people exist in each of the populations under consideration, so that our choices directly affect them, then we believe that we should benefit the worst-off. In this way, we simultaneously apply both an increasing and a decreasing transformation to the marginal value of individual welfare; however, the two are applied to different people so that they are no longer inconsistent.

Such views require the establishment of identity across populations. This means they do not apply to cases where no identity can be established across populations (such as cases involving future generations or distinct populations). It also means that they cannot be expressed in terms only of individual welfare levels. If we adopt a narrow person affecting view of one, but not the other, we will still face inconsistencies when dealing with people to whom both views apply. If we adopt a Narrow Person Affecting view of both, so that no person can have both views applied to them at the same time it will be hard to distinguish between which views should apply to each person.

Furthermore, if we do try to apply personal identities to the populations in the Mere Addition Paradox in order that we might use these principles to explain its intransitivity we get counterintuitive results. Since we believe that population A is better than C because of its perfection it holds that the people in A must not also be in C. However since we value C more than B because we want to benefit the worst off in B it follows that these people must also be in C. The intransitivity of our judgements cannot therefore only be explained by the fact that the worst off people in B are not also in A, it also requires that the best off people in B do not also exist in A as well. However, this violates the conditions necessary for the transition from A to B to be classed as a case of ‘mere addition’.

**Multiple Goods Prioritarianism / Perfectionism**

Finally, we can avoid the inconsistency by interpreting one or both views as claims about the value of something other than welfare. For instance, we might claim that perfectionism is a view about a particular good (the best things in life) which is closely associated with, but not the same as, a very high welfare level.

In “Overpopulation and the Quality of Life”, Parfit acknowledged the possibility that the best things in life may be present in lives at any welfare level. However having acknowledged this fact he, and his successors, seem simply to ignore this possibility. If Perfectionism is not a claim about welfare at all, but about a particular sort of good that may or may not be present in lives at a wide range of welfare levels, then it is perfectly compatible with prioritarianism, since prioritarianism makes no reference to enjoyment of the best things in life.

Similarly, prioritarianism can be understood either as a view about the priority given to satisfying the claims of the worst-off or about the value of things that are associated with lower welfare levels, such as suffering or frustration. The prioritarian claim that “benefiting people matters more the worse off these people are” (Parfit (1997) 213) can be, and often is, identified as a claim about giving welfare to people, i.e. improving their welfare level without changing their identity. However, this is not the only interpretation of prioritarianism that can be produced. A second interpretation is that prioritarianism makes a claim about the strength of the claims individuals have to our moral concern. One advantage of this interpretation that is that it has been shown by Michael Otsuka and Alex Voorhoeve to deal more effectively with separateness of person’s objections to prioritarianism (Otsuka and Voorhoeve (2009) Rendall 2011 …).

There is one final advantage to making prioritarianism a view about the strength of people’s claims, rather than the value of welfare. This is that it allows for other views about the strength of competing claims to be taken into account alongside prioritarianism. This might involve other reasons that individuals have a claim against us that will be largely sympathetic to the prioritarian worldview, for instance that benefiting people matters more the worse off they are relative to others, or the more suffering they experience in their lives. It also might involve reasons that are very tangential to prioritarianism, such as that a person could, or could not, be better off than they are (i.e. Claims that derive from the modal features of a life). Finally it might even involve reasons that largely run counter to the prioritarian world view, for instance whether people enjoy some of the best things in life.

There is another alternative interpretation of prioritarianism that avoids its conflict with perfectionism. This is the view that there are certain features of lives which are ‘barely worth living’ that are especially bad, just as there are features of very happy lives that are very good. These might include a significant proportion of negative welfare components such as suffering, frustration or the absence in this life of the good things that anybody might expect to have or achieve, such as personal relationships. We should benefit people with lower levels of welfare as a priority on this view, not because there marginal welfare is more valuable in itself but because doing so is most likely to remove some of these bad features. Neither of these interpretations of prioritarianism is incompatible with perfectionism, even in the form of welfare perfectionism

These ‘multiple goods’ interpretations of prioritarianism and perfectionism cannot be expressed in terms only of individual welfare levels, since they make reference to goods that may be present at many different welfare levels. Nevertheless, they seem closer to the views associated with these labels when they are not being used in the field of population axiology.

**The implications of person affecting or multiple goods views**

It seems to me that these last two interpretations of prioritarianism and perfectionism are the most likely candidates for explaining the supposed intransitive judgements about the populations in the mere addition paradox. However, they are not without their challenges. In particular, both of these views have profound implications for other aspects of population axiology.

Both person affecting and multiple goods interpretations of prioritarianism and perfectionism can only be applied in situations where we have additional information about the lives in each population, such as the strength of people’s claims against us or the various components of their welfare. Without this information, these theories can say little if anything, and what they can tell us will always be contingent upon these further facts.

This implies that our intuitions, which seem to make use of these principles in arriving at the kinds of intransitive judgement that we commonly experience, are poorly suited to the populations as they are specified in the mere addition paradox. In fact additional information often is given about these populations, but this is treated as colour or as merely tangential to their value. The method of producing a population axiology that I have pursued in this paper is not meant to rely on such additional information at all.

I therefore conclude that my arguments present us with a challenge. Either we must revise our approach to population axiology so that it takes account of additional information that is not provided in examples like those in the mere addition paradox. Since our judgements depend upon additional information that is not being specified, we must abandon any hope of incorporating these judgements into our population axiology. Without them however, it is hard to see what sort of methodology we might find that would allow us to make any progress in this field at all.

**Conclusions**

Explaining the intransitivity of the 'Mere Addition Paradox' turns out to be surprisingly difficult. An appeal to the three values of equality, perfection and utility cannot be used to explain it, even if the value of equality is taken to be essentially comparative.

If we are to try and explain the intransitivity of our judgements regarding this paradox by appealing to perfectionism and prioritarianism then we face one of the following three alternatives:

1. We hold clearly inconsistent views about the value of welfare – prioritarianism and perfectionism
2. We hold views that imply arbitrary and counterintuitive results – priofectarianism or an extreme view
3. We hold views that cannot be clearly expressed in terms of individual welfare alone such as person affecting or multiple goods views of prioritarianism or perfectionism

Of these alternatives, option 3 seems to me by far the most coherent and reasonable response. However, it requires population axiologists to either abandon their pretence that the value of populations can be described only in terms of individual welfare levels and the number of people at each level or abandon their attempt to base population axiology on our intuitive judgements about different populations, such as those of the mere addition paradox[[3]](#footnote-4).

**References**

Arrhenius, G., Forthcoming, Population ethics, a challenge to morality, Oxford: Oxford University Press.

Otsuka, M, and Voorhoeve, A., 2009, “Why It Matters that Some Are Worse Off than Others: An Argument against the Priority View,” Philosophy and Public Affairs 37, pp. 171–199.

Parfit, D., 1984, Reasons and Persons, Oxford: Clarendon Press.

Parfit, D., 1986, “Overpopulation and the Quality of Life.”, in P. Singer, ed., Applied Ethics, Oxford: Oxford University Press, 145–164. Reprinted in J. Ryberg and T. Tännsjö (eds.) 2004.

Parfit, D., (2012, “Another Defence of the Priority View,” *Utilitas* 24, pp. 399-440

Parfit, D., Forthcoming, “Towards Theory X part 2”

Temkin, L., 1998, “Equality, Priority and the Levelling Down Objection.”, in M Clayton and A Williams, ed., The Ideal of Equality, Basingstoke: Palgrave, 126-162.

Temkin, L., 2011, Rethinking the Good, Oxford: Oxford University Press.

1. As such it is only vaguely related to other theories known as perfectionism [↑](#footnote-ref-2)
2. This is sometimes termed ‘Millian Superiority’ see Arrhenius (Forthcoming ) chapter 6 [↑](#footnote-ref-3)
3. I am grateful for feedback on this paper from Veronique Murnos-Darde, Mike Otsuka, Alex Voorhoeve, Larry Temkin and an audience at the London Moral Philosophy Workshop 2015 [↑](#footnote-ref-4)