**What is repugnant about the repugnant conclusion?**

In this essay I will consider a question in the field of population ethics that I see as inadequately addressed by the existing literature. If the Repugnant Conclusion is 'repugnant' then what is the source of its repugnance? I consider this question to be significantly different from two more prominent questions, namely 'should we accept the Repugnant Conclusion?' and 'how can we avoid the repugnant conclusion?'. However I shall argue that the answers to these questions may well be closely related.

There are potentially many ways of explaining the alleged repugnance of the repugnant conclusion, however in this essay I shall pursue only one, namely that this conclusion, implied by several theories in population ethics, has one or more features that we find very hard to accept. I find that pursuing this possibility demonstrates that it could be very difficult to explain the repugnance of the repugnant conclusion in simple terms, but that the difficulty is remarkably similar to that of avoiding the repugnant conclusion. I shall however assume throughout this essay that the repugnant conclusion is, in-fact, repugnant.

**The field of population ethics and the Repugnant Conclusion.**

In this essay I shall understand population ethics as simply all ethical judgements in which the identity of the people we affect is changed by our actions. This implies that we cannot make people better or worse off, instead we can only make (i.e. cause to exist) better or worse off people[[1]](#footnote-2). This implies that we must reject the so called 'person affecting view' that something is good if it is good for someone and bad if it is bad for someone. It also removes the moral relevance of the 'separateness of persons' because since we cannot make individuals better or worse there is no moral sense in which we could treat judgements about a single individuals differently from those about separate people.

All principles and intuitions that we wish to employ in population ethics must therefore be stated in strictly non person-affecting terms. These restrictions greatly reduce the breadth of moral theory we can employ within the realms of population ethics. Amongst other things it seems to several significant moral objections to consequentialism, which has subsequently become the dominant theory within population ethics.

One form of intuition that remains a problem for consequentialist theories is that there are some conclusions, implied by particular theories that are 'very hard to accept'. Chief amongst these is the 'Repugnant Conclusion', which is implied by total utilitarianism.

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” (Arrhenius (2000) 248)*

This can be represented graphically as follows:

Where the width of each box represents the size of the population, the height represents the welfare of people in that population and the sigh < represents the relationship better than.

The Repugnant Conclusion works as a counterexample to our intuitive judgements of goodness, and hence we find it very hard to accept. Whatever it is about this judgement that makes it very hard to accept I shall call its repugnance. I take this to be similar to the feature of many conclusions in person affecting ethics that we also find 'very hard to accept', such as the conclusion that we should kill one person to save five. However I do not believe that it is clear exactly what it is about the repugnant conclusion that gives it this repugnance. To try and discover that is the aim of this essay.

My main concern in this essay will be to consider the nature of the repugnance of the Repugnant Conclusion. However, since I shall refer to them later on in this essay I wish to flag up two more conclusions implied by other theories in population ethics that seem to be equally 'hard to accept', sometimes labelled the 'sadistic conclusion' and the 'anti-egalitarian conclusion':

The Sadistic Conclusion - *“When adding people without affecting the original people's welfare, it can be better to add people with negative welfare rather than positive welfare.” (Arrhenius (2000) 251)*

The Anti-egalitarian Conclusion - *“A population with perfect equality can be worse than a population with the same number of people, inequality, and lower average (and thus lower total) positive welfare.” (Arrhenius (2000) p258)*

I shall not comment in this essay on whether any of these conclusions are easier to accept than others, but rather will simply assume that all three of them have the property of being 'very hard to accept', i.e. they all have the feature of repugnance. Whether each of them is repugnant in its own way, or whether it is due to something they share is an important issue I shall not consider here.

**What is it about the repugnant conclusion that is so hard to accept?**

If we accept that the Repugnant Conclusion is repugnant it is a useful and legitimate question to ask why it is so. In this essay I shall consider only two possible answers to this question. Firstly I shall consider whether the only answer we can give is that the repugnant conclusion is in itself repugnant. Secondly I shall consider whether there is some particular feature of the repugnant conclusion that is the source of its repugnance.

The Repugnant Conclusion is just repugnant

To say that the Repugnant Conclusion is 'just repugnant' is to say that we cannot give any further explanation of why it is 'very hard to accept', beyond the fact that it exists as a counterexample to our intuitive judgements of goodness. This seems to me to be implicitly assumed by those who have considered whether the repugnant conclusion is repugnant. For instance Michael Huemer speaks of a single 'unrepugnant intuition' which he sees as conflicting specifically with the Repugnant Conclusion to make it appear very hard to accept, although he concludes that we should in-fact accept the repugnant conclusion and revise our intuitions.

One reason why I do not wish to take this approach is precisely because it makes the Repugnant Conclusion much easier to accept. As Huemer argues, one of the reasons for revising an intuition is if it is “a 'bare' intuition, lacking significant support from other intuitions, and lacking a satisfactory theoretical explanation” (Huemer (2008) 907). Huemer concludes that this is true of whatever intuition lies behind the Repugnant Conclusion, but he puts in no real effort into determining what this intuition is or whether it is really the case that it is a 'bare intuition' without 'significant support from other intuitions' or a 'satisfactory theoretical explanation'. If supporters of the Repugnant Conclusions repugnance merely argue that it is by itself 'very hard to accept' they would be conceding this argument, but that does not mean it must be conceded.

A second reason why I do not wish to take this approach is that it seems to make the Repugnant Conclusion much easier to avoid. It would, I think, be no real weakness to total utilitarianism if it was conceded that it produced the correct results in population ethics only for populations that are not of dramatically different sizes, as they are in the Repugnant Conclusion. Whilst not attractive at first sight, such an approach to avoiding the Repugnant Conclusion has several key advantages. Firstly it ties in with our sense that comparisons of the value of populations are made more difficult when they are of different sizes, but that for populations of a similar size they are perfectly possible. Secondly, the kind of judgement implied by the Repugnant Conclusion is practically impossible, and would never trouble our ethical theories in any conceivable policy decision. Finally, it matches a 'bare intuition' with a 'bare exclusion' recognising that, if we really cannot find any theoretical explanation for the repugnance of the Repugnant Conclusion it does not seem sensible that we should re-design our entire moral theory to avoid it.

Even if we did wish to avoid the Repugnant Conclusion by a more sophisticated method than a simple exclusion from total utilitarianism it should further be noted that arguing that the Repugnant Conclusion is just repugnant gives us not further clue to understanding how best to avoid it, and whether the various problems people have found with doing so are systematically related, or merely a set of unrelated principles thrown up by our intuitive judgements of goodness.

What features of the Repugnant Conclusion could explain its repugnance

There are several approaches to explaining the repugnance of the Repugnant Conclusion. In this essay I shall consider only one of these, namely to search for a feature of the Repugnant Conclusion that is the source of its repugnance and that it shares in common with other judgements that we find very hard to accept.

Examining the Repugnant Conclusion suggests three such features that may be the source of its repugnance. Firstly it represents an extreme trade-off between maximising the average welfare of a population and its size, whereas our intuitive judgements of goodness may only allow for such trade-offs to a limited extent. Secondly, it represents an extreme trade-off between lives that are much better and those that are much worse, whereas our intuitive judgements of goodness may only allow for such trade-offs to a limited extent. Thirdly it represents the value of very happy lives or lives that are barely worth living as being directly comparable, whereas our intuitive judgements of goodness may include some qualitative difference between the values of such lives. To put these three features more succinctly, the repugnance of the Repugnant Conclusion may be caused by its implied evaluations of quantity of life compared to quality of life, the relative value of better and worse lives or the absolute value of either very happy lives or lives that are barely worth living. Any one of these features could be sufficient to explain the repugnance of the Repugnant Conclusion and I shall consider each of them in turn.

**Three features of the Repugnant Conclusion**

The relative values of quality and quantity of life

Total utilitarianism implies that the value of utility is irrespective of its distribution. In particular that it does not matter whether a great many people have a small amount of utility or that a few people have a great deal. Thus any reduction in the quality of life of people in a population can be compensated for, in terms of the overall value of that population, by an increase in its size. However, our intuitive judgement of goodness may not share this implication. In particular it may be that intuitively we value populations only according to the (average) level of quality of life of their members or that we make only a limited adjustment to the value of a population according to its size. If this were the case then we would not be willing to accept that a significant increase in the size of a population could increase its value so that even if people in that population had their quality of life substantially reduced it might increase in value. This would be sufficient to explain the repugnance of the Repugnant Conclusion.

If this were the source of the repugnance of the Repugnant Conclusion then it would follow that many other judgements that involved significant trade-offs between quality and quantity of life of this kind would be equally hard to accept. For instance, if this was the source of the repugnant Conclusion then the following judgement should also be hard to accept



Note that in this comparison the first population has a much higher total utility, due to the significant number of people who are all living lives that are worth living. However the number of people whose lives, though worth living, are not so good means that the average quantity of life is lower than the second population, where a small number of people whose lives are not worth living reduces the average quality of life significantly, but not to the same extent

However, far from this judgement being hard to accept in itself, the alternative judgement, that the second population is better than the first, implies the 'Sadistic Conclusion', since it would demonstrate that it may be better for a population to contain people whose lives are not worth living than a large number of people, all of whom live lives worth living. If we find the Sadistic Conclusion hard to accept then we could not also find this judgement hard to accept. Even if we do not find the Sadistic Conclusion hard to accept this judgement still appears quite attractive and far from repugnant.

The absolute value of very happy lives or those barely worth living

Total utilitarianism implies that the value of all lives comparable on the same scale, so that a sufficient number of any lives that are worth living could always be more valuable than some number of very happy lives. However, this may not be our intuitive view. Whilst we might agree that both very happy lives and those that are barely worth living are 'good', or at least not 'bad', we may intuitively feel that no amount of lives barely worth living could be better than any number of very happy lives. One reason we might believe this is because we either view very happy lives as having a lexically superior value or that we view lives that are barely worth living as having a lexically inferior value, or indeed no value at all. I shall not distinguish between these views. Instead I will simply refer to lives above or below a critical level; irrespective of whether this represents the boundary between positive value and no value or between lexically superior and inferior value or is the level between happy and very happy lives or lives barely worth living and lives that are more worth living.

If this were the source of the repugnance of the Repugnant Conclusion then it would follow that many other judgements that involved some lives that were either very happy or barely worth living (depending on our view) would be equally repugnant. For instance, if this was the source of the repugnant Conclusion then the following judgement should also be hard to accept



Note that in this comparison the two populations are of exactly the same size and the first has both a higher total and average utility. However, the second population contains some lives above the critical level and the other contains only lives below the critical level. If it were true that we preferred any number of lives above the critical level to any number below that level then the second population, which has some lives above this level, would clearly be preferable to the first, which did not, and we would find this judgement very hard to accept.

However, far from this judgement being hard to accept in itself, the alternative judgement, that the second population is better than the first, implies the 'Anti-egalitarian Conclusion', since it would imply that a population with perfect equality can be worse than a population with the same number of people, inequality, and lower average (and thus lower total) positive welfare. If we find the Anti-egalitarian Conclusion hard to accept then we could not also find this judgement hard to accept. However, again, even if we do not find the Anti-egalitarian Conclusion hard to accept then this judgement still appears quite attractive and far from repugnant.

The relative value of very happy lives and lives barely worth living

It therefore appears unlikely that the source of our sense of repugnance could be a failure by total utilitarianism to fully respect the absolute value of either very happy lives or those that are barely worth living. However, there is another potential source of repugnance; that total utilitarianism fails to fully respect the different relative values of these two types of lives in respect of each other. It may not be that there is any critical level between lives whose values are lexically superior and inferior, or otherwise qualitatively incomparable, but that we still do not intuitively agree that there is any number of lives that are barely worth living that could be more valuable than some number of lives that are very happy ones.

One problem with such a view is that it is theoretically not possible to hold a consistent axiology in which some goods are lexically superior to those that are much worse than themselves, but not to others that are only slightly worse. As Arrhenius has pointed out, if we assume that between two goods or lives there is no relationship of lexical superiority then it follows that some amount of one of them is better than some amount of the other. Thus if a qualitative theory attempts to hold that lexical superiority only holds between two goods that are separated by a large gap, call them A and Z, there must be some amount of a good M such that this is better than any amount of A, where M is worse than A. There must also be some amount of another good N such that any amount of Z is always better than this. However, there can be no amount of M, such that this is worse than any amount of Z, or any amount of N, such that this is better than any amount of A. If there were then it would be necessary to conclude that some amount of Z could be better than any amount of A through transitivity, which would violate the relationship of lexical superiority. It further follows that, no matter how similar they are, no amount of N could be better than any amount of M. i.e. M must be lexically superior to N, even if N is only slightly worse than M (Arrhenius (2005) p106).

However, this may not in-fact be the case, so long as it is not the case that all populations can be compared as better than or worse than others, but that some can only be compared as being of imprecisely equal value. Derek Parfit has claimed that in this case it is then possible for A and Z to be in a relationship of lexical superiority and N and M not, so long as it is indeterminate whether any number of N lives could be better than some number of A lives (i.e. we can at most determine that the two populations would be imprecisely equally valuable) and similarly for any number of Z lives and M lives (Parfit (Forthcoming) 7).

If we agree with Parfit that this is the case then it follows that the source of our repugnance could indeed be that total utilitarianism violates our intuitive judgements of the relative values of lives in a way we find very hard to accept. If this were he source of the repugnance of the Repugnant Conclusion then it would follow that many other judgements that involved two populations, one of which with some lives that were much better than all those in the other population would be equally repugnant. For instance, if this was the source of the repugnant Conclusion then the following judgement should also be hard to accept:



Note that in this comparison the two populations are of exactly the same size and the first has both a higher total and average utility. However, the second population contains some lives that are significantly better than all those in the first population (the two lines here represent the boundaries of what would constitute a 'significant' difference, i.e. One that could make one level of quality of life lexically superior to another). If it were true that we preferred any number of lives that are 'much better' to any number that are 'much worse', then the second population would clearly be preferable to the first and we would find this judgement very hard to accept.

However, far from this judgement being hard to accept in itself, the alternative judgement, that the second population is better than the first, again implies the 'anti-egalitarian conclusion', since it would demonstrate that a population with perfect equality can be worse than a population with the same number of people, inequality, and lower average (and thus lower total) positive welfare. Some people have found this particular instance of the anti-egalitarian conclusion less than hard to accept, and certainly easier to accept than the Repugnant Conclusion (Parfit (forthcoming) 23 - 27). However is still appears, on balance, harder to accept than its alternative, which this theory of repugnance claims should be very hard to accept. Once more, even if we do not find the anti-egalitarian Conclusion hard to accept, in this or any other instance, this judgement still appears quite attractive and far from repugnant.

**repugnant, or merely Repugnant**

My attempt so far to examine the source of the repugnance of the Repugnant Conclusion is by no means exhaustive. There remain several potentially fruitful sources of the repugnance of the Repugnant Conclusion left to examine (presumably the collective pronoun for such groups is a 'doctorate' e.g. a doctorate of problems?). However I find two features of my findings to be potentially interesting.

Firstly, as I have taken pains to illustrate, the problems with finding the source of repugnance of the Repugnant Conclusion shares a great deal of similarity with the problems of finding a way of avoiding it. The most likely features of repugnance would, it seems, lead us to conclusions that violate the most likely theories proposed to avoid the Repugnant Conclusion. This suggests to me that there may be some systematic failure in attempts so far to dealing with the repugnant conclusion by coming up with ingenious methods of avoiding it, without properly considering what the source of its repugnance might be.

Secondly, attempts to deal with the Repugnant Conclusion by avoiding it have an ambiguous connection with questions about whether or not it is actually repugnant. Even if a satisfactory moral theory that avoids the repugnant conclusion can be found it would still be a valid question whether or not it is necessary to adopt such a theory if the Repugnant Conclusion does not need avoiding after all. On the other hand, if we find a convincing source of the repugnance of the Repugnant Conclusion then this would seem to provide stronger reasons for accepting that it is indeed repugnant, whilst if we do not then its repugnance would seem to be founded on “a bare intuition, lacking significant support from other intuitions, and lacking a satisfactory theoretical explanation” as Huemer describes, and would thus be much easier to dismiss.

Whilst I have only scratched the surface of considering what the potential source of the repugnance of the Repugnant Conclusion might be, I therefore conclude that there is a strong case for trying to answer this question, and that it may turn out to be just as important as whether the Repugnant Conclusion is repugnant or how, if it is, we can best avoid it.

**Bibliography**

Arrhenius, G. (2000), “An Impossibility Theorem for Welfarist Axiology”, Economics and Philosophy, 16: 247–266.

Arrhenius, G. (2005), “Superiority in Value”, Philosophical Studies, vol. 123. Reprinted in M. Zimmerman and T. Rønnow-Rasmussen (eds.), Recent Work on Intrinsic Value, Springer Verlag, 2006.

Huemer, M. (2008), “In Defence of Repugnance,” Mind, 117 (468): 899–933.

Parfit, D. (Forthcoming) “Towards Theory X: Part Two”, author's rough draft

1. I am assuming here that existence does not constitute a benefit in itself. [↑](#footnote-ref-2)