Why We Ought to Accept the Repugnant Conclusion

Torbjörn Tannsjö
Stockholm University

I. INTRODUCTION

Derek Parfit has famously pointed out that ‘total’ utilitarian views, such as classical hedonistic utilitarianism, lead to the conclusion that, to each population of quite happy persons there corresponds a more extensive population with people living lives just worth living, which is (on the whole) better. In particular, for any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living. This world is better if the sum total of well-being is great enough, and it is great enough if only enough sentient beings inhabit it.¹ This conclusion has been considered by Parfit and others to be ‘repugnant’.²

In the present context, how are we to understand ‘repugnant’? I suggest that the word means something like obviously false. This means that, if a moral principle which strikes us as quite plausible in its own right has been shown to lead to this conclusion, it should, only for this reason, be read out of court. This interpretation is certainly suggested by the word ‘repugnant’. And unless we make this (epistemic) interpretation it becomes incomprehensible why arguments from plausible principles leading to the repugnant conclusion should be considered paradoxical.³

Is the repugnant conclusion ‘repugnant’, then? I think not. Several philosophers have acknowledged and accepted the repugnant conclusion.⁴ I have myself defended the claim that this conclusion is

¹ From the point of view of hedonistic utilitarianism, the repugnant conclusion should really be stated in terms of sentient beings, not in terms of people. It is quite possible, from the point of view of hedonistic utilitarianism, that we should prefer a world with many sentient non-human animals who lead lives just worth living to a world with very happy, though not so many, people.


³ Parfit’s own comment on his terminology is: ‘As my choice of name suggests, I find this conclusion very hard to accept’, ibid. And he does take it to be a strong desideratum that a moral principle (X) be found which, when applied to problems of population ethics, does not imply it.

⁴ In ‘A Set of Solutions to Parfit’s Problems’, Nous, xxxv (2001), 232, Stuart Rachels cites Sikora, Anglin, Ng, Attfield, Ryberg, Norcross, and Fotow as philosophers who, in addition to the present writer, have accepted the repugnant conclusion. This list could easily be made longer.
acceptable.5 My argument has been accepted by Christian Munthe in his dissertation,6 and by Jesper Ryberg in his dissertation and elsewhere.7

In this article I will once again state my argument why the repugnant conclusion is not repugnant, and I will try to answer some objections that have been raised against it.8 My objective with this article is however twofold. First of all, I want to argue that, even if, to some people, the repugnant conclusion may at first sight seem unacceptable, it is not repugnant, i.e. it is not obviously wrong. I will restate my argument to this effect and rebut objections that have been raised against it. Second, I will put forward three arguments to the effect that the repugnant conclusion follows from any plausible moral principle, when applied to problems of population ethics. This means that, if a moral principle does not imply the repugnant conclusion, we must suspect that there is something seriously wrong with it. These arguments are not new, they are already present in the literature, but, because the repugnant conclusion has been considered obviously wrong, their true significance has not been acknowledged.

In my discussion I will focus on cases where people (and other sentient beings) lead lives that are positively worth living. This is in line with how Parfit has stated the repugnant conclusion. However, it should be noted in passing that, on hedonistic utilitarianism, if each of the many people who live lives just worth living gets his or her situation worsened only a little bit, so that each one comes to live, instead of a life just worth living, a life that is just worth not living, the result will be a world that is worse than a world where ten billion people do all live terrible lives. This may be considered to create a problem in its own right. Having also defended the ‘ultra-repugnant conclusion’ that, in order to improve, in a sub-noticeable way, the well-being of a great many people, who already live happy lives, we may have to torture one person, I have no problem with this conclusion either.9 However, in the present context I will just leave the problem of evil to one side and focus on the standard formulation of the ‘repugnant’ conclusion. Even those who find that there are amounts of suffering so terrible that they cannot be counterbalanced by any decrease in

5 ‘Who are the Beneficiaries?’, Bioethics, vi (1992).
6 Christian Munthe, Livets slut i livets början (The End of Life at Life’s Beginning), Stockholm, 1992, ch. 5.
7 Topics on Population Ethics, Copenhagen, 1996. See also his ‘Is the Repugnant Conclusion Repugnant?’, Philosophical Papers, xxv (1996).
8 Even if, on my view, ‘the repugnant conclusion’ is a misnomer I will, for the sake of simplicity, use it. I feel free do so since this manner of speaking can hardly be considered biased in favour of the thesis I want to defend.
9 Cf. my Hedonistic Utilitarianism, ch. 5. The term ‘the ultra-repugnant conclusion’ was suggested to me by Derek Parfit.
less serious harm, let alone by any increase in happiness, may accept my argument that there is nothing repugnant about the 'repugnant' conclusion, as it has been stated by Derek Parfit. Upon closer examination this conclusion may even be seen to be, not only perfectly acceptable, but also indeed irresistible.

II. THE REPUGNANT CONCLUSION AND COMMON SENSE

The repugnant conclusion has been considered to be at variance with common-sense morality. This is a mistake. There exists no answer from common-sense morality to the problem posed by the repugnant conclusion, for this conclusion is the result of a very advanced thought-experiment, and the typical common-sense reaction to such very advanced thought experiments is to deny their moral relevance. Common sense does not enter into this kind of speculation. So even if the repugnant conclusion gets no support from common-sense morality, it is not threatened by it either. The repugnant conclusion is simply irrelevant to common-sense morality, which deals exclusively with more mundane situations, situations more 'down to earth'. Any philosopher who has tried to talk to lay people about the repugnant conclusion must have come across the reaction that this is simply of no interest to them, it being a mere thought-experiment typical of philosophers and totally irrelevant to all practical purposes. I have certainly met with this reaction many times myself.

What contributes to the observation that our common-sense morality is silent about the repugnant conclusion is not only the fact that the conclusion is derived from a thought-experiment, but also that it is derived from a very abstract thought-experiment. In its abstract formulation it is difficult even for a philosopher, well-versed in thought experiments, to form any firm intuition in relation to it. We need to know more about what a life well worth living, a life barely worth living, a life not worth living, and a life worth not living really amount to, in order to be able to take up any stance whatever to the repugnant conclusion.

The repugnant conclusion bears little connection to common-sense morality, then, but it is still of the utmost relevance to theoretical ethics. For it is reasonable to hold that an acceptable moral theory should give the right answers, not only in all actual cases, but also in hypothetical cases, at least such hypothetical cases as are not only logically but also nomologically possible. We should ponder some possible cases then, even if, technically speaking, they are not possible. Otherwise we will have problems improving on our ethical theory. Some of our arguments for and against ethical theories must come
from thought-experiments, or we will not have any solid ground for our choice of moral principles at all.¹⁰

Before we go into my argument why the repugnant conclusion is not repugnant and why, as a matter of fact, we ought to accept it, there are some possible misunderstandings of the conclusion that should be diagnosed and put to one side. Even if these misunderstandings are elementary, they are very common indeed, and it is important to guard oneself against them.

III. SOME MISTAKEN BELIEFS ABOUT THE REPUGNANT CONCLUSION

First of all, we must remember that the repugnant conclusion stresses a mere possibility. Note the reference in Parfit’s statement of it to an ‘imaginable’ population. Even if we accept the repugnant conclusion, we need not endorse all aspects of the actual increase of the world population. The actual increase in the world population that we have seen during the last century, say, may well have meant, at least in some cases, a loss of welfare; and some further actual increases of the world population may mean that even more is detracted from the sum total of well-being. I think here in particular of the ‘addition’ of those children who are born into extreme poverty. It is not far-fetched to assume that these children lead lives that are worth not living and that their addition detracts from the sum total. Their addition does not mean that we are approaching what Parfit has called a ‘Z’ world, where the sum total of well being is maximized. Moreover, the cost (to many existing people) of adding even some lives worth living may be so great, that the sum total of well being is affected for the worse. This may be true of some people living good lives in the rich part of the world. Their lives are worth living but, in living them, these people consume scarce material resources that could have been put to more efficient use elsewhere. Even a person who accepts the repugnant conclusion may be concerned about aspects of the actual population growth, then.

Second, when taking a stand on the repugnant conclusion, we should be careful not to ask ourselves, in what world we want to live, in a world where a few very happy persons live, the A-world, or, in a Z-world where very many, moderately happy persons, live. It is very natural to prefer to live in the world where each person is very happy. This does not answer the question of which of these worlds is the better

¹⁰ This is not the place to defend this claim in detail. I do so, however, in ch. 2 of my Hedonistic Utilitarianism as well as in ‘In Defense of Theory in Ethics’, Canadian Journal of Philosophy, xxv (1995).
one, however. The question 'In which world would I like to live?' is clearly biased. If at all we should discuss the matter from the point of view of what world we would opt for, if we were offered a choice, and were in a Rawlsian manner forced to make our choice behind a veil of ignorance. And somehow our very existence should be at stake in our choice.

Is it possible to take the decision about the number of people who will live behind a veil of ignorance? Derek Parfit has argued that this is not possible:

We can imagine a different possible history, in which we never existed. But we cannot assume that, in the actual history of the world, it might be true that we never exist.\footnote{Reasons and Persons, p. 392.}

Even if this is true, there is a possibility of constructing a suitable contractual situation. We could profit here from our bias for the future. This means that we may hold it to be an open question whether, after the contractual situation, we will continue to exist or not. When my option has been made then either the veil of ignorance is simply lifted and I have to live with my option — or I, the contracting party, am taken away. If I choose the wrong option it may well be that, all of a sudden, I perish. I never get out of the original position.

How in more detail are we to conceive of the probabilities? There is no fixed answer to this question. Following Rawls, at any rate, we should model the conditions of the original position to satisfy our intuitions about justice as fairness. Then I suggest the following. The risk that I will perish is lower if I opt for a more extensive population. If I opt for the most extensive, possible population, I know that I will continue to exist.

Even if this does not settle the matter completely, I find the following line of argument quite plausible: Suppose we have to choose between a world where 10 billion people live very happy lives and a world where \(10^{100}\) billion people live lives that are just worth living. Suppose also that no world with more than \(10^{100}\) billion people is possible. Now, if, behind a veil of ignorance, I opt for a world where 10 billion very happy people live, the probability that I will be one among them may quite reasonably be assumed to be 10 billion to \(10^{100}\) billion, i.e. negligible. If, on the other hand, I opt for the larger population, then I am certain that I will (continue to) live. On this interpretation of the original position, I would not hesitate to opt for the larger population.\footnote{Ragnar Frånén and Niklas Juth have both objected to the probability distribution here suggested. They find the distribution quite arbitrary. I tend to believe that the suggested distribution does reflect our intuitions about justice as fairness. However,
But perhaps it could be objected to this description of the original position that it is biased in favour of this answer because it calls forth our irrationally strong fear of death.\textsuperscript{13} However, there is another possibility of describing the original position. Everyone who would exist if either A or Z came about could be represented by an advocate whose goal is to further his or her interests maximally behind the veil of ignorance. This has the advantage that the decision to make is about whole lives. And presumably, the advocate would then choose Z (both maximin and maximizing expected utility would imply this).\textsuperscript{14}

And, without allowing this piece of reasoning to decide the matter, I am inclined to say that this interpretation of the original position captures the idea of justice as fairness. So the repugnant conclusion is something Rawls, and those who adopt what has come to be known as the ‘priority’ view of justice, should be prepared to accept. And irrespective of whether we accept this piece of reasoning or not, we should be careful to avoid being misled by the question, ‘In which world would I like to live?’, since this question is clearly irrelevant to the acceptability of the repugnant conclusion.

Third, a likely misunderstanding in our apprehension of the repugnant conclusion has to do with the fact that our actual moral sense seems to be based on identification. However, our capacity to identify with others is limited. Most of us care about our own family, and those who are near and dear to us. We take less interest in our fellow countrymen but more interest in them than in people living far away from us. However, it is widely recognized that we ought to care about strangers. We ought to generalize our sympathy even to them. We have extra difficulties in doing so when it comes to very large numbers of people. Very large numbers mean very little to us. However, large numbers do matter. In the same manner that we generalize our sympathy to strangers we ought (mechanically, if necessary) to generalize our sympathy to large numbers of people, even to all the people living in Parfit’s Z-world. If we do we may have to accept the repugnant conclusion after all.

I turn now to my own main argument in defence of the claim that the repugnant conclusion is not repugnant, after all.

\textsuperscript{13} I owe this objection to Stuart Rachels.

\textsuperscript{14} I owe this answer to the objection to Nils Holtug.
IV. THE ACTUAL VALUE OF OUR ACTUAL LIVES

Many people, who have considered the repugnant conclusion, seem to believe that they are living lives that are very well worth living and that an enormous population of people living lives just worth living would look something like a vast concentration camp. I believe this to be a mistake.

The view I am prepared to defend is somewhat pessimistic but still, I am afraid, realistic. My impression is that if only our basic needs are satisfied, then most of us are capable of living lives that, on balance, are worth experiencing. However, no matter how ‘lucky’ we are, how many ‘gadgets’ we happen to possess, we rarely reach beyond this level. If sometimes we do, this has very little to do with material affluence; rather, bliss, when it does occur, seems to be the ephemeral result of such things as requited infatuation, successful creative attempts, and, of course, the proper administration of drugs. Most of the time we spend waiting for all sorts of things and events. We often wait in vain. And when Godot arrives, if eventually he does, he is not always such a great acquaintance to make. Note also that many of the good things in life come with a price tag to be paid in terms of suffering. In many situations, we find ourselves momentarily below the line where our lives are worth living. Moreover, many people probably live lives that, on the whole, are worth not living. When this is acknowledged, the repugnant conclusion does not seem repugnant any more.

Even if this argument is extremely simple and rather obvious, I for a long time believed that I was the first to put it forward in print. However, I now realize that this is a mistake. The argument was previously stated by John Mackie. According to Mackie, in a rarely discussed paper, a possible approach to the problem formulated by Parfit, is to say that a ‘level that is really marginally better than non-existence must already constitute a high degree of flourishing, and beyond this little further improvement is possible’.

This is well in line with how I have myself argued.

Now, if people live lives worth not living, why don’t they kill themselves? The reasons for this may be complex. These people may hold false expectations about the future. Or they may feel obliged to others to stay on for some while, taking care of their children, for example. Or, they may fail to face up to realities. They live in what Sartre used to call mauvaise foi.

From an evolutionary perspective, a tendency to hold on to life is

---

what should be expected of us. For, to be sure, it must have survival
value to cling to life, even if the situation is desperate.

Moreover, we must be careful not to think of a life not worth living
as, necessarily, a terrible life. Even if there is a rather sharp line
between lives worth living and lives worth not living, dropping below
this line need not mean any very significant fall. A life not worth living
may be a life almost worth living. And the difference between a life just
worth living and a life just worth not living is trivial. It is better not to
live a life just worth not living than to live it, but, if one does, it's not
such a big deal. Having lived such a life, when death is approaching, a
person may well believe that he or she has lived a life well worth living.
For, remember, our memory is highly selective. However, if this obser-
vation is correct, we must suspect that our view of the value of the life
we live is often too rosy.

Note also that, when I lead a life just worth living, it is possible that
I drop to a lower state of well-being, without coming to lead a life that
is worth not living. I may come to live a life which is not worth living,
but a life that is also not worth not living. The mere addition of such a
life neither adds to, nor detracts from, the sum total of well-being. And
note that the people who live in the Z-world live lives that are better
than these lives.

If this observation is correct moreover, we should be prepared to
admit that the best world that we could possibly create is probably
very crowded. However, we need not fear that this is a world so
crowded that those who happen to live in it live lives much worse than
the ones we are living right now. Even if they live lives just worth
living, they probably live lives pretty much like the lives we already
live. And by 'we' I denote me who is writing this and you who read
it, i.e. affluent Western people who do not need to worry about proper
schooling for their children, old age, or health care. I do not refer to
people living in abject poverty and true misery, or to people who are
terminally ill and are dying from painful diseases, who, by all prob-
ability, live lives worth not living.

Now, if it is correct to say that the Z-world is a world where every-
body lives roughly at our standard level of well-being, then it tran-
spires that the repugnant conclusion is not repugnant after all. The
intuition that the repugnant conclusion is repugnant withers. This
intuition was nourished by the false belief that, in the Z-world, people
are living lives like the people we Western people see only in certain
television programmes, or when we travel to Africa, people plagued
by famine and infectious disease. When we give up this false belief
we can see the repugnant conclusion as an, if unsought for, yet accept-
able, consequence of hedonistic utilitarianism (and other related 'total'
views).
V. OBJECTIONS

I have come across two main arguments against my view that the repugnant conclusion is not repugnant. Here is the first one, put forward by Nils Holtug.\(^{16}\) This is how he writes about my treatment of the subject:

Firstly, I doubt Tännsjö’s claims about the value of our lives. For instance, it seems to me that my life could be significantly worse than it actually is and yet worth living.

Now, is it true of you and me that we could lead significantly worse lives and yet live lives worth living? I admit that, if this is true, then my position must be wrong. But I do not think that this is true. It is not clear what kind of change of our situation Holtug has in mind, but this is an example I have come across in conversation. Suppose that, by accident, I lose a limb. I will then live a life that is worse than the one I am living right now, but still a life worth living.\(^{17}\)

I am suspicious about this piece of moral psychology. What would happen, I submit, is rather something along the following lines. When the accident has just happened, I am taken over by despair. For a while I live a life worth not living. This does not mean that I ought to end my life. For, as soon as I have adapted to the new situation, I have probably developed a new image of myself, developed a different character, and made my interests conform to my new capabilities. All of a sudden I find that my life is different, all right, but just as good as the one I used to live before the accident.

This is not to deny that things can happen to me, such that I end up with a much lower quality of life, and where I am not capable of adapting to the new situation. I think of examples such as when I catch an incurable and progressive disease with (painful) symptoms. But then I do not find it far-fetched to assume that I have come to live a life worth not living, i.e. a life that is, from my personal point of view at least, worth ending (which does not mean that I should end it, of course; it may be of importance to my children, say, that I stay around for a while).

It may be retorted to this that my answer is not really an argument, but just a flat denial that we can live lives that are, hedonistically speaking, worse than the lives we live, without our lives being worth not living. What kind of argument could settle the matter, then? What we should do, it seems to me, if we want to know more about this kind of moral psychology, is to ask people who know what they are talking

---


\(^{17}\) I owe this example to Włodek Rabinowicz who has stated it in conversation.
about. We should ask people who have acquired various different kinds of disabilities how they consider the value of their lives, as compared to the situation before they acquired their disabilities.

As a matter of fact, having often been involved in discussion with members of the disabled people's organizations, I have sometimes asked people with acquired (not congenital) disabilities how they assess the quality of their lives, before and after the accident, and I have always received the same answer to my question: 'When the accident had just occurred, I thought it was all over with me. I felt that my life had no meaning any longer. However, once I had adapted to the new situation, my life turned out to be no worse than it was before I acquired my disability. I live a different life now, but not a worse life.'

I tend to believe that this answer is both sincere and true. Note that this answer seems to fit into a realistic evolutionary picture as well. If our 'normal' hedonistic situation is one where we keep our heads just above the water, this provides the best possible incentive for us to carry on with our swimming. I admit, however, that there is room for further research here.

Now, having put forward this argument in many different contexts, I know that there are some people who want to stick to a different picture after all. These people are telling different stories about their lives, indicating that they may have a lot to lose while, for all that, living lives well worth living. Some of them tell me about their past, when they used to live much better lives than the ones they live right now. They have lost their health, their loved ones, and so forth. Yet, for all this, they still live lives well worth living. Even so, these people should pause to think, it seems to me. They should acknowledge that it is indeed very difficult to plot one's own life on a scale upon which well-being is assessed. It may be true that, on the whole, their lives are very much worth living. And yet, for all that, even these people, I submit, must be somewhat reluctant when they put forward this claim. Even these people should feel that, possibly, they did not lead such great lives in the past. Possibly, they live now at roughly the level at which they used to live, having lived for a while in between a life worth not living. It is indeed difficult to plot one's life securely on the level of well-being assessed on the relevant scale.

It should perhaps also be noted that there are some people who tell quite a different kind of story about their lives, indicating that what we may call my 'near-zero' hypothesis is much too optimistic. They suspect that they themselves, and most other people as well, live lives below zero.¹⁸ And they cite the same kind of 'negative' evidence as I do,

¹⁸ David Benatar and an anonymous reviewer for Utilitas have both made this very helpful comment.
to the effect that we should not trust our preferences on this matter. I concede that what we can call the ‘below-zero’ hypothesis is compatible with this hypothesis. For reasons to do with evolutionary biology I personally have difficulties accepting it. However, this very difficulty in plotting our lives on the scale is an argument as such, I submit, why the repugnant conclusion cannot really be obviously false. The graphic representation of the A-world and the Z-world respectively may be too much of an idealization. For all we know, our lives may be close to the ones lived by people inhabiting the Z-word, and the lives of the people inhabiting the A-world may be out of reach of our moral imagination. We may even be living below zero. And, even if we happen to live lives somewhat above the level on which the people in the Z-world live, the lives of the people inhabiting the A-world may be so different from the ones we live that we cannot really identify with them. The usual graphical representation of the case indicates that the people in the A-world lead lives many times as happy as the lives lived by the people in the Z-world. Can we really make sense of this claim? Remember that the claim that the repugnant conclusion is obviously false is a very strong one indeed, so the negation of it should not be that difficult to embrace.

I now turn to the second objection. It too has been raised by Holtug: Secondly, it seems to me that he [Tännjö] does not really address what is repugnant about the repugnant conclusion. What is repugnant is the combination of two features; people’s lives in the larger population are barely worth living, and people’s lives in the smaller population are much better. Tännjö only addresses the former feature, and this feature, taken by itself is not repugnant (whatever ‘barely worth living’ may turn out to involve, people’s lives are worth living in the larger population). So it seems to me that Tännjö does not explain away the repugnancy.

A similar objection has been put forward by Gustaf Arrhenius. This is how Arrhenius puts it:

The unacceptability of the Repugnant Conclusion doesn’t depend on the welfare of actual people. It is surely a logical and nomological possibility that people could enjoy very high welfare and we have no problem imagining such lives ... Let’s assume that Tännjö et al. are right and that the current world population consists of people with very low positive welfare. Which of the following two futures would be the best? In the first scenario we have a massive expansion of the population size but all the people still have very low positive welfare. In the second scenario, the population size remains the same but we have a major increase in people’s welfare such that everybody enjoys very high welfare. The answer seems obvious.¹⁹

As we saw, in his second objection Holtug concedes that there is nothing repugnant as such about a very extensive population with

lives just worth living, and I take Arrhenius to concur in this assessment. This concession is very important, it seems to me. For, as I have insisted above, I do have the uneasy suspicion that some people who find the repugnant conclusion repugnant do so because, for snobbish reasons, they hold a very extensively populated world, with people leading lives just worth living, in contempt. So I am happy to notice that this is not how Holtug or Arrhenius conceives of the situation. They both agree that the people living in the Z-world lead lives worth living, and Arrhenius does concede, at least for the sake of argument, that these lives may be pretty much like the ones we live. What is repugnant, they claim, is not the Z-world as such, but the claim that such a world could be better than a world with a more restricted population living lives very much worth living (with creatures living lives much better than the ones we live). We should focus on the comparison between the two worlds.

Now, I am prepared to focus on the comparison, once I have noted that the value of each of the two worlds we are being invited to compare must be of relevance to the outcome of it. And note that what has been considered repugnant by Parfit et al. is not the mere fact that a world with moderately but very many people, can be better than a world with not so many, but even happier persons. Some may find this wrong but no one would find it repugnant. It is the comparison with the last, Z-world, in Parfit’s series of comparisons that has been singled out in the discussion. What has been considered repugnant is that such a world could be better than the A-world, a less populated world with very happy people. But note that if these very happy people are too few, the alleged repugnance seems to evaporate once again. Remember that, according to Parfit, it takes a world of 10 billion very happy people to establish the repugnance of the repugnant conclusion. When all these clarifications have been made, and once it is conceded that a Z-world may be pretty much of the same kind as our actual world, I cannot help feeling that the repugnant conclusion is not at all repugnant, i.e. it is far from obviously false that the Z-world may be the better one.

Note also that the claim made by Arrhenius, that it is a logical and nomological possibility that people could enjoy very high welfare and that we have no problem imagining such lives, is problematic. It is certainly a logical and nomological possibility that there could be some sort of beings capable of leading much better lives than the ones we live, but it is far from certain that these lives would be recognizably human lives. But then it becomes difficult for us to identify with them, and once again our moral intuition falters.

Or, even if they are human, it is doubtful whether we can really imagine what such lives would be like to live, for those who live
them. Are we to imagine a world where people use innocuous drugs, enhancing the quality of their lives, without destroying their capacity in the long run for leading good lives? Or are we to imagine a state of infatuation that never goes away or turns into ordinary love?

But I have conceded that our lives consist of bad moments, neutral moments, and good ones. Why not then conceive of the A-world, with very happy people, as a world where everybody lives a life composed exclusively of days just like our best days?²⁰

Of course, we could perform such an abstract thought-experiment. However, once again, the kind of life we are now imagining does not strike me as a recognizably human life. I have difficulties in identifying with the creatures living, from the point of view of their hedonic status, such homogenous lives. How is it possible to feel happiness, if you have made no sacrifices in order to get there? How is it possible to sense the joy of a successful achievement, if you did not have to strive in order to make it? Or, to focus on the Z-world, how is it possible to stay in the neutral state, without dropping below it, when there is no improvement to look forward to?

One way of making sense of the thought-experiment would be to think of the very happy people as each day falsely believing that, the day before, they made great sacrifices, in order to achieve what they now feel they are achieving. I suppose we have to imagine something of this sort, if I am correct when I claim that we ordinary people rarely rise behind a level where we live lives just worth living. But is this world the better one? I think many will hesitate to judge that it is. Quite to the contrary, they find this world scary. Their reluctance to endorse this kind of world may reflect a mere anti-hedonistic prejudice, of course. But even if it does, it is far from obvious that this world is the better one. The A-world seems to be inhabited by a kind of utility monsters. And we are at a loss when it comes to assessing the relative value of their lives.

But Arrhenius has also put forward another version of the repugnant conclusion, which he thinks I should at least find repugnant:

Tännsjö is a hedonistic total utilitarian. The welfare of a life is determined by just summing the utilities of the happy and unhappy moments in life. Consider a population that consists of very short lives, say a minute of slight happiness. According to Tännsjö’s theory, these lives enjoy positive welfare. It is hard to deny that such lives have very much lower welfare than the lives led in the privileged parts of the world. Irrespective of whether there are possible lives with very high welfare, Tännsjö’s theory implies the following recasting of the Repugnant Conclusion: For any perfectly equal population with the same welfare as the people in the privileged parts of the world, there is a

²⁰ This has been suggested in correspondence by Stuart Rachels.
population of lives consisting of just one minute of slight happiness, which is better.21

Again, when considering an example such as this one, I must admit that I feel not only that common-sense morality fails us, but that my own moral intuition falters. What are we to say about this example, then? I am inclined to say that, being human, I cannot conceive of this kind of momentary life. At least it does not strike me as an example of a human life. It is certainly true that we human beings can feel happiness for just a short moment. However, unless a human being lives for quite a long time, it is not possible for this being to develop the capacity for such experiences. And in order to lead a life that is, on balance, worth living, I must both experience moments that are, as such, worth not living, and other moments that are, as such, well worth living. A life that goes on for only one minute is not a recognizably human life.

However, this should not settle the matter. Perhaps there are creatures who can experience happiness in a life that goes on for just one minute. Then, if similar creatures are replacing them, once they cease to exist, the result may well be a world with a considerable sum total of well-being. What are we to say of a world where many individuals live lives like this and experience each a moderate amount of happiness?

I would say that, if the sum total of well-being in this world is high enough, it would be an example of objectionable ‘speciesism’ not to be prepared to ‘abdicate’ in the interest of this kind of world. Once again, we must guard against the mistake of thinking: In which world would I like to live? In this case, the better world is not meant for creatures like you and me.

It is hard to imagine what it would be like to be a creature who lives happily for one minute, but this is clearly beside the point. And, to make a more realistic interpretation of this last question: What if, through genetic engineering, we could prolong our life-spans, in a manner that makes our ordinary life-spans look like brief moments in comparison? What if we could thus, without improving the momentary quality of any life, make each of our lives contain a more substantial amount of well-being? What if the alternatives are, on the one hand, for us not to let go but to stay on and live very long lives, and, on the other hand, to allow nature to take its course (so that we die of old age) making room for an enormous lot of people, each living a life of the kind we do? We assume the momentary quality of life to be the same in both scenarios, and we assume the sum total of well-being to be larger in

the scenario where many people live shorter lives (i.e. lives of the kind we live right now). How should we choose?

On egoism, we would have good reasons of course to cling to our lives. And those who find the repugnant conclusion repugnant may concur. At last we have found a realistic way of seeing how we could improve the value of the lives of existing people. By prolonging our lives we make them contain much more happiness on the whole. In our thought-experiment we abstract from boredom, of course. The assumption that we will go on with our lives without becoming bored is not realistic, I think, but an assumption that does not destroy the example. Even if we doubt that we will not be bored to death after a few hundred years, we can imagine and understand what it would be like, going on with our lives without becoming bored. Those who find the repugnant conclusion obviously repugnant must feel that a world where existing people accumulate, in each life, much more well-being, than the people in the other population, who live much shorter lives, is preferable. But is it?

I find this in no way obvious. Note that, when we live for a very long time, it is plausible to assume that, eventually, we will lose all our memories of earlier parts of our lives. In order to stay the same persons there must be some continuity in our mental development, of course, but such continuity is certainly consistent with a gradual loss of memory of certain parts of a life. And without a considerable amount of loss of memory we would become bored, I submit. But then, what's so great with a long and happy life, even from an egoistic perspective? And note furthermore that what we are here discussing are problems of moral philosophy, not of prudence. Once again we must guard ourselves against asking: 'In what world would I like to live?'

There are no firm intuitions to rely on with respect to thought-experiments like the one just constructed. It may even be that some feel that we are here touching upon an intra-personal version of the repugnant conclusion. Or so those may be tempted to say who find the repugnant conclusion in its standard version repugnant. Perhaps they would say that a life that contains less well-being, because it goes on for a shorter time, may yet be better than a much longer life, containing a greater sum total of well being, provided the former life is of a much better momentary quality?

I do not know. But I do feel that now things are getting messy. It would be premature to claim that any position in the discussion is obviously the correct one. And personally I tend to side here with total hedonistic utilitarianism. But this indicates that, if possible, we should accept a densely populated world where every sentient being, who is living, lives only for one minute, but lives for this minute a life worth living. We should not only accept such a world, but also prefer
this kind of world to any world containing a lesser sum total of well-being, even one containing many people like you and me. If, by pressing a button, I could transform the existing world into a world sparkling with this kind of moderately happy life I would for egoistic reasons hesitate to do so, but I doubt that, from a moral point of view, I would have any good reasons to hesitate. Be this as it may, it is by now obvious that there is no obvious answer to the question what to do in a situation like this one.

VI. IN DEFENCE OF THE REPUGNANT CONCLUSION

It seems to me that my (and Mackie’s and Ryberg’s) argument has survived the criticism it has been submitted to. The repugnant conclusion is not repugnant, after all. The conclusion is not obviously wrong. We see that this is so when we realize that a life just worth living may be of roughly the sort as the one we already live. Therefore, the ‘repugnant’ conclusion is a misnomer. No moral principle should be read out of court simple because it can be shown to imply the repugnant conclusion. But if this is true, then, if we find that there are good reasons to do so, we should also accept the repugnant conclusion. And it is not difficult to find good reasons in support of the repugnant conclusion. The best arguments in defence of the conclusion are well-known from, or at least hinted at, in the recent literature on population ethics. I extract one of them from Arrhenius’s dissertation. The second can be found in Parfit’s book Reasons and Persons and the third one is very close to an argument in the same book.

Parfit takes his arguments as providing us with ‘paradoxes’, and Arrhenius takes his argument to be part of an ‘impossibility theorem’ for population ethics, but in this they are both mistaken, relying on the false assumption that the repugnant conclusion is really repugnant. If we let this false assumption go, then, indeed, what we are faced with are three very simple and extremely good arguments to accept the repugnant conclusion.

Let me start with the argument I extract from Arrhenius. It can be found in one of his ‘impossibility’ theorems for population ethics. As a point of departure for one of these theorems, Arrhenius states what he calls ‘the quantity condition’. According to this condition, for any pair of worlds A and B, with n people in the A-world, and m people in the B-world, all living lives worth living, but where the people of the A-world lead slightly better lives than the people of the B-world, the B-world is at least as good as the A-world, provided the number m is large enough as compared to n.

We should accept this, I submit. And we should even strengthen the condition and accept that the B-world referred to is better than the
A-world. Then, of course, through successive reasoning we can go from a world with 10 billion people leading extremely good lives (because their lives are extremely long, for example), to a better world with a much more extensive population leading only somewhat worse lives (only slightly shorter lives, say), right down to the Z-world (where an enormous population lead lives just like the ones we live right now). And this piece of reasoning carries us up to the very last letter in our moral alphabet, the world Z.

Why ought we to accept the (strengthened) version of the quantity condition? I feel tempted here to quote Genesis. Even if Adam and Eve were leading fantastic lives in the Garden of Eden, the world was not perfect. Not only could the world have been made better through mere addition of people, if God had bothered to create more of our kind. Even at some cost, Adam and Eve themselves should replenish the earth.

The very idea of a universe without sentient life strikes us as terrible. A world with human life, and other kinds of sentient life on Earth only, is better than a universe with no life at all. Yet we still gaze at the starry heavens above us and fear that we are alone. All those stars, all that matter (and anti-matter) and all that energy, but no sentient life, we fear. Such a waste of material resources! How utterly meaningless if such a vast universe were to exist only to host, for a short moment, us! We hope to find life elsewhere. And if there is none, we will not rest satisfied, I conjecture, until we have colonized as much as we can of the universe. No sacrifice will seem too hard for this purpose. In this cosmic perspective, even very large numbers such as $10^{100}$ begin to make some sense to us. We can conceive of them as an area of space that has been successfully colonized.

Is this endeavour reasonable? I think it is, at least it is up to the point where we are beginning to make sacrifices in order to add lives that are not worth living. The total view explains and rationalizes this (somewhat moderated) human zeal. To the extent that we add creatures living lives worth living, our ambition to replenish the universe not only is part of our quest for meaning, but also means that we comply with our duties as moral agents.

In my opinion, the argument just stated is completely convincing. However, some may feel that it is not. They may then turn to the next argument, put forward by Parfit. Judging from the aforementioned paper by Mackie, we find this argument in Parfit’s original statement of what he has called ‘the mere addition paradox’, and he mentions it in passing also in Reasons and Persons. The argument has two premises and leads to the repugnant conclusion:

---

22 Ibid., p. 430.
(1) The mere addition to the world of a person leading a life worth living makes the world better.

(2) By levelling out inequalities in a manner that improves the sum total of well being, since those who are worse off gain more from the redistribution than those who are better off lose from it, means, if all other things are kept equal, an improvement.

By 'mere' addition of a person we mean the addition of a person to an existing population, which in no way affects the well-being of the existing population. These two premises lead to the repugnant conclusion. Parfit himself shows this in an informal way with the help of the following well-known figure:

### 142. MERE ADDITION

**Consider these alternatives:**

![Diagram](image)

You start with A, a world of 10 billion very happy persons (living, for example, extremely long lives of the same quality as the ones lived by you and me), add some people and arrive at world A+, which is better than A (because of the truth of (1)). You now compare A+ with divided B and find that divided B is better than A+ (because of the truth of (2)). But divided B is no worse than B so, because of the transitivity of 'better than', B is better than A+ and therefore also better than A. You now take B as your point of departure in a new comparison and repeat the argument. Eventually you will end up with Z, or this is what Parfit claims, at any rate. The truth seems to be that you can carry the argument on *almost* to the point where you reach Z, but not right to this point, at least not as Z has been conceived of here (where people lead lives that are *just* worth living). For if people live lives *just* worth living, the possibility of gaining total welfare by adding people who are *less* well off does not exist, so the very last step in the argument does
not go through. Perhaps this explains why Parfit does not speak of lives 'just' worth living but prefers to speak more loosely of lives that are 'barely' worth living. However, if the argument can be carried on almost up to Z, then it is hard to believe that the last step is so special. So the argument does indeed indicate even that Z is better than A.

Are (1) and (2) true, then? I think so. I think few would object to the truth of (2). But even (1) is true. As a matter of fact, I must say that here I find the word 'obvious' to be in place. (1) is obviously true. I even find the denial of (1) morally repugnant. If the addition of one person does not affect the well-being of existing persons, why deny this additional person a life?

It is certainly true that, if we do not add this person there is no one there to complain. Yet, for all that, if we do add this person, and ask him or her if he or she is grateful for having come into existence, the person will probably answer in the affirmative. This may have something to do with a tendency to overestimate the quality of the life we lead, of course, but, in this case, the feeling of gratitude is perfectly well placed. This person does live a life worth living (like the one you and I live). Such a want of generosity, if we don't welcome such a creature!

In a similar vein, if we do not create a person who would have led a terrible life well worth not experiencing, there is no one there to thank us. Yet, for all that, we were right in not creating this person. It would have been wrong to do so. Some people have wanted to avoid this very natural moral symmetry — because they have wanted to avoid the repugnant conclusion. They have argued that, while we must not create a person who lives in misery we need not create a person who would live a life worth living. Once we realize that the repugnant conclusion is not repugnant, however, we may safely hold on to the symmetry.\textsuperscript{25}

Again I find the argument just stated very convincing. However, those who do not share this intuition may consult the third and last argument, also put forward by Parfit.

I suppose that those who do not accept the argument just stated fail to accept it because they are reluctant to accept (1). They are not prepared to grant that the mere addition of happy people to the world makes the world a better place. But Parfit has shown that we can do without (1). The following is not exactly a restatement of Parfit's argu-\textsuperscript{25} Not all philosophers have sought to avoid the moral symmetry. Many have been prepared to accept it. And Nils Holtug has argued convincingly that we ought to accept it. Cf. for example his 'Utility, Priority, and Possible People', \textit{Utilitas}, xi (1999), and 'On the Value of Coming into Existence', \textit{Journal of Ethics} (forthcoming).
ment, but something roughly to the same effect, and somewhat simpler and more complete.\(^2\) Consider the following picture.

\[\text{A-world} \quad \text{B-world} \quad \text{C-world} \quad \text{...}\]

\[\text{A-people} \quad \text{B-people} \quad \text{C-people}\]

Once again we start with the A-world, where 10 billion people lead extremely good lives (the A-people). We now add 10 billion people, the B-people, leading lives not as good but very good nevertheless (much better than the ones we live). This addition of the B-people is not an example of 'mere' addition, however. It so happens that the addition of the B-people has the effect that the A-people come to live even better lives. This means improvement.

It is certainly true that, by adding the B-people we introduce inequality into the world. However, few would object to the introduction of this kind of inequality. The A-people profit from it. And the B-people have their existence to thank for it.

We now consider the B-world and realize that, if we add another 20 billion people to it – we may speak of these additional people as the C-people, leading lives well worth living, but not as good as the lives lived by the B-people – we will improve the lot of the B-people. Now

\(^2\) The argument draws on pp. 433–41 of Reasons and Persons. However, in Parfit’s statement of the argument, the first step (from A to B) is omitted; hence he speaks of the argument as leading to a ‘new’ repugnant conclusion. The argument supports the original conclusion as well, however. The argument has rarely been discussed, mainly because it is presented in a very complicated manner by Parfit himself. However, the argument is discussed by Stuart Rachels, in his ‘A Set of Solutions to Parfit’s Problems’, Nous, xxxv (2001), 227–9. In order to resist the argument Rachels jettisons the idea that ‘better than’ is a transitive relation, a price too high for anyone who, like the present author, believes that, if one state of affairs is better than another state of affairs, then this is so because there are definite amounts of value in the two states and more of value in the former than in the latter.
there will be a cost, paid by the A-people. However, the gain to the B-people is larger than the loss to the A-people, and it means that the difference between the A-people and the B-people will be levelled out. Then, according to (2), this change means (as such) an improvement. And it is bought only at the 'cost' of introducing people who are perfectly happy to exist. But this cost is well worth paying.

It is obvious how this argument could be carried forward, down the moral alphabet, almost to the Z-world. And once again it is hard to believe that there is something very special about the very last step. But this means that we have an extremely good reason to accept the repugnant conclusion.

VII. CONCLUSION

The repugnant conclusion is not repugnant, i.e. it is not obviously false. We realize this when various possible misunderstandings have been set to one side and when we have acknowledged that a life just worth living is not very different from the one we already live. However, having acknowledged this, we ought to admit also that, if we find good reasons to accept the repugnant conclusion, we should do so. But there exist some very strong arguments in defence of accepting the repugnant conclusion. Rather than thinking that it be a desideratum, then, that a plausible moral theory, when applied to problems of population ethics, should not imply the repugnant conclusion, we should be suspicious of any putative moral principle that does not have this implication.26

Torbjorn.Tannsj@philosophy.su.se

26 Many people have given helpful comments on this paper. I owe special thanks to Ragnar Francén, Nils Holtug, Niklas Juth, Hans Mathlein, Stuart Rachels, and Folke Tersman.