The Repugnant Conclusion: An Assessment and Critique of Utilitarianism as Applied to Future Populations

A thesis presented to
the faculty of
the College of Arts and Sciences of Ohio University

In partial fulfillment
of the requirements for the degree
Master of Arts

Andrew Ryan Douglas Stauffer

August 2012

© 2012 Andrew Ryan Douglas Stauffer. All Rights Reserved.
This thesis titled

The Repugnant Conclusion: An Assessment and Critique of Utilitarianism as Applied to Future Population

by

ANDREW RYAN DOUGLAS STAUFFER

has been approved for

the Department of Philosophy

and the College of Arts and Sciences by

_____________________________

Alyssa R. Bernstein
Associate Professor of Philosophy

_____________________________

Howard Dewald
Interim Dean, College of Arts and Sciences
ABSTRACT

STAUFFER, ANDREW RYAN DOUGLAS, M.A., August 2012, Philosophy

The Repugnant Conclusion: An Assessment and Critique of Utilitarianism as Applied to Future Populations

Director of Thesis: Alyssa R. Bernstein

The Repugnant Conclusion is a problem for utilitarianism put forward by Derek Parfit that addresses population size and future persons. For any population made up of happy individuals, there is a much larger population made up of barely happy individuals. On total utilitarianism, the latter population is better than the former population, so long as the latter has greater total utility. In this thesis I examine and critique three responses to Parfit’s problem and argue that average utilitarianism solves the problem. First, I examine Tännsjö’s solution to the problem, which is to accept it. Second, I investigate Wolf’s negative utilitarianism. Third, I consider and defend on behalf of the utilitarian Neumann’s average utilitarian response. Last, I give non-utilitarian arguments for rejecting average utilitarianism, focusing on reproductive rights.

Approved: _____________________________________________________________

Alyssa R. Bernstein.

Associate Professor of Philosophy
ACKNOWLEDGMENTS

I would like to give special thanks to my thesis committee members which include Dr. Bernstein, Dr. Lebar, and Dr. Carson, the professors in the department of philosophy, and especially my fellow graduate students.
# TABLE OF CONTENTS

| Section                                                                 || Page |
|------------------------------------------------------------------------|------|
| Abstract                                                               | 3    |
| Acknowledgments                                                        | 4    |
| List of Figures                                                        | 6    |
| Chapter I: Introduction: The Repugnant Conclusion                      | 7    |
| Solutions                                                              | 10   |
| Chapter II: Tännsjö’s Acceptance of the Repugnant Conclusion           | 11   |
| Tännsjö                                                                 | 11   |
| The Actual Value of Our Actual Lives                                  | 16   |
| Objections                                                             | 17   |
| Conclusion                                                             | 23   |
| Chapter III: Wolf                                                       | 24   |
| Popper, Obligation, and Beneficence                                    | 28   |
| Popperian Consequentialism and the Repugnant Conclusion                | 30   |
| Conclusion                                                             | 34   |
| Chapter IV: Average Utilitarianism                                     | 35   |
| Parfit on Average Utilitarianism                                       | 36   |
| Neumann on Average Utilitarianism                                      | 38   |
| Hurka and Anglin on Average Utilitarianism                            | 43   |
| Conclusion                                                             | 47   |
| Chapter V: Non-Utilitarian Reasons for Rejecting Average Utilitarianism| 49   |
| Unacceptable Conclusions                                               | 49   |
| Sidgwick                                                               | 51   |
| Brink                                                                  | 55   |
| Conclusion                                                             | 60   |
| Works Cited                                                            | 61   |
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure I</td>
<td>Populations A and B</td>
<td>7</td>
</tr>
<tr>
<td>Figure II</td>
<td>Populations A and Z</td>
<td>9</td>
</tr>
<tr>
<td>Figure III</td>
<td>Killing and Utility Levels</td>
<td>39</td>
</tr>
</tbody>
</table>
CHAPTER I: INTRODUCTION: THE REPUGNANT CONCLUSION

In *Reasons and Persons*, Derek Parfit argues that total utilitarianism, the view that “the best outcome is the one in which there would be the greatest quantity of happiness—the greatest net sum of happiness minus misery,”\(^1\) leads to what he calls the Repugnant Conclusion. This is, “for any possible population of at least 10 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.”\(^2\)

Parfit explains the Repugnant Conclusion in detail. In Fig. I: Populations A and B below, Parfit compares “the outcomes of two rates of population growth.”\(^3\) These are two possible future populations.

---

\(^2\) Ibid. 388
\(^3\) Ibid. 385
The quality of life (or how well off the population is) is shown vertically and the number of people in the population is shown as the width of each block. In Fig. I, A has half as large a population as B, but the population as a whole is better off, on average. When Parfit writes of the people in B being less well off than people in A, he means that “[t]here might be worse housing, overcrowded schools, more pollution, less unspoilt countryside, fewer opportunities, and a smaller share per person of various other kinds of resources”. This refers, loosely, to “the quality of life”.

Which population should a utilitarian try to bring about or favor? In one way, A would be better than B. This conforms to the moral intuitions of most people. It would be better for the individuals in population A since they would all be better off. This is implied by the hedonistic version of the impersonal average principle, the view that “the best outcome is the one in which there is the greatest average net sum of happiness, per life lived.” Nonetheless, Parfit and others quickly dismiss average utilitarianism, for reasons to be given below, in favor of total utilitarianism. However, this leads to the Repugnant Conclusion.

According to total utilitarianism, population B in Fig. I would be better than population A. And another slightly larger population, C (not illustrated), containing people with lives slightly less worth living, would be better than B. This reasoning would continue until we reach a population Z that is very large and is made up of individuals whose lives are barely worth living. Under total utilitarianism, Z would be

---

better than A. This is, essentially, the Repugnant Conclusion⁶ and is illustrated in Fig. II: Populations A and Z below.⁷

![Figure II: Populations A and Z](image)

The repugnancy of this conclusion arises from the idea that an immensely large population made up of individuals with lives that are barely worth living (Z) is better than a much smaller population made up of individuals with lives very much worth living (A). Parfit describes the Z-lives as “painless but drab.”⁸ He writes,

> There is nothing bad in these lives; but there is little happiness, and little else that is good. The people in Z never suffer; but all they have is muzak and potatoes. Though there is little happiness in each life in Z, because there are so many of

---

⁶ In *The Methods of Ethics*, Henry Sidgwick takes a similar route as Parfit regarding population. He writes, “the point up to which, on Utilitarian principles, population ought to be encouraged to increase, is not that at which average happiness is the greatest possible . . . but that at which . . . happiness reaches its maximum.” Henry Sidgwick, *The Methods of Ethics*. Hackett, 1981. 415-416.


these lives Z is the outcome in which there would be the greatest total sum of happiness. Similarly, Z is the outcome in which there would be the greatest quantity of whatever makes life worth living.\(^9\)

There are a number of reasons why the Repugnant Conclusion is repugnant. For example, the idea that we make the world a better place by adding more people instead of alleviating the suffering of those already alive is repugnant. This kind of calculation seems to misapply the utilitarian calculation of the value of individual lives. The focus is on the overall utility of the population as a whole and not on the value of individual lives that make up the population.

In this essay I aim to 1.) clarify and respond to three possible utilitarian solutions to the Repugnant Conclusion, concluding that average utilitarianism is sufficient for solving the problem, and 2.) consider non-utilitarian arguments against the average utilitarian solution, specifically, one which has to do with the freedom to refrain from reproducing.

**Solutions**

The three utilitarian solutions I propose to examine and critique are a) the acceptance of the Repugnant Conclusion by redefining a life worth living (Tännsjö), b) the view that the minimization of suffering, not the maximization of happiness, is the main utilitarian goal (Wolf), and c) average utilitarianism (Neumann). I will argue that the former two utilitarian proposals cannot adequately resolve the problem of the Repugnant Conclusion. However, average utilitarianism can solve the problem.

\(^9\) Ibid.
CHAPTER II: TÄNNSJÖ’S ACCEPTANCE OF THE REPUGNANT CONCLUSION

One way to address the Repugnant Conclusion is to accept it. Instead of trying to elude the Repugnant Conclusion, total utilitarians can simply concede that it follows from their moral theory. This is the route taken by Torbjörn Tännsjö, a hedonistic total utilitarian, who argues that the Repugnant Conclusion is really not as bad as Parfit and others make it out to be. While his views are ultimately flawed, he provides useful insight into how utilitarians can, or need not, view the Repugnant Conclusion. I consider his arguments next.

Tännsjö

In “Why We Ought to Accept the Repugnant Conclusion”, Tännsjö writes, “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.” He prefaces his argument by emphasizing three mistaken beliefs that many philosophers have, which contribute to faulty thinking about the Repugnant Conclusion. Once these mistaken beliefs are corrected and the actual value of our actual lives is taken into consideration, the repugnancy of the Repugnant Conclusion disappears, or so Tännsjö claims.

The first common mistake in reasoning, Tännsjö writes, is failing to “remember that the repugnant conclusion stresses a mere possibility... Even if we accept the repugnant conclusion, we need not endorse all aspects of the actual increase of the world.

---

Comparing multiple populations is different from deliberating about which should be brought about. When comparing two populations, utilitarians are not considering how the populations might have arisen; they are simply comparing hypothetical populations. Even if Z is better than A, utilitarians must consider what would be at stake if a policy were proposed for bringing about population Z. So it is worth considering Tännäsjö’s distinction between accepting the Repugnant Conclusion in theory and actually carrying out the policy based on that theoretic judgment.

However, utilitarians may not be able in all cases to defend the distinction between possibility and actuality. Presumably they would take all possible factors, such as the effects on the environment, the treatment of women, and the reactions of people, into consideration when calculating utility. Nonetheless, when utility would, if the goal were reached, be increased, the means to attain that goal are justified on utilitarian grounds. On a total utilitarian view, we would be obligated to bring about or produce a certain population if in doing so we would be increasing total utility. So, Tännäsjö’s view that the Repugnant Conclusion is merely a possibility and does not require anything of the utilitarian in practice is not as helpful for the total utilitarian case as he thinks it is.

The second allegedly mistaken belief about the Repugnant Conclusion that Tännäsjö addresses has to do with unbiased choices about existence in possible populations. He states that, “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

\[\text{Ibid. 221}\]
happy persons live.”

12 On Tännsjö’s account, although I would be happier in the A-world, this does not mean that the A-world is better than the Z-world. Tännsjö writes, “The question ‘In which world would I like to live?’ is clearly biased.”

13 What we ought to do is to “think of ourselves 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.”

14 It would be better for me, behind the Rawlsian veil of ignorance, if I were to choose the larger population where everyone is moderately happy. This is so because I am more likely to be brought into existence with this population. As Tännsjö writes, “The risk that I will perish is lower if I opt for a more extensive population.”

15 Parfit objects to this line of reasoning. When addressing how the veil of ignorance might be used in reasoning about population policy, Parfit considers whether we (the parties behind the veil of ignorance considering populations) should assume that we will exist regardless of which population we pick. If we assume that we will exist regardless of the population, it would be best for us to choose population A. However, this assumption violates the requirement of impartiality which the veil of ignorance is supposed to support.

16 Parfit writes, “If we assume that we shall certainly exist whatever principle we choose, this is like assuming, when choosing a principle that would disadvantage women, that we shall certainly be men.”

17 However, if we do not assume that we will exist regardless of the population we choose, it is better if we choose

---

13 Ibid.
14 Ibid. 221-222
15 Ibid. 222
17 Ibid.
population Z instead of A, because we are more likely to exist in the Z-population. Is there good reason to make a decision about population size while assuming that we do not know whether we will ever exist? Parfit thinks not. He states, “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. We therefore cannot ask what, on this assumption, it would [be] rational to choose.”

Parfit writes that when we are discussing this way of determining which population is better, the veil of ignorance “is not impartial unless we imagine something that we cannot possibly imagine.”

As a response to Parfit’s argument, Tännsjö alters the thought experiment. He notes that even if Parfit’s argument is true, we can imagine a different suitable contractual situation. He writes,

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 make the wrong option it may well be that, all of a sudden, I perish. I never get out of the original position.

On Tännsjö’s view, given these circumstances, one would do best to choose population Z because there would be more of a chance of further existence. But while this may avoid Parfit’s criticism that we cannot assume a history of the world in which I do not exist, this thought experiment does not make sense. It is not clear what anyone would be agreeing to, nor is it clear when or why there would be sufficient reason to use such a veil of ignorance. So, I turn to the third of Tännsjö’s considerations.

---

18 Ibid.
19 Ibid.
Tännsjö points out that it is difficult for us to identify with large numbers of people. He writes, “Very large numbers mean very little to us. However, large numbers do matter.”21 We care about people close to us, like family, friends, and neighbors. We can sympathize with them more easily than we can with large groups of people. That much is understandable. By ‘large group’, I take Tännsjö to mean any number of people larger than the number one interacts with on a daily basis. This could range from the number of people in a small town to the number of people in population Z. ‘Large’ is a relative term here. Tännsjö continues, “In the same manner that we generalise our sympathy to strangers we ought to generalise 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.”22 If Tännsjö’s point is that people tend to care less about people in population Z simply because there are so many of them, this seems to go against his argument. It seems that people are repelled by population Z because they care about the people in that population. But perhaps Tännsjö intends to show that the mere fact that population Z has more people in it than population A does not entail that the people in Z are any less morally important. However, this point would seem to support a different idea, namely, that utilitarianism is concerned with individual people rather than increasing the aggregate utility of the population, and so would not imply an obligation to bring about a Z-size population. Tännsjö does not make his point very clear, but I take it that he is saying that we should try to expand our moral circle. To this, most moral

---

21 Ibid. 223
philosophers can agree. However, this point does not relate specifically to the problem of the Repugnant Conclusion.

Tännsjö’s argument as to why we should accept the Repugnant Conclusion is now somewhat clearer. Even if the above clarifications are not all accepted, nevertheless we have gained insight into Tännsjö’s main argument. I turn to that next.

The Actual Value of Our Actual Lives

Tännsjö accepts the Repugnant Conclusion because he imagines what lives in a Z-population might be like as well as how much they would differ from lives in an A-population. He believes that, on the whole, A-lives and Z-lives do not really differ very much. Tännsjö’s view is that people have lives worth living as long as basic needs are met. While this is not a very specific description of what a life in Z is like, it is, nonetheless, helpful in leading us to consider necessary conditions for a life worth living. Let us say then, that a life worth living is one in which the individual’s basic needs, such as drinkable water and adequate food, are met most of the time. Is a life with such needs met barely worth living or well worth living? For Tännsjö, there is not much difference between the two. He eloquently makes his pessimistic point:

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 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 isn’t always such a great acquaintance to make. Note also

\[23\] Ibid.
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.\footnote{Ibid.}

Tännsjö writes that, while many of us have lives that we consider well worth living, they are on the whole, not so different from lives that are barely worth living. Even the best possible life includes unfulfilled expectations and periods of suffering. Additionally, modern gadgets cannot really make the difference between an A-life and a Z-life. According to Fig. II (page 10), the people in population A are six or seven times happier than people in population Z. But what does it mean for someone to be six or seven times happier than someone else who is also happy (albeit barely happy)? This is not clear, and on Tännsjö’s view, there is no such disparity between actual people’s lives. According to Tännsjö, once we realize that there is not much difference between A-lives and Z-lives, we can see that the Repugnant Conclusion is not repugnant at all. The problem of the Repugnant Conclusion is thus resolved or avoided. However, Tännsjö’s view is open to objections. I consider the most important of them next.

**Objections**

Tännsjö considers two objections to his argument that A-lives are not very different from Z-lives. The first objection is essentially that his claim about the value of lives is not true. Some may assert that their lives could be significantly worse and yet still be worth living. If this is true, it would show a significant disparity in quality between A- and Z-lives, thus establishing Tännsjö’s claim that they are not different to be
false. Tännsjö cites Nils Holtug as one who offers this objection. Holtug states, “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.” For example, the difference here may, perhaps, be roughly between the life of a middle-class American and the life of someone in a poor country who works for a couple of dollars a day and has no access to basic medical care when he needs it. Tännsjö writes, “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.” However, Tännsjö insists that this view is mistaken.

To defend his view, Tännsjö appeals to cases of people who have come to be disabled by losing a limb. When asking people in such circumstances how they assess the quality of their lives, Tännsjö has always received the same answer. They say, “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.”

This response works well for someone living in a wealthy nation where there are ways to integrate oneself back into society after becoming disabled. But in some other

---

26 Tännsjö, 225.
27 Ibid.
areas of the world, someone who loses a limb may not be able to get medical attention, physical therapy, or other assistance to cope with their loss.

Furthermore, arguably the life of a person who loses a limb, cannot get proper medical care, and is unable to function on a day-to-day basis, is not worth living. Imagine the life of someone who lost both legs, has no wheelchair or prosthetic legs, and has no one to help him function. Additionally, imagine that his legs were lost early in his life. Such a life would be difficult at best. This might not be a Z-life, but instead a life not worth living. The problem here is that, although a life definitely worth living and a life definitely not worth living are easy to distinguish, a life barely worth living is difficult to conceptualize with precision. While Tännsjö evidently believes that the necessary conditions for a life to be barely worth living are also sufficient conditions for a life to be well worth living, I disagree: there is a difference between the two.

There are multiple counterexamples to Tännsjö’s claim that A- and Z-lives are the same or even similar. Consider the life of a person who, at twenty years old, has been sentenced to life in prison for a murder he did not commit. He lived a couple of decades of a flourishing life but spent the next sixty (or so) years in prison, where he was treated humanely. Nonetheless, his life is worth living, although barely worth living. A hedonistic utilitarian like Tännsjö might respond that this is not a life worth living because the positive utility of the first twenty years does not outweigh the sixty years of negative utility. In that case another example may suffice. Imagine the life of a person who is severely mentally retarded. He has a fairly high level of subjective well-being so that, on a hedonistic utilitarian view, he has a life worth living. Should a utilitarian judge
that a Z-world full of severely mentally retarded people is better than an A-world full of people who do not have any form of mental retardation? Tännsjö may think so. If so, then I would disagree but not be able to convince him to agree with me—and vice versa. Such a view is simply counterintuitive to me.

Tännsjö could deny that these Z-lives would be worse than A-lives by appealing to some kind of objective criterion of utility. However, there are multiple other counterexamples that do not involve subjective utility. Consider someone born into an abusive family without any opportunities to cultivate her interests. She does not have regular access to medical care when she needs it and, furthermore, acquires a serious heart disease, such as cardiomyopathy, at the age of fifteen. However, she is not constantly suffering. She does not have to beg for food but she has barely enough to feed herself. The state of her country is such that it is nearly impossible for her to get an education and a decent job. Nonetheless, we can say that she has a life worth living. After all, she is not constantly suffering and the positive utility that she gains from, say, meaningful relationships, compensates for any suffering she does experience. This serves as a counterexample to Tännsjö’s claim that a life well worth living and a life barely worth living are similar or even indistinguishable.

Furthermore, even if Tännsjö’s story of someone who loses a limb works in favor of his conclusion, it is only one example. As I have shown, there are multiple examples of lives that are worth living yet significantly worse than other lives that are also worth living. All in all, Tännsjö’s reasoning is more speculative than seriously argumentative and it is difficult to see how he could back up his claim that A- and Z-lives are as similar
as he asserts. However, while I disagree with Tännsjö that A- and Z-lives are so similar, I do not think that they are as disparate as the Repugnant Conclusion literature has made them out to be. Here Tännsjö has offered some important insight into the problem.

A second objection that Tännsjö considers is that he has not looked at the whole picture when addressing what is repugnant about the Repugnant Conclusion. It is not that Z is repugnant in itself; it is the judgment that Z is supposedly better than A that is repugnant. Tännsjö cites Nils Holtug as raising this objection. Holtug writes, “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. Tännsjö only addresses the former feature, and this feature, taken by itself is not repugnant. So it seems to me that Tännsjö does not explain away the repugnancy.”

The judgment that population Z is better than population A is what generates the repugnancy.

Tännsjö responds to this objection by reinforcing his claim that lives in A really are not so very different from lives in Z. He writes of the Z-world:

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… [O]nce 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.²⁹


This reasoning is not sufficient to show that the judgment that population Z is better than population A is not repugnant. Tännsjö does not make clear to the reader what kinds of lives he is thinking of when he writes that the Z-world is similar to the actual world. There is certainly some difference between qualities of lives in our actual world. And even if the Z-world is the same as the actual world, it does not follow that the repugnancy totally disappears. If the difference in happiness between A and Z is not extreme, as Tännsjö thinks, we can imagine that an A-life is the life of a modern Westerner who has multiple opportunities and has had them her whole life. She has health care and a job that she loves. A life in the Z-world may be one like the example above of a person who is born into an abusive family without any opportunities to cultivate her interests and who acquires a serious illness when she is fifteen. To judge a large population of these Z-people as better than a smaller population of the A-people would be morally repugnant. It is simply counterintuitive. If utilitarians care at all about lives of individuals and how they live those lives, this conclusion should be unacceptable to them. We need not compare a world containing one god-like being who is impossibly happy with a world full of one trillion people who live nearly tortured lives. Such a comparison is not necessary in order to establish the repugnancy of the Repugnant Conclusion. However, Tännsjö has shown that the Repugnant Conclusion is not as repugnant as we might think. If a life worth living is one in which basic needs are satisfied, then there is not such an extreme difference between the A- and Z-lives. This cushions the blow to utilitarianism dealt by the Repugnant Conclusion, but it does not block it altogether.
Conclusion

The success or failure of Tännsjö’s argument depends on the standard for a life worth living. Given his position as a hedonistic total utilitarian with a pessimistic view about lives worth living, Tännsjö has shown that the Repugnant Conclusion is not as repugnant as Parfit thinks. However, he cannot adequately address the objections raised by critics that part of what is repugnant is that the utilitarian judges the Z-population to be better than the A-population. And even though Tännsjö stresses that the Repugnant Conclusion expresses a *possibility* and does not require anyone to *actually* try to bring about population Z, he does not successfully support this claim. Tännsjö has shed light on a different way of viewing the Repugnant Conclusion, but his argument rests on his conception of a life barely worth living. The conception of such a life is addressed by Clark Wolf, who focuses his version of utilitarianism on negative utility rather than distinctions between total and average utility. I turn next to an assessment of Wolf’s version of utilitarianism, how it relates to the Repugnant Conclusion, and how he interprets the idea of a life barely worth living.
CHAPTER III: WOLF AND POPPERIAN CONSEQUENTIALISM

Wolf’s utilitarian response to the Repugnant Conclusion is an interesting one because he reframes the conclusion in order to resolve the problem. First, I will describe Wolf’s view, which he calls Popperian consequentialism. Second, I will explain Wolf’s attempt to solve the problem of the Repugnant Conclusion. And lastly, I will argue that while Wolf does avoid some problems that derive from the problem of the Repugnant Conclusion, his theory does not solve the problem and is stuck with further trouble: what Parfit calls the Absurd Conclusion.

Wolf sets out six basic ideas that are important in his view of utilitarianism. First, he defines the Misery Principle.

Misery Principle: If people are badly off, suffering, or otherwise remediably miserable, it is not appropriate to address their ill-being by bringing more happy people into the world to counterbalance their disadvantage. We should instead improve the situation of those who are badly off.\(^{30}\)

The Misery Principle, says Wolf, rules out many defended theories of population choice, including average and total utilitarianism.\(^{31}\) What Wolf considers most important is the alleviation of suffering. The suffering of some cannot be reduced by simply making other people happier (or adding happy people to the population.) This is the crux of Wolf’s view and it provides a foundation for most of his argumentation.

The second idea Wolf introduces is that we should reject commensurabilism, defined as follows:

---


\(^{31}\) Ibid. 64
Commensurabilism: [the assumption] that well being is commensurable with ill-being, and that the two can be weighed against one another in a grand aggregate.\textsuperscript{32}

This idea relates to the Misery Principle, which holds that suffering and pleasure are to be regarded differently, and furthermore, that the suffering of others is more morally important than the happiness of others. To clarify commensurabilism, Wolf introduces a third and fourth idea; he distinguishes between two types of utilitarianism, Positive Utilitarianism (PU) and Negative Utilitarianism (NU):

\textit{PU}: Actions are presumptively obligatory to the extent that they promote well-being.

\textit{NU}: Actions are presumptively obligatory to the extent that they mitigate ill-being.\textsuperscript{33}

Wolf favors NU and rejects PU for the following reasons.

First, while PU is insatiable, NU is satiable. Wolf writes, “the [PU] requirement is insatiable – no matter how much well-being there is, PU implies that it is presumptively obligatory to produce more. On the other hand, NU is satiable – if ill-being were eliminated, then NU would imply no further obligation.”\textsuperscript{34} It would eliminate numerous obligations that would otherwise exist, since people can always be made happier. Utilitarianism would be less demanding and utilitarians would be more likely to fulfill their obligations. However, and this is a practical point against Wolf, while NU \textit{is} satiable, given the number of suffering people on Earth, it can be regarded, practically speaking, as insatiable. No matter how much suffering we alleviate, there always seems to be more. Another problem, mentioned by Wolf, is that “one good way to minimize

\textsuperscript{32} Ibid., 66
\textsuperscript{33} Ibid., 67
\textsuperscript{34} Ibid.
misery would be painlessly to extinguish all conscious life from the universe.”³⁵ There can be no suffering in a world without beings capable of suffering.³⁶ Wolf writes, “A world without beings capable of misery would be a world in which misery was maximally mitigated, so NU might be taken to recommend that we exterminate the miserable instead of helping them.”³⁷ Wolf gives this as a reason to question negative utilitarianism but does not attempt to refute it. One may ask why a utilitarian ought to ameliorate the suffering of those in pain rather than painlessly euthanizing them. It would certainly be easier and provide a more certain way of mitigating suffering. I think this is a serious problem for Wolf. However, I suppose he (or any other utilitarian who subscribes to NU) could argue that the euthanization argument misses the point. Perhaps, euthanizing everyone in order to eliminate suffering eliminates the purpose of utilitarianism. While it is true that negative utilitarians want to mitigate suffering, they do not want to totally ignore positive utility. Mitigating suffering is the most important, but increasing positive utility is important as well. This can only be done if there are people in existence. But, if Wolf takes this route, he would no longer be a negative utilitarian. He would be a positive utilitarian.

The second reason Wolf gives for favoring NU over PU is that “unlike PU, NU recommends that it is presumptively obligatory to make people happy (if they’re not), but

---
³⁵ Ibid.
³⁶ This raises some further questions: Is an empty planet better or worse (or equally good) as compared to a planet full of life? If there is no living thing in existence to value life, is there really any value? Or rephrased, can life be valuable without something living valuing it? These questions are interesting but beyond the scope of this paper.
not obligatory to make happy people.”

This is a good point. Whereas PU holds that people are presumptively obligated to create (additional) happy people, NU does not assert any such presumptive obligation. This is an agreeable implication of NU, and thus, Popperian consequentialism.

The third reason for favoring NU over PU is as follows. Wolf writes, “If some people are miserable while others are well-off, NU assigns an obligation to aid those who are miserable, even if we could instead dramatically improve the lives of those [who] are well-off. But PU would recommend aiding the miserable only if the amount of well-being provided for those who are badly off is greater than the well-being we could instead provide for the well-off.” However, Wolf does not take into consideration that most of the time the most practical way to increase happiness is by helping those who are suffering, not by making the marginally happy people happier. So practically, PU and NU would often commit one to the same action much of the time. For example, one may have the money to buy a new Corvette for someone who has a car and is moderately happy, and this would bring about quite a bit of happiness for the person. However, to bring about the most happiness for the money, one could donate it to a poor African country for the construction of a hospital. A Corvette may last many years, but it ultimately gives only a few people happiness. A village hospital can bring more happiness to more people over more years. Still, there are cases in which it is possible to generate more happiness by making happy people happier than by making unhappy people happy, and so PU and NU do not always prescribe the same actions in all cases.

---

38 Ibid., 68.
39 Ibid.
In these cases, NU may not prescribe the action that most think would be the morally best.

Popper, Obligation, and Beneficence

To support his view, Wolf cites Karl Popper as a philosopher against commensurabilism and as an advocate of NU. Popper writes, “…all moral urgency has its basis in the urgency of suffering or pain. I suggest, for this reason, to replace the utilitarian formula ‘Aim at the greatest amount of happiness for the greatest number’ … by the formula ‘The least amount of avoidable suffering for all’.”40 Popper also writes that there is “no symmetry between suffering and happiness, or between pain and pleasure.”41 This Popperian view is a kind of negative utilitarianism. Wolf argues from this perspective for two types of imperatives (these are the fifth and sixth ideas), the Negative Principle of Obligation [NPO] and the Positive Principle of Beneficence [PPB].

**NPO:** Actions that reduce (or minimize) misery are *prima facie* obligatory.

**PPB:** Actions are good if they increase well-being. Actions are better or less-good depending on the extent to which they promote well-being.42

Wolf describes Popperian consequentialism as a theory that “regard[s] the negative principle as assigning *obligations* while the positive principle applies to beneficence.”43 There are two attractive features of Popperian consequentialism, according to Wolf. First, “Popperian consequentialism comports well with many of our pre-theoretic

---

41 Ibid., 284.
43 Ibid.
judgments. Most people regard it as at least presumptively obligatory to address the suffering of others, but would not regard themselves as blameworthy if they were to pass up a chance to improve the situation of other well-off people around them. This is a valid point and I think it is a good aspect of Popperian consequentialism.

Second, Wolf argues,

Popperian consequentialism accounts well for commonsense judgments about genesis choices—choices that involve bringing people into existence. It is commonly judged that we have at least a prima facie obligation not to bring into the world people who will be miserable… But while we may regard it as (at least) presumptively obligatory not to conceive a child who will be miserable, we do not typically regard it as obligatory to conceive children whenever the children we conceive would have an excellent chance to enjoy a good life. At most, we regard it as permissible or (at most) ‘good’ to have children who face good prospects for a happy life. No one has an obligation—even a prima facie obligation—to have children.

That Popperian consequentialism does not obligate people to have children is certainly a reason to favor the theory. Many people would concur with the conclusions that are initially implied: we have an obligation not to bring miserable people into the world and it is not obligatory to produce happy children (or to produce children at all). This is, of course, an empirical question that I will not go into here. However, while Popperian Consequentialism may conform with many people’s moral intuitions about alleviating suffering and may agree with their views about whether there is an obligation to reproduce, there is an implication that conflicts with many people’s moral intuitions.

This is a familiar objection to utilitarianism: Do we have an obligation to do all we can to help those who are suffering? Would a couple be obligated, instead of buying a vacation for their family, to give that money to charities, fund-raisers, and organizations

---

44 Ibid.
45 Ibid., 72
that help developing nations? On Popperian Consequentialism, the answer to these questions seems to be ‘yes’. Popperian Consequentialism thus falls prey to the criticism that it is too demanding. However, there are utilitarian replies to this criticism.\footnote{See Peter Singer, “Famine, Affluence, and Morality”, \textit{Philosophy and Public Affairs} 1, no. 3 (1972). 229-243.}

Wolf writes that there is an obligation not to produce children who would have miserable lives, and that there is no obligation to produce children who have a good chance at having excellent lives. However, he does not write about children who have lives that are not excellent but still worth living. The same question arises, as about the children who have a good chance at having an excellent life: should we regard producing children who have a chance at a life barely worth living as being permissible or (at most) not worse than not having them? It is hard to say what Wolf’s answer to this question would be, but I imagine he would say producing these children is permissible, though not obligatory since not bringing such a person into the world does not mitigate suffering.

Next, I turn to Wolf’s solution to the Repugnant Conclusion.

\textbf{Popperian Consequentialism and the Repugnant Conclusion}

The way Wolf attempts to avoid the Repugnant Conclusion is by re-framing the conclusion so that it is less repugnant. Wolf writes from the outset that “Popperian consequentialism does not entirely succeed in allowing us to avoid the repugnant conclusion, but perhaps it is possible to undermine the repugnance by re-framing the conclusion itself.”\footnote{Clark Wolf. “O Repugnance, Where is Thy Sting?” in \textit{The Repugnant Conclusion. Essays on Population Ethics}, eds. Jesper Ryberg and Torbjörn Tännsjö (N.p. Kluwer Academic Publishers, 2004), 73.} He later writes that “NU does not by itself imply Parfit’s repugnant
conclusion. But neither NU nor Popperian Consequentialism will rule it out either. Thus, by itself the Popperian view gives us no reason to prefer small well-off populations to large marginally well-off populations. This objection to the Popperian view is less serious if Wolf can show, as Tännsjö attempted to do, that the Repugnant Conclusion really is not so repugnant. He endeavors to do this by defining what it means for a life to be barely worth living. If he can show that a life barely worth living is not so bad, he can possibly avoid the repugnance.

Wolf reframes the Repugnant Conclusion:

*The Slightly-Less Repugnant Conclusion:* For any possible population of at least ten billion, 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 just barely good.

Wolf asks, “Which lives are just barely good?” The barely good life is one which is above what Wolf calls the ‘neutral level’.

This concept is a vague one. In order to determine what a just barely good life is, Wolf considers multiple proposals of lives that would be marginally above a neutral level. One of the initial proposals that he favors is the following:

*Proposal 4A:* Life is marginally above the neutral level just in case if you knew that any child *someone else* conceived would enjoy at least that level of well-being, you would not regard that person as having a reason (deriving from consideration of the child’s welfare) not to conceive a child.

This proposed definition of a barely good life avoids issues of preference for one’s own children over another’s, which is good because it is more impartial. Wolf writes, “As we

---

48 Ibid.
49 Ibid., 74.
50 Ibid.
51 Ibid., 77.
consider Proposal 4A, it may be possible to abstract from the partiality we feel on behalf of our own children. And perhaps this is a fair measure of the limit at which life is good.”

If we think of a life barely worth living (or a barely good life) in these terms, and imagine population Z with a very large number of people, all with this kind of barely good life, the Repugnant Conclusion does not seem so repugnant. One way of testing whether we think a life is really good, is to test it against the Repugnant Conclusion itself. Wolf writes, “If we find the repugnant conclusion repugnant, perhaps it follows that the lives we are considering aren’t really marginally good after all.” This idea hints at another definition of a barely good life, the one which Wolf ultimately favors:

Proposal 5: At some level of well-being, it is no longer repugnant to think of numerous people or even innumerable people living at that level, no longer odd to think that we might reasonably be indifferent between a world in which fewer people were even more blissful, and a more numerous world in which people were only just so blissful. Life is marginally above the Neutral Level when it is at least as good as this.

This definition raises the standards of what a barely good life (or a life barely worth living) is. However, this definition raises a problem. From whose perspective is the judgment of what is repugnant made? The perspective of an affluent Westerner is very different than the perspective of someone from a developing nation who is uneducated, has minimal healthcare, has a large number of children, but is nonetheless happy. The definition of a life barely worth living is somewhat unclear in Parfit’s work, but it does not have to be precise in order to lead to the Repugnant Conclusion. When trying to

---

52 Ibid.
53 Ibid., 77-78.
54 Ibid., 78.
define what a life worth living (or a barely good life) is from a utilitarian standpoint, one runs into more problems. For example, how should utility be interpreted? Wolf is able to avoid giving a list of necessary conditions for a life barely worth living by, instead, giving comparative definitions. Given his Proposal 5, Wolf is able to make the Repugnant Conclusion less repugnant.

However, even if the repugnance is avoided, Wolf recognizes that there is a further problem. Both he and Parfit see that by raising the standards of what a life worth living is, the Repugnant Conclusion is avoided but an altogether different problem follows:

_The Absurd Conclusion:_ Suppose that, in some history of the future, there would always be an enormous number of people, and for each one person who suffers and has a life that is not worth living, there would be ten billion people whose lives are worth living, though their quality of life is not quite as high as the Valueless Level. This would be worse than if there were no future people.\(^5^6\)

Why might this conclusion be absurd? There are a number of reasons. Wolf writes, “Popperians must accept the implication that the empty world is at least as good, from the perspective of minimizing misery, as any possible world. More counterintuitively perhaps, the Popperian view ranks the empty world as better than any world in which there is at least one person whose life is marginally bad—the actual world, for example.”\(^5^7\) Furthermore, many will argue that a world without any people is a waste or has no value. To this question I ask, to whom is it a waste? To whom has it no value? If there are no people in existence then the planet cannot be valuable because there is

---

\(^5^5\) Wolf’s Neutral Level and Parfit’s Valueless Level are the same.


nobody alive to experience the lack of value. Wolf writes, “Because I find it quite counterintuitive to regard well-being as impersonally valuable, I do not find the Absurd Conclusion excessively absurd.” I tend to be sympathetic to this view. However, as a human, I would be distraught if all living things that are capable of suffering (including humans) were going extinct so quickly as to bring about an empty world in the next decade. But this is not a problem we face today. I would be even more distraught if governments worldwide put policies into practice that actively euthanize all living things that are capable of feeling pain or pleasure in order to bring about the empty world. But this goes further than simply comparing two possible worlds. When comparing the actual world with the empty world, to prefer the empty world is not such a negative view.

Conclusion

To conclude, Wolf does give some acceptable and important guidelines for acting with regard to producing children. Unlike Tännsjö, Wolf succeeds in making the Repugnant Conclusion less repugnant by redefining a life worth living. He does this by raising the standard of what a barely good life is. However, in doing so, negative utilitarianism is left with the problem of the Absurd Conclusion. While there is no inherent utilitarian problem with accepting Popperian consequentialism, it does not, by itself, avoid the Repugnant Conclusion. Wolf gives the utilitarian no reason to prefer population A to population Z.
CHAPTER IV: AVERAGE UTILITARIANISM

The prima facie advantage of average utilitarianism is that the required transition from population A to B (and eventually to Z) is blocked. Since the average, not total, happiness of a population is what is to be maximized, the average utilitarian would prefer A over B. Gustaf Arrhenius et al. state, “That the average well-being is all that matters implies that no loss in average well-being can be compensated for by a gain in total well-being. Thus, in Figure [II], A is preferable to Z, i.e. the Repugnant Conclusion is avoided.”\(^{58}\) This seems to make practical sense. We want people to be happy and no amount of barely happy people can be as good as population with members who live very happy lives.

However, Parfit, among others, rejects average utilitarianism. In this section I will examine works by Parfit, Neumann, Hurka, and Anglin to see whether we should agree with Parfit in rejecting it. First, I will explain why, according to Parfit, average utilitarianism should be rejected as a solution. But as I will then show (second), Neumann’s version of average utilitarianism addresses Parfit’s criticisms of average utilitarianism. Third, I will consider Thomas Hurka’s response to Neumann’s kind of average utilitarianism. Fourth, I will show how, on Bill Anglin’s view, average utilitarianism sometimes has implications that are similar to the Repugnant Conclusion. Finally, I will conclude that while average utilitarianism may be able to avoid not only some of the implications of total utilitarianism but also Parfit’s criticisms, it nonetheless has some counterintuitive implications.

Parfit on Average Utilitarianism

Parfit defines the average principle as follows:

The Impersonal Average Principle: If other things are equal, the best outcome is the one in which people’s lives go, on average, best.\(^{59}\)

The hedonistic version of this principle, Parfit writes,

If other things are equal, the best outcome is the one in which there is the greatest average net sum of happiness, per life lived.\(^{60}\)

I begin by giving an overview of Parfit’s two main criticisms of average utilitarianism: his ‘Eve and Adam’ argument and his ‘How Only France Survives’ argument.

Parfit writes, “On the Average Principle, the best history might be the one in which only Eve and Adam ever live. It would be worse if, instead of Eve and Adam, a billion billion other people lived, all with a quality of life that would be almost as high.”\(^{61}\)

Parfit thinks that this claim is difficult to believe, but that it follows from average utilitarianism.\(^{62}\) Regardless of whether the claim does follow from average utilitarianism, there is a further, absurd implication of average utilitarianism. Parfit states, “Suppose Eve and Adam lived these wonderful lives. On the Average Principle it would be worse if, not instead but in addition, the billion billion other people lived. This would be worse because it would lower the average quality of life.”\(^{63}\) These first two objections are based on a temporally neutral version of average utilitarianism\(^{64}\). What this means is that when calculating average utility, one must not only look at who will be affected at the time of

\(^{60}\) Ibid.
\(^{61}\) Ibid., 420.
\(^{62}\) Ibid.
\(^{63}\) Ibid.
\(^{64}\) I will call average utilitarianisms that are not temporally neutral, ‘non-temporally-neutral’.
or after an action, but also take into consideration those alive before the action is
undertaken. As will be shown shortly, implications of temporally neutral versions of
utilitarianism are very different than non-temporally neutral versions.

Turning to his second (‘How Only France Survives’) criticism, Parfit considers
that what matters for the average utilitarian is “the average quality of life of all presently
existing and future people.”65 Parfit compares two possible futures. He writes,

In one possible future, the worst-off people in the world soon start to have lives
that are well worth living. The quality of life in different nations then continues
to rise. Though each nation has its fair share of the world’s resources, such things
as climate and cultural traditions give to some nations a higher quality of life.
The best-off people, for many centuries, are the French.

In another possible future, a new infectious disease makes nearly everyone
sterile. French scientists produce just enough of an antidote for all of France’s
population. All other nations cease to exist. This has some bad effects on the
quality of life for the surviving French. Thus there is no new foreign art,
literature, or technology that the French can import. These and other bad effects
outweigh any good effects. Throughout this second possible future the French
therefore have a quality of life that is slightly lower than it would be in the first
possible future66.

In this second possible future, the French have a slightly lower quality of life than they
have in the first possible future. However, Parfit stipulates, the average happiness of
everyone in existence would be much higher as a result of the non-existence of people in
the rest of the world.67 On the Average Principle, the second future is preferable. Parfit
writes of all the non-French individuals: “If these billions of people lived, their lives
would be well worth living, and their existence would be better for the French. On the
Average Principle, it would be worse if these people lived, simply because the French

66 Ibid.
67 Ibid.
have even better lives.”

This is a strong argument against average utilitarianism, but it can be rebutted.

**Neumann on Average Utilitarianism**

In *Killing and Average Utility*, Neumann defends a specific formulation of average utilitarianism. He writes that average utilitarians “define the average utility of an act as its total utility divided by the number of those who possess at least one of two qualifications: (i) they exist when the act is performed, (ii) they exist after the act is performed.” As Parfit notes, thinking this way allows us to avoid the Eve and Adam problems. Instead of calculating possible histories, the average utilitarian calculates only utility as it applies to people existing at the time of the action and after the action. Thus, a history in which the happiest people lived in the past would not require the average utilitarian to eliminate the current population because that population brings down the average happiness of people who have ever existed or will ever exist. This is also a sensible way to think about utilitarianism. There is no possible way to take into consideration the happiness of individuals thousands of years ago. If average utilitarianism required this, the theory would be useless.

However, Parfit argues against the average principle when it “refers only to the people who are alive after we have acted.” He thus argues against non-temporally neutral versions of average utilitarianism, of which Neumann’s is one, as having the

---

68 Ibid.
71 Ibid., 386.
ludicrous conclusion that we ought to kill “all but the most ecstatic.” But Parfit understands non-temporally neutral average utilitarianism differently than Neumann does, and the way he understands it leads to the obligation to kill all but the happiest people. Neumann takes into consideration people who exist at the time of an action. Parfit, however, describes the average principle as only applying to “people who are alive after we have acted.” This is a small but important difference. To see this, consider a population containing John (ten units of happiness), Mary (nine units of happiness), and Luke (eight units of happiness). On Parfit’s definition of average utilitarianism, if Luke is killed, the average happiness goes from 9 to 9 ½. By killing Mary as well, the average happiness increases to 10. To see this, consider Fig. III: Killing and Utility Levels:

<table>
<thead>
<tr>
<th></th>
<th>Initial utility levels</th>
<th>Utility levels after the 1st killing</th>
<th>Utility levels after the 2nd killing</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Mary</td>
<td>9</td>
<td>9</td>
<td>—</td>
</tr>
<tr>
<td>Luke</td>
<td>8</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Figure III: Killing and Utility Levels*

On Parfit’s view, since Luke and Mary cease to exist, they are not counted in the utility levels. However, on Neumann’s kind of average utilitarianism, there would be zeros in place of the dashes, indicating that Luke and Mary’s lives are counted at the time

---

72 Ibid.
73 Ibid. Italics added
74 Michael Neumann, “Killing and Average Utility,” *Analysis* 40, no. 1 (January 1980): 36. (I have added the title ‘Killing and Utility Levels’ to this figured.)
of the killing. The average utility levels after the first killing would then be $6 \frac{1}{3}$ instead of $9 \frac{1}{2}$ as on Parfit’s view. If Mary and Luke were both killed, the average utility level would be $3 \frac{1}{3}$ instead of 10. This is sufficient for showing that the average utilitarian is not obligated to kill all but the happiest people. This meets the first of Parfit’s two criticisms of average utilitarianism. What about the second? I will now provide a response to Parfit’s France example on behalf of the average utilitarian.

Parfit’s France example shows that average utilitarianism has undesirable implications even if you subscribe to Neumann’s non-temporally-neutral version of average utilitarianism. However, I think this thought experiment fails to show that average utilitarianism has particularly negative results in practice. Imagine the population of France years and years after everyone else on the planet has become sterile. This future may be preferable to the first possible future that Parfit gives in his example. The happy French would continue to reproduce and there would be more than just moderately happy people on the planet. This future may happen to be better in multiple ways. For example, the happy French may spread their society and thus there would be more people with better schools, better healthcare, more opportunities, etc. This really is not such a bad outcome. The only really unpleasant aspect of this population would be the means to arrive at it. The fact that the majority of the planet became sterile and stopped reproducing is a sad story. However, saying that one possible story about the future (where the consequences of any means are not taken into consideration) is better than another is not the same as prescribing that we actively bring about that population. The consequences of the means to attain a certain population must be considered as well.

---

75 Ibid., 36.
Parfit’s thought experiment asks us to imagine that the sterilization is permanent and was not caused by humans trying to bring about a particular population. The consequences of the case in which only the French survive would be much different if the sterilization of everyone was purposively done with the purpose of population control or eugenics. This brings me to my next point.

Recall that, in Parfit’s France example, the first possible future is one where the worst-off people on the planet start to have lives well worth living although the best-off people for many centuries are the French. In the second possible future, only the French survive a world-wide infectious disease. The quality of life in France is slightly lower than in the first possible history. While Parfit’s France example has the strange conclusion that utilitarians should favor the second future over the first, his view does not prescribe that utilitarians actively sterilize people. There is a distinct difference between the consequences of lives ending by accident and the consequences of people purposefully ending lives; this is a distinction that many utilitarians fail to draw. For example, if a tsunami hit a developing country whose citizens were living sub-par lives and were a burden on other nations, the loss of those citizens might, in the long run, be a good thing on average utilitarian grounds. Let us assume that it is. However, proposing a governmental policy to painlessly eliminate those same citizens would be morally very different. It is necessary to keep in mind here that not everyone is a utilitarian, let alone an average utilitarian. Because this is the case, one should assume that there would be many people deeply troubled by the systematic, yet peaceful, killing of the citizens of the

---

76 The reasons may include: developed nations no longer have to send food, medicine, soldiers for defense from other countries, doctors, etc. to help this developing nation.
developing nation. Because of this, the average utility would be lowered. But imagine plausibly or not that all of these possible consequences are taken into consideration when formulating the proposal for the governmental policy to painlessly end the lives in the developing nation full of barely happy people. Even if everyone in that country were an average utilitarian, if their government systematically ended the lives of the unhappy, everyone would be in constant distress for fear of slipping into an unhappy state for some reason and wondering whether they would be killed next. This would cause anxiety about death and would result in everyone feeling the need to constantly maintain a happy mood. This may not even be possible. In a country (or world, if we consider a world-wide population policy) where the unhappiest (or those whose happiness is less than the level of the most ecstatic) risk death, anxiety undermines happiness. Thus, by requiring painless killing of the unhappiest people, this average utilitarian policy would cause many people to suffer and thus lower the average happiness of the population. Regarding Parfit’s France example, even if the average utilitarian ought to prefer the second future to the first if nothing else is considered, the example fails to show how average utilitarianism has counterintuitive results in practice. Once all the consequences that can be taken into consideration are, it is reasonable to expect that average utilitarianism would not require the painless killing of all but the happiest. The net utility would not be higher than average. Parfit has shown that average utilitarianism advocates a counterintuitive preference, not a counterintuitive obligation to adopt a certain policy.

77 The people living in the countries doing the killing may come to distrust their governments. They may see the systematic killing of all the people in the developing country as wrong. They would therefore lose trust in their own government, fear the possibility of war, or be unhappy because any number of other results.
Nonetheless, there are further counterintuitive implications of average utilitarianism, which I will address next.

Hurka and Anglin on Average Utilitarianism

In “More Average Utilitarianisms”, Hurka addresses Neumann’s kind of average utilitarianism. 78 Hurka states that Neumann’s formulation tells us to maximize the average of the total happiness in every life (or part of a life) lived from now on into the future by persons who satisfy one of two conditions: (i) they will exist now, when we are performing our actions; (ii) they will exist in the future, after we have performed our actions. 79

Hurka’s main problem with this formulation is that, if accepted, it requires people today to “let the human race die out if people will be less happy in the distant future than they will in the near future.” 80 This would imply certain actions on a collective, governmental level. To get this point across Hurka presents a hypothetical situation. He writes,

Imagine that everyone alive in the world today has an extremely happy life ahead of him, and will enjoy a total of 100,000 units of happiness between now and his death (this is on a scale where the best we do today is 1,000 units of happiness in a life.) But any people born after today will enjoy only 90,000 units of happiness in their lives. Then even though these future people will lead very happy lives [Neumann’s average utilitarianism] says we must not produce them, but let the human race die out. 81

On Neumann’s view, we ought to let the human race die out if it were known that future people would live slightly less than presently average (but still very happy) lives. This is a difficult problem for the average utilitarian. In response to this criticism, I argue on behalf of the average utilitarian that one cannot know with complete certainty how

---

78 Hurka calls it A7’.
80 Ibid., 117.
81 Ibid.
happy future people will be. And even if it were possible know that future people would be slightly less happy years later, the average happiness might rise again. Imagine a case where economists predict a world-wide depression for the next decade. The average quality of life will decrease as a result, but this will be only for a decade or so. Imagine that it will take roughly fifty years for the average happiness to rise back up to where it was before the depression. Note that standard of living and happiness are not the same thing, but they are often correlated. Should we, individually and collectively, take this into consideration when deciding how to act? It is unclear how far we each ought to look into the future when determining the utility of an action. However, this is a problem for any utilitarian view and gives us no reason for rejecting average in favor of total or negative utilitarianism.

Furthermore, while the human race dies out, the average happiness of everyone still alive may increase. It would take years and years for the human race to die out. If it were possible for various governments to put such a policy into practice, many of the world’s problems might be mitigated. There would be fewer consumers, fewer people to feed and employ, fewer people for whom to provide drinkable water and health care, etc. After a number of years during which the human population diminished, the average happiness of people in existence might eventually rise above the level it was at originally. Governments or institutions that subscribe to average utilitarianism would then, presumably, stop trying to let the human race die out and would thus not actually bring about the end of humanity. So while average utilitarianism has the undesired
consequence that it requires letting the human race die out, it would, in practice, require such action and so not as unacceptable as Hurka thinks in this regard.

However, in *The Repugnant Conclusion*, Anglin presents a number of arguments that show that various aspects of average utilitarianism imply repugnant conclusions. I will only address the argument relevant to Neumann’s average utilitarianism. Anglin gives an example:

> [S]uppose that there are one thousand people who will exist and enjoy a net one unit of happiness each regardless of what the average utilitarian does and suppose that he can add to this population any number of people who will each enjoy a net of 1.1 units of happiness. Since the quotient

\[
\frac{1000 \times 1 + n \times 1.1}{1000 + n}
\]

increases as \( n \) increases, it follows that the average principle utilitarian will be obligated to add as many people as possible, an infinite number if he can.\(^{82}\)

Let us consider two possible populations each containing one thousand people with one unit of happiness. In the first population we add five hundred people (who we know will have 1.1 units of happiness). Multiply the original 1000 people by one (for the one unit of happiness). Then add in 500 people times 1.1 (for the 1.1 units of happiness). Then divide that number by the sum of the total population (1500). The average happiness is thus 1.03; up .03 points from before the five hundred people were added. In the second population 5000 people are added, thus (by the same math) raising the average happiness to 1.08.

Population 1 \( \frac{1000 \times 1 + 500 \times 1.1}{1000 + 500} = 1.03 \)

---

The average utilitarian would be obligated to bring about the second, rather than the first population. This reasoning gets reiterated until an immensely large population is reached. The average utilitarian is thus obligated to bring about a larger population so long as she knows that each person brought into existence will have more than one unit of happiness. There is a reproduction obligation in these cases. Furthermore, there is an obligation not to reproduce so long as the child brought into existence will live a life worse than average (even if their life will be well worth living). This is not the same as the Repugnant Conclusion. Nonetheless, some utilitarians may not like these conclusions. Anglin writes, “Thus if the adherent to the average principle finds the [total utilitarian] conclusion repugnant for the fact that sheer numbers matter, then he will have to find his own principle repugnant too.” Anglin has shown average utilitarianism to be repugnant in the sense that it obligates us to produce people for the sake of population, not population for the sake of people. In other words, the average utilitarianism is obligated to bring people into the world, not for the sake of that person or for other people, but in order to increase the average utility.

I have two comments regarding this argument. First, it is true that the average utilitarian would be obligated to produce a larger population, but the individuals in that population would not be any worse off than the individuals in the initial population. In fact, they would be better off. The average utilitarian does not arrive at population Z, but rather, a population with a high average utility (like population A), but with a greater

\[
\text{Population 2} \quad \frac{1000 \times 1 + 5000 \times 1.1}{1000 + 5000} = 1.08
\]
number of people. What can be construed as repugnant is, as Anglin notes, the fact that the average utilitarian is obligated to produce people for the sake of a population. This is a problem with utilitarianism and not solely average utilitarianism. Therefore, it is not a problem for average utilitarianism in comparison to other utilitarian theories. Second, there is rarely a case where we know that all the people produced will have more units of happiness. We can speculate about future happiness, but since there are so many conditions that affect happiness, to say whether or not a person produced now would have more or fewer than average units of happiness would be nearly impossible in practice.

Theoretically, average utilitarianism implies that we sometimes ought to increase the population. This is when doing so would raise average utility. And we ought to refrain from reproduction when doing so would lower average utility. This would require individual and possibly collective action on a governmental level. Average utilitarianism solves the problem of the Repugnant Conclusion but is left with these other conclusions which do not respect people as autonomous individuals.

Conclusion

Parfit is quick to dismiss average utilitarianism and his arguments against it are not very convincing. Neumann is able to avoid Parfit’s criticisms by appealing to his own kind of average utilitarianism, which takes into consideration only individuals affected in the future and at the time of the action. Anglin and Hurka present the best arguments against average utilitarianism. Anglin’s point about increasing population is
about as strong as the point Hurka makes about letting the human race die out. Both are contingent on knowledge of the future, and since that knowledge is difficult to have, the theories would not frequently establish obligations. Thus, the arguments are not too threatening for Neumann’s view.

In the end, average utilitarianism solves the problem of the Repugnant Conclusion. We have no reason to prefer total utilitarianism over average utilitarianism on the grounds that Parfit, Anglin, and Hurka have urged us to consider. While both average and total utilitarianism have undesirable consequences, average utilitarianism has less. Total utilitarianism implies the Repugnant Conclusion and an obligation to reproduce while average utilitarianism implies only an obligation to sometimes reproduce. All in all, if utilitarians want to avoid the Repugnant Conclusion they ought to be average utilitarians but they must recognize that there are other implications they might find undesirable.
CHAPTER V: NON-UTILITARIAN REASONS FOR REJECTING AVERAGE UTILITARIANISM

Up until now I have shown that the best utilitarian solution to the problem of the Repugnant Conclusion is to subscribe to average utilitarianism. However, even if average utilitarianism can solve the problem of the Repugnant Conclusion, the theory has undesirable consequences from a non-utilitarian perspective. I turn next to a criticism having to do with utilitarianism and reproductive rights. I then consider whether appealing to a Sidgwickian position which views utility not as the sole motive for action but only the standard of evaluation of action can adequately respond to the criticism. I conclude that it does not justify rights. I then consider whether David Brink’s utilitarianism, which considers certain things as inherently valuable, can justify the right to refrain from reproduction. I conclude that it cannot.

Unacceptable Conclusions

On average utilitarianism, when it can be reasonably expected that a child brought into existence would have a better than average life, it is obligatory to bring that child into existence. This would mean that the many people who live in affluent societies, are sufficiently healthy and well off and are capable, and are capable of reproducing ought to do so. This is so because it is almost certain that many children born into affluent societies will live better than average lives compared to the children in the rest of the world. There are, of course, children in affluent societies who are born into abusive families or who have unloving parents. Total utilitarianism implies a similar conclusion.
Recall that whenever it is expected that a child would have a life worth living it should be brought into existence because doing so would increase total utility. Both average and total utilitarianism limit the freedom of the individual in making reproduction decisions. This especially limits the freedom of women by requiring them to reproduce so long as average utility will be increased by doing so. Average utilitarianism solves the problem of the Repugnant Conclusion but is stuck with the problem that it obligates people to reproduce.

In discussing reproductive rights I focus on the freedom to refrain from reproduction. There should not be a moral obligation to reproduce on the basis of utility maximization. If there were such an obligation, respect for the autonomy of the individual would be violated (i.e. the person would not be treated as an autonomous individual) and this is not part of an acceptable moral theory. There are, of course, certain situations where people may have obligations to reproduce. What I have in mind here is a situation where the human race is going extinct. I set this case aside for the purposes of this paper. When I argue for reproductive rights here I focus on individuals today, when there is no immediate risk of human extinction.

I turn next to two conceptions of utilitarianism that may be able to justify the right to refrain from reproduction. I begin with Sidgwick’s view.
Sidgwick

By allowing actions to be based on a diversity of motivations, utilitarians can avoid denying individuals the right to refrain from reproduction. Here I have in mind a point that Sidgwick makes. He writes,

[T]he doctrine that Universal Happiness is the ultimate standard must not be understood to imply that Universal Benevolence is the only right or always best motive of action. For…it is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim: and if experience shows that the general happiness will be more satisfactorily attained if men frequently act from other motives than pure universal philanthropy, it is obvious that these other motives are reasonably to be preferred on Utilitarian principles.  

What Sidgwick is stating here is that while utilitarianism may be the standard by which we should judge actions, it is not necessarily the best motive on which to act.

Furthermore, since a utilitarian calculation is based on expected or probabilistic utility, people will not always know which actions will raise utility.

For example, imagine that a young man, Henry, goes to visit his elderly grandfather who is near death. Henry expects utility to be maximized because he knows that his lonely grandfather will be happy to see him, his parents will be glad that Henry spent time with his grandfather, and Henry believes he will be rewarded in his grandfather’s will. Unfortunately, while Henry does care about his parents’ wishes, he does not really care much for his grandfather and only really desires to promote utility, his grandfather’s included. He is not motivated to visit his grandfather because of a sense of love, gratitude or loyalty; or even a desire to get to know his grandfather better before he passes. Henry simply visits on the expectation that he is promoting utility. The

---

85 From now on I use “criterion” instead of “standard” so as to stay consistent with current literature.
grandfather is able to tell that Henry’s visits are not sincere and he is greatly disappointed about it. In the end, Henry’s grandfather leaves him nothing in the will. If Henry had acted not on the principle of utility but, instead, from love, gratitude, or loyalty to his dying grandfather, he would have been happier and would have perhaps been more inclined to leave Henry money in the will. Assuming the grandfather did leave Henry money, utility would have been maximized. Of course, Henry might have used the money in such a way that utility was not promoted, but at least there is the possibility of utility promotion if Henry had acted on other motives. This example illustrates the point that taking utility maximization to provide the criterion of evaluation but not in all cases the proper motive for action can better enable people to maximize utility.

Instead of advocating that people use the principle of utility as their motive for action, Sidgwick claims that the common sense morality of one’s day and age is the morality most likely to promote greatest utility. The morality to conform to is not the esoteric morality of philosophers, but the morality that the general population finds intuitive. Here, I am considering the utility that would result if people generally accepted and internalized the rules over a span of time. A utilitarian should promote the common morality rather than actively encouraging morals based on utility-maximization, as doing so may not maximize utility.\(^86\) However, Sidgwick’s distinction between motive and criterion does not always yield a utilitarian justification for a right to choose whether to reproduce: it does not yield such a requirement if no such right is commonly accepted. Some societies have as their norm an obligation to reproduce. For example, in certain religious groups it is understood that one should reproduce if one can, and small families

\(^86\) Ibid., 475-476.
are looked down upon. On the one hand, in those societies, appealing to non-utilitarian motives would not always establish the right to refrain from reproduction. On the other hand, societies without the right to refrain from reproduction seem few and far between. In any case, the Sidgwickian split at least provides justification to refrain from reproduction at least some of the time. But this is not a satisfactory utilitarian solution as it does not justify reproductive rights.

There is certainly much debate as to whether Sidgwick’s view is correct or is even a moral theory. Samuel Scheffler describes one criticism of the view that utility ought to be seen as the criterion but not the moral motive for action. Writing on the motive/standard-of-action distinction, he states, “non-consequentialists characteristically argue that there is something irrational or otherwise objectionable about a principle of right action which, from its own point of view, may not provide an acceptable source of motives or a permissible basis for decision-making.”\(^87\) This is a substantial criticism, but Scheffler does not believe it shows utilitarianism to be self-defeating or irrational, “any more than the principle that students ought to stay calm during an important examination would be proven irrational if it should turn out that students who made calmness their conscious aim ended up tenser than they would otherwise have been.”\(^88\) Raising another criticism, he writes, “It may be said that even if utilitarianism is not irrational, any moral theory capable of requiring its own suppression surely violates one of the conditions of adequacy for such theories.”\(^89\) This is a worthwhile criticism which supposedly shows utilitarianism to be counterintuitive because a moral theory should not have to be

\(^88\) Ibid. 46
\(^89\) Ibid. 46.
suppressed, but promoted instead. Related to this criticism is another regarding the motive/criterion of action distinction. If it is the case that the maximization of utility is the criterion of action and not the motive for action, then utilitarianism provides no motive for action. It seems that for a moral theory to be adequate, it must provide one with a motive for action and not simply a criterion by which to retrospectively judge actions. I think that this criticism is a considerable objection to the motive/criterion-of-action separation. However, even if the Sidgwickian utilitarian could answer it, she would still be stuck with the problem that she does not have sufficient justification for the right to refrain from reproduction.

As I have shown, a Sidgwickian approach is a worthwhile candidate for justifying the right to refrain from reproduction on utilitarian grounds. However, there are some considerable objections. Even if the utilitarian could adequately respond to the above objections, utilitarianism would still not provide justification for reproductive rights. The fact that not all societies have the right to refrain from reproduction as their norm shows that Sidgwickian motive utilitarianism does not always provide justification for the right to refrain from reproduction. However, if motive utilitarianism cannot justify this right all of the time, are utilitarians really acknowledging the right to refrain from reproduction as a right? I think not. If utilitarianism grounds an obligation to reproduce on utility maximization, utilitarianism does not provide justification for the right to refrain from reproduction. I turn next to Brink’s view of utilitarianism and consider whether or not he succeeds in justifying the right to refrain from reproduction. Brink, who calls himself an
‘objective utilitarian’ (OU), sees some things other than pleasure as objectively valuable. His view therefore promises not to require people to reproduce on utilitarian grounds.

Brink

Describing OU, Brink writes,

OU will not interfere greatly with an agent’s important projects and commitments… More important, as OU recognizes, special value attaches to people’s having and pursuing personal projects and to the development and maintenance of closer personal relationships involving mutual concern and commitment. The moral importance of an agent’s personal projects and commitments insulates agents from certain other-regarding demands.90

It seems that, since choices regarding reproduction are apparently justified under this description, the right to refrain from reproduction must be granted. It seems that the utilitarian can respond to my argument that his theory requires individuals to reproduce when doing so will increase average (or total) utility, by arguing that utility can be construed so as to take into consideration the individuals’ preferences and desires, including preferences about reproduction.

However, that construing utility as to take into consideration preferences and desires adequately justifies the right to refrain from reproduction turns out not to be the case. If two people have different desires and the first person’s desires promote more utility then, on utilitarianism, the fulfillment of the preferences of the first person should be encouraged and promoted and can even override the fulfillment of the second person’s preferences. So, the desires of the second person are not being regarded as important and, hence, the second person is not being treated as inherently valuable.

It should be noted that something is not good simply because it is desired or desirable. And when something is good or valuable it is not merely because of the desire to have or to do that something. Some desires are for morally wrong things. For example, one may have the desire to kill another without adequate moral justification. Also, some things are desirable for reasons that are independent of the fact that we desire them. I see the individual as having a higher moral status than what the utilitarian seems to grant. Individual freedom and autonomy are inherently good. John Christman writes, “Individual autonomy is an idea that is generally understood to refer to the capacity to be one's own person, to live one's life according to reasons and motives that are taken as one's own and not the product of manipulative or distorting external forces.”  

Individuals are valuable in themselves and not merely as a means for promoting utility, even if utility is construed so as to take into consideration preferences and desires. That is, individuals ought to be treated as autonomous individuals and the utilitarian does not do this.

The objective utilitarian can respond that she has taken autonomy into consideration when making a utilitarian calculation and can avoid my criticism. Brink writes, “OU attaches important value to the pursuit and realization of personal projects [and] freedom or autonomy is required for realizing these values. Thus, OU will give high priority to personal autonomy.”  

He continues, “Because of the importance of autonomy and basic well-being, any individual’s interest in these goods will outweigh the

---


value of large magnitudes of other goods.”

So it seems that the utilitarian has a way to justify the individual’s autonomy.

In response I argue that while autonomy may be worth more than other things in the utilitarian calculation, there are still times when it is possible to increase the autonomy of more individuals by limiting the autonomy of a single individual. In that case, the utilitarian will demand that the autonomy of the individual be ignored and not respected. But this is treating the individual not as an autonomous being, but as a means to the end of greater utility, even if the autonomy of individuals is included in the utility calculation. In every case, the individual is being treated as a means to the end of greater utility and this violates autonomy. Morality, properly understood, should require that people be treated as ends in themselves. Although the utilitarian may claim to take autonomy into consideration when calculating utility, he ultimately does not respect the individual as an end in himself. That is, the individual is still simply being compared with other individuals to the degree that she maximizes utility. That is, he is being treated as a means to an end. Here I have in mind a Kantian conception of respect.

Robin Dillon explains this type of respect. He writes,

To respect persons is thus to regard them as absolutely, unconditionally, and incomparably valuable, to value them in themselves and not just in comparison to others or insofar as they are valuable to someone or could be useful as a means for furthering some purpose, and to acknowledge in a practical way that their dignity imposes absolute constraints on our treatment of them.

This is what I have in mind when I write of respect for individuals.

---

93 Ibid., 270.
The utilitarian may retort that the utility to be maximized can be understood as respect-constrained utility and that the autonomy of the individual is therefore respected. Brink writes, “[I]n order for the pursuit or realization of a project to be of value, that project must, among other things, respect other people at least in the minimal sense of not causing significant and avoidable harm.” But this does not help the utilitarian case. Even if respect is considered in the utilitarian calculation, it is only a superficial respect for the individual. When respect is something to be maximized, there will always be instances when the respect of the individual can be violated if it promotes more respect elsewhere. The utilitarian treats the individual with respect only insofar as doing so will maximize utility (even when respect constitutes utility). For example, consider a female, who wants to become pregnant, and her partner, a male, who does not want her to become pregnant. They have discussed their desires and have been trying to reach an agreement about what to do. Imagine that the female partner intentionally misuses her birth control in order to become pregnant. She has, thus, disrespected her partner and used him as a means to an end. Furthermore, since the child to be brought into existence provides opportunity to respect a person, in the utility calculation the female partner’s actions are justified. The objective utilitarian may respond that utilitarians should take into consideration respect for persons and treat individuals not merely as means. In that case, the objective utilitarian is no longer utilitarian.

So far I have attempted to show that the utilitarian does not truly treat individuals with respect as inherently valuable, autonomous individuals with important preferences and desires of their own. The utilitarian may respond that my reasoning is not correct.

---

95 Ibid., 264.
and that objective utilitarianism can provide better justification for moral intuitions such as the right to refrain from reproduction and the inherent value of respect and autonomy. However, it is not clear whether or not the objective utilitarian has the moral intuition that persons should be treated with respect as autonomous individuals, as I do. If the utilitarian agrees with me, and argues that she does respect the autonomy of individuals in the same way that I do and independent of utilitarian calculations, then she is no longer holding a utilitarian position. Furthermore, there would be no advantage to utilitarianism. Utilitarianism is supposedly an easily applied moral theory because one can do a utility calculation and know which action to do. There is supposedly specificity of what is to be maximized and a determinate answer to what one ought to do. But if the utilitarian construes utility in such a way that respect and autonomy are inherently valuable, it seems impossible to claim any advantage to utilitarianism. Utilitarianism would not be distinct from other kinds of moral reasoning.

On the other hand, the utilitarian could respond to my question about moral intuitions concerning respect and autonomy by suggesting that it is not clear that it is wrong to violate someone’s rights or autonomy. In this case, I think utilitarianism must be rejected. In this chapter I have been exploring ways in which utilitarianism can conform to my convictions that certain moral rights, in particular the right to refrain from reproduction, should not be violated. Utilitarian prescriptions can conflict with such rights. This is why I reject utilitarianism. In the end, utilitarianism does not justify the right to refrain from reproduction.
CONCLUSION

To conclude, I have explained the problem of the Repugnant Conclusion and critiqued three responses to it. Average utilitarianism succeeds in solving the problem of the Repugnant Conclusion but there are non-utilitarian arguments against it regarding reproductive rights. The right to refrain from reproduction is justified on a non-utilitarian view because it is necessary in order to respect persons as autonomous individuals. Appealing to Sidgwick’s view that utility provides a criterion but not a motive for action is one way for a utilitarian to answer these criticisms. However, as I have noted, this approach is not without problems of its own and it does not solve the problem of the Repugnant Conclusion. Also, appealing to objective utilitarianism does not provide sufficient justification for reproductive rights. The average utilitarian can solve the problem of the Repugnant Conclusion but cannot provide an adequate justification of rights, including the reproductive rights.
WORKS CITED


