

4-2019

Egoism and the Repugnant Conclusion

Nathaniel Anderson

Follow this and additional works at: <https://scholarworks.wm.edu/honorsthesis>



Part of the [Applied Ethics Commons](#), and the [Other Philosophy Commons](#)

Recommended Citation

Anderson, Nathaniel, "Egoism and the Repugnant Conclusion" (2019). *Undergraduate Honors Theses*. Paper 1355.
<https://scholarworks.wm.edu/honorsthesis/1355>

This Honors Thesis is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.

Egoism and the Repugnant Conclusion

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Arts in the Philosophy Department from
The College of William and Mary

by

Nathaniel Anderson

Accepted for High Honors
(Honors, High Honors, Highest Honors)

Philip Swenson
Dr. Philip Swenson, Director

M. Swetnam-Burland
Dr. Molly Swetnam-Burland

Noah Lemos
Dr. Noah Lemos

Williamsburg, VA
May 3, 2019

Nathaniel Anderson

Honors Thesis: Egoism and the Repugnant Conclusion

The repugnant conclusion has flummoxed philosophers since its appearance in such works as Derek Parfit's *Reasons and Persons*. The conclusion, that, for any world of people living very good lives, there is another, better, world of far more people living far worse lives, struck Parfit as deeply unintuitive, which drove him to label it repugnant.¹ In section 1, I will explain the intuitions and arguments underlying the repugnant conclusion and discuss why several proposed solutions fail to satisfactorily avoid the repugnant conclusion. In section 2, I will argue that ethical egoism provides a plausible way to dodge the repugnance of the repugnant conclusion by denying that the repugnant conclusion gives rise to any ethical implications. Section 3 will be devoted to answering some of the most serious objections to ethical egoism, while an entire section, section 4, will address the particular objection that collective action dilemmas represent a fatal flaw in ethical egoism. Finally, section 5 will discuss the one-person repugnant conclusion and argue that it is not a fatal problem for the egoist response.

The repugnant conclusion can come about from Parfit's impersonal total principle, which states that, other things being equal, the outcome which contains the most of whatever makes a life worth living is the best outcome.² This principle is a natural response to the challenge implicit in Parfit's non-identity problem, and thus should not be discarded out of hand.³ If adopted, however, it makes explicit that any amount of quality of a life may be compensated for by a sufficient amount of lives of lower quality, which leads directly to the repugnant conclusion. A second, and stronger, argument for the repugnant conclusion comes from the so-called Mere Addition Paradox. Parfit

¹ Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), 388

² Parfit, *Reasons and Persons*, 387

³ Parfit, *Reasons and Persons*, 357-361

argues that adding people with lives worth living, even if they are only barely worth living, does not lower the value of a world. Thus, given a world A of 10 people living very good lives (+100), there is a world A+ containing ten people living very good lives (+100) and ten people living lives barely worth living (+1) which is not worse than A. In turn, there is a world B with 20 people living lives that are better than the average value of the lives in world A+, perhaps +55, which is better than world A+ because it is more equal and has a higher amount of whatever makes a life worth living. World B seems to be better than world A because B is better than A+, which, in turn, is not worse than A. Repeated applications of the principles, addition of people with lives barely worth living followed by leveling which increases the total value of lives, will eventually produce a world Z with a vast number of people living lives barely worth living. Therefore, the plausible principles of the mere addition paradox also produce the repugnant conclusion.⁴

Section 1A-The Bounding Response

One style of approach is to limit the amount of value which certain types of lives can contribute to the value of a world. Critical level theories, which postulate that lives below a certain threshold have value for the people who live them but not for the world, are one example. In “Resolving the Repugnant Conclusion,” Tyler Cowen develops another approach which suggests that bounding offers a way to avoid the repugnant conclusion.⁵ Bounding refers to setting a limit on a certain type of value. Cowen suggests that bounding the amount of value which lives barely worth living contribute to the value of the world blocks the repugnant conclusion in both the total utility derivation and the mere addition derivation. He further suggests that boundedness makes

⁴ Parfit, *Reasons and Persons*, 419-441

⁵ Tyler Cowen “Resolving the Repugnant Conclusion” in *The Repugnant Conclusion: Essays on Population Ethics*, ed. by Torbjorn Tannsjo and Jesper Ryberg (Dordrecht: Kluwer Academic Publishers, 2004), 81-97

sense in light of the value of societal flourishing or civilization which have holistic value. While he is correct that imposing a limit on the amount of value which certain types of lives can contribute to the value of a world, his bounding leads to severely unintuitive conclusions of its own, namely that a much smaller population with slightly better lives is better than a much larger population with slightly worse lives.

The standard account of the life barely worth living, suggested by Parfit himself in *Reasons and Persons*, is the life of Muzak and potatoes. Cowen argues that we should bound the amount which such lives can contribute to the value of the world. He suggests that each tier of life might have a different bound, so that lives of muzak, potatoes, and yearly tennis matches might be able to add more value than the lives without the tennis matches.⁶ Thus, a world of fifty billion people enjoying muzak and potatoes might reach the maximum value of multiplying such lives and be worth only twenty billion units despite each life being worth more than four tenths of a unit to the person who lives it, while a world of fifty billion people, half of whom experience the yearly tennis match, might be worth forty-four billion units, twenty billion for the mere muzak and potato people, and another twenty-four billion representing the maximum value which could be contributed by the additional tennis match people. The bounded theory, however, maintains that lives worth living never have negative value, although the marginal value of each additional life may approach zero at some point. Thus, such a system of bounding has advantages over critical level theories which postulate that lives under the critical level detract from the value of a world, because it does not imply the sadistic conclusion, namely that a world with a certain number of people with lives not worth living could be better than a world with more people with lives worth

⁶ Cowen, "Resolving the Repugnant Conclusion," 85-86

living.⁷ The bounded theory, then, can be thought of as a stronger version of the critical level theory.

The bounded theory, however, runs into the same problems with thresholds which all critical level theories are susceptible to. Once a world reaches the bound for the type of life the people in that world have, its value caps out. That means that a world, A, of a trillion people getting muzak and potatoes might be worse than a world, B, of twenty-five billion who have the additional yearly tennis match. The marginal value of the tennis match, however, seems far too slight to imply that a life with it in world B is more than forty times better for the world than the life without it in world A. Now, it is true that thresholds exist in all theories which postulate that lives can be worth living or not worth living, but in most such theories the threshold is set at the point at which a life becomes worth living. While there is a similarly odd difference in value between lives just barely worth living and lives just barely not worth living, that threshold is less unintuitive than the thresholds in the bounded theory because there are presumably many more of them, and they lack even the limited explanatory force afforded by the vague idea of worth living versus not worth living. It seems possible that the threshold in most theories between worth living and not worth living could be clarified, perhaps by a question such as if you knew before your life began exactly what it would hold, but you would forget once your life began, would you choose to live that life? Cowen might reasonably argue that I have made the thresholds seem too slight, that a tennis match cannot be enough to elevate the muzak and potatoes life to another tier, but it seems that at some point something as small as a tennis match must make the difference between one level and the next. As Tim Mulgan has identified, the problem with such puzzles is that step differences in

⁷ Gustaf Arrhenius, "Critical-Level Theories" in *Future Generation: A Challenge for Moral Theory*, (uu-diva-portal.org, 2000), https://www.academia.edu/3488242/Future_generations_A_challenge_for_moral_theory, 72-86

degree must eventually become differences in kind, but such step differences always seem too small to justify such a lexical shift.⁸ Thus, Cowen must offer some explanation for what justifies the difference in valuation between two tiers of lives.

Another objection comes from considering the kinds of actions which can improve the value of a world. Imagine a world, A, with fifty billion people living the drab muzak and potatoes life. Now imagine a world, B, which is identical except that ten of these people live isolated on an island where they have no contact with the rest of the population. These ten people brought muzak recordings and potatoes with them to the island and in all respects their lives are like those of the rest of the population. One of the people on the island, however, decides to murder the others and takes great pleasure in the commission of the deed so that his life is eight times more valuable for him than each of the lives of the others was for each of them. The other nine people do not realize that the murders are taking place and are in no way pained by the murders, and thus the only consequence of the murders is to end their lives. It would seem that the murders have raised the value of world B over the value of world A. The marginal value of nine people getting muzak and potatoes is insignificant because the worlds have reached, or at least come asymptotically close to, the maximum value which muzak and potatoes can provide, while the value of a single person living a good life because of the pleasures of accomplishing his goals is undiminished by bounding because he is the only person living that category of life. It seems unacceptable that a theory would suggest that murder of the innocent can raise the value of the world, particularly when it reduces total utility, and thus bounding the value of muzak and potatoes seems problematic.

A number of replies are open to Cowen. The first reply is that the amount which the lives of murderers can contribute to the value of the world is bounded at zero. Thus, the murderer

⁸ Tim Mulgan, "Two Parfit Puzzles" in *The Repugnant Conclusion: Essays on Population Ethics*, ed. by Torbjorn Tannsjo and Jesper Ryberg (Dordrecht: Kluwer Academic Publishers, 2004), 24-26

reduces the value of the world by murdering because he eliminates the small but nevertheless positive value of nine lives of muzak and potatoes and adds no other value to the world. Initially this approach seems to solve the problem, but now imagine a world, C, just like world B except that the murderer takes no pleasure in his actions. The day after the murders he just sits around eating his breakfast of potatoes and listening to muzak as if nothing has happened. It seems that world C is worse than world B. All things being equal, it is better if one world contains more pleasure than another. That response may seem like question begging. It may seem like a mere assertion of the repugnant conclusion. If the *ceteris paribus* clause is taken at its most strict, however, the repugnant conclusion is not an inevitable result. My assertion only applies to worlds with the same population, the same amount of virtue and vice, the same relation of desert to receipt, and so on, and thus does not represent a mere endorsement of the repugnant conclusion. It seems that if one world contains more of a good and is in no other way different than another, it must be better than that other. Thus, bounding the value which the lives of murderers can contribute to the value of the world at zero leads to another unintuitive result.

Following G. E. Moore, Cowen could counter with the argument that taking pleasure in the bad is intrinsically bad, and thus that the fact that the murderer is pleased by his murder makes the murder even worse for the world.⁹ While it may be true that taking pleasure in a bad thing is intrinsically bad, it seems to me that it would not be a large enough intrinsic bad to outweigh the intrinsic good of the episode of pleasure. Furthermore, the murderer's pleasure could be renamed or reexplained in a way that would not make it a pleasure in the bad, perhaps taking pleasure in the fulfillment of a goal or plan. Many moral theories suggest that people have a duty to improve the world, so much of the murderer's pleasure could even be coming from the belief that he has

⁹ G. E. Moore, Chapter VI: The Ideal in *Principia Ethica*, (Buffalo: Prometheus Books, 1988), accessed online at <http://fair-use.org/g-e-moore/principia-ethica/>, Sections 124-125

improved the world by elevating his life into another tier. Under such a description, it seems even more unlikely that the intrinsic badness of taking pleasure in an evil would outweigh the intrinsic goodness of both the initial pleasure in the murder and the subsequent pleasure in improving the value of the world, the value of which would be amplified as a pleasure taken in the good.

A third option available to Cowen would be to appeal to the *ceteris paribus* principle implicit in comparisons of worlds. Arguments about the repugnant conclusion tend to hold all other things equal to avoid debates about whether moral actions or the correspondence of desert and receipt contribute to the value of a world. Cowen could thus argue that the murders violate the *ceteris paribus* condition by adding moral wrongs to the world, and that the disvalue of those moral wrongs exceeds the positive value of the murderer's life. This is a reasonable response, but it requires that the amount of moral and immoral actions alter the value of the world, which is far from clear. It also requires that the murders were a moral wrong, which might not be the case if they were motivated only slightly by the disposition to be pleased by murder and far more by the desire to improve the value of the world. Furthermore, the case can be amended again to avoid a moral wrong while retaining the unintuitive conclusion. Imagine a world, D, like all of the other worlds except that the other people have not been murdered but annihilated by a blast of cosmic radiation which has also caused the growth of peaches on the island. The remaining man on the island might, as a result of the addition of peaches to his diet, live a life that was twice as good for him as his previous life, and which made it into the next tier of lives in terms of its value for the world. Here the *ceteris paribus* requirement seems to be infringed far less, but the unintuitive conclusion, that the deaths of nine people can be good for the world even when the total utility of the world also declines, remains true.

Section 1B-The Perfectionist Response

Another style of response is to maintain that certain goods are so good that a world without such goods is worse. Parfit argues that a world without the best things in life, such as the music of Mozart, cannot be as good as a world which contains only inferior goods, such as the music of Haydn, even if those goods are more plentiful.¹⁰ The response from the best things in life fails, however, to avoid the repugnant conclusion in cases in which people's lives are barely worth living because their receipt of the best things in life is almost cancelled out by their receipt of many pains.¹¹ It also involves dubious claims about the lexical superiority of higher quality goods over lower quality goods, leading to a total neglect of quantity in favor of an exclusive focus on quality.

In "Can We Avoid the Repugnant Conclusion," Parfit claims that we can assume that the repugnant conclusion focuses on a world of the type that he calls "Drab Z." He says that in Drab Z, people live lives barely worth living because they receive exceedingly minimal goods but do not receive anything that detracts from the values of their lives. The people in Drab Z lack the better things in life, and thus the lexically prior lives of people in world A who live lives very much worth living make world A clearly better than Drab Z.¹² Parfit conveniently ignores the world which he calls "Roller Coaster Z." In Roller Coaster Z, people live lives that are barely worth living because they contain ills almost equal to the substantial goods that they contain.¹³ The lives in A do not seem to be lexically prior to the lives in Roller Coaster Z because the lives in the latter contain the same high-quality episodes present in the A lives, but with the addition of very

¹⁰ Derek Parfit, "Overpopulation and the Quality of Life" in *The Repugnant Conclusion: Essays on Population Ethics*, ed. by Torbjorn Tannsjo and Jesper Ryberg (Dordrecht: Kluwer Academic Publishers, 2004), 16-20

¹¹ Parfit arbitrarily assumes a bleak lives version of the repugnant conclusion in *Reasons and Persons*, 388, but the conclusion does not seem to depend on that assumption.

¹² Derek Parfit, "Can we Avoid the Repugnant Conclusion," *Theoria* vol. 82, Issue 2, May 2016, 118

¹³ Parfit, "Can we Avoid the Repugnant Conclusion," 118-119

low-quality moments which detract from the value of a life. Without lexical priority, the vast number of lives in Roller Coaster Z collectively seem to contain more value than the much smaller number of lives in World A. If Roller Coaster Z is better than World A, then the repugnant conclusion holds because it claims that for every world of people living moderately good lives, there is *some* world of many more people living lives barely worth living. It is not an objection to the repugnant conclusion that not all worlds of the generic Z type are better than world A so long as at least one world of the Z type is.

Furthermore, the claim that lives of higher quality are lexically prior to lives of minimal quality is also dubious. Parfit criticizes what he calls the “Linear Model.”¹⁴ The linear model describes a theory of value whereby different episodes of good or ill can be compared to one another exactly because their values correspond to positions on a number line. Thus, an exceptional ice cream cone might be exactly 4.8 times better than a mediocre oatmeal raisin cookie. Parfit claims that this view is naïve and proposes that it fails when confronted with cases such as deciding whether Einstein or Bach was a greater genius. In such cases, we might conclude that Bach and Einstein are equally good, but that might entail a contradiction when we compare both of these two to some third person like Oppenheimer. We might think that Einstein was a greater genius than Oppenheimer, but that Oppenheimer was not less of a genius than Bach which would entail both that Bach was as great a genius as Einstein and less of a genius than Einstein. Parfit proposes that Bach and Einstein might be imprecisely equally good, so that “not less of a genius than” does not mean “at least as much of a genius as.” The weaker “not less than” is not transitive so that Bach could be not less of a genius than Einstein and Oppenheimer could be not less of a genius than Bach without Oppenheimer being not less of a genius than Einstein. Such a non-transitive

¹⁴ Parfit, “Can we Avoid the Repugnant Conclusion,” 114-115

system requires that Einstein, Bach, and Oppenheimer cannot be placed on a numerical scale, because if they could be so placed, then any relations between them would be transitive.

While imprecision seems reasonable in the evaluations of concepts such as greater genius than, it only seems applicable there because greater genius is an incomplete concept. Temkin describes an incomplete concept as one that “allows us to make some comparisons between [geniuses] as to which is greater but not others... due to the roughness or complexity intrinsically involved in the notion.”¹⁵ Parfit explicitly claims that incompleteness is not the source of the problem but offers no defense of his claim. Genius plausibly describes many different phenomena. In contrast, it is plausible that good might only describe one type of phenomenon. Monist theories argue for the existence of a single bearer of intrinsic value, such as pleasure. If something like pleasure is the only good, then it seems that any two episodes could be compared through the concept of desirability. Heathwood argues that “a sensation S, occurring at time t, is a sensory pleasure at t iff the subject of S desires, intrinsically and de re, at t, of S, that it be occurring at t.”¹⁶¹⁷ On Heathwood’s view, if an agent desires A more than B, then it seems that A must be more pleasurable than B. Exact relations could be found by assessing how much of some very minor pleasure or pain (which Heathwood calls uncomfortableness)¹⁸ would be required for the agent to forego some experience. If I will only forego an ice cream cone for at least three chocolate bars, then it seems plausible that the ice cream cone is exactly as good as three chocolate bars. Similarly, parity between pleasures and pains could be explored by asking how much of some good would

¹⁵ Larry Temkin, “Intransitivity and the Mere Addition Paradox,” *Philosophy and Public Affairs* vol. 16(2) 1st April 1987, 145

¹⁶ Chris Heathwood, “The Reduction of Sensory Pleasure to Desire,” *Philosophical Studies* 133 (2007), 32

¹⁷ Fred Feldman suggests a similarly reductionist account of at least sensory pleasures, but he posits that they are reduced to attitudinal pleasures which, nevertheless, seem to embody much of the same pro-mental states which Heathwood invokes. *Pleasure and the Good Life* (Oxford: Oxford University Press, 2004), 79-83

¹⁸ Heathwood, “The Reduction of Sensory Pleasure to Desire,” 41-44

be required in order for you to submit to some evil. We might ask how many ice cream cones I would have to give you in order for you to let me punch you in the face and reasonably expect an answer.

Parfit's response, then, seems to assume some non-monistic account of the good. When we introduce comparisons between different types of goods, it seems much more plausible that we might not be able to stipulate how much of one equals some amount of another. Parfit's response, however, is inadequate insofar as it assumes that monism is false. The repugnant conclusion is axiology agnostic: it is applicable for all descriptions of the good. If a response can only satisfactorily avoid the repugnant conclusion on a limited number of accounts of the good, then it is inherently unsatisfactory in the absence of a proof for the truth of one of the descriptions of the good for which the response obtains. Thus, because Parfit fails to take into account the problem posed by Roller Coaster Z, and because his approach only seems to address the repugnant conclusion on non-monist theories of the good, it is not adequate in and of itself. Someone might try to salvage Parfit's response by maintaining that if monism leads to the repugnant conclusion, then we have additional reason to believe pluralism. If, however, the ethical implications of the repugnant conclusion are as minimal as I maintain that they are, then avoiding the repugnance of the repugnant conclusion does not offer any meaningful advantage for pluralism.

Section 1C-Justicism

Fred Feldman proposes a solution to the repugnant conclusion, Justicism, which posits that the value of a world is determined not only by how much of what makes a life good people get, but also by how much the people who get it deserve it.¹⁹ Feldman further claims that all people, in

¹⁹ Fred Feldman, "Justice, desert, and the Repugnant Conclusion," 204

virtue of being people, deserve a substantial amount of pleasure.²⁰ When a person receives what is good, and deserves it, the value of the good for the world is increased, while when a person gets far more than he deserves, or fails to get anywhere near what he does deserve, then the value of the world is reduced.²¹ The addition of desert to the value of a good for the world enables Feldman to say that the world composed of quintillions of people eking out a tiny amount of good is not merely not the best world, but in fact is abhorrent. Each person seems to get far less than he deserves, and each person who falls far short of what he deserved makes the world worse, so the world is in fact extremely bad. The world of excellence, on the other hand, is made even better by the addition of desert, because the good that each person receives is amplified by the fact that the person deserved it.

While Feldman thus adroitly circumvents the main thrust of the repugnant conclusion, his theory leads to some odd conclusions of its own. For one, it seems that the world of infinite people barely getting any good could still be better than the world of excellence if the infinite people deserved what they were getting. The implication is that a world of people with an iota of desert getting a little good might be better than a world of saintly people getting the same good. Furthermore, good people living lives worth living might make the world worse if their desert was sufficiently higher than their receipt. If people realized that they were unlikely to get much of what makes a life good, then in the interest of raising the value of the world, it would behoove them to think wicked thoughts, and possibly even to take wicked actions, so that the gap between their desert and their receipt would be eliminated. Of course, it might be the case that trying to perform morally bad actions in order to make the world better is itself morally good so that these people would only make their desert higher by trying to lower it for such a reason. It seems

²⁰ Fred Feldman, "Justice, desert, and the Repugnant Conclusion," 209

²¹ Fred Feldman, "Justice, desert, and the Repugnant Conclusion," 204-205

counterintuitive to believe that the world might be made better by vice and worse by virtue, and thus there is some grounds for dismissing Feldman's theory.

Feldman attempts to answer this objection by arguing that in order for the desert of the people in the infinite world to be so low, then they must have done some atrocious things, which would massively lower the values of the lives in that world. In his example, players in a game may inflict pain in order to lower their desert. They each inflict substantial pain in the first round, while their desert is still high, and thus the value of the world is greatly reduced. Only in the second round is there any gain in value, when each person gets the iota of pleasure he still deserves, and it fails to compensate for the bad of the first round.²² Feldman's answer is deceptive, because in his example each player gets on net a tremendous amount of uncompensated pain, whereas in the infinite world, it is stipulated that everybody gets a slight net pleasure. If instead of inflicting pain, the players in Feldman's example had tried to inflict pain and failed, then they would have similarly lowered their desert without violating the stipulations of the example, but, in that case, each person would have contributed only the slight net positive of the second round, and thus if there were an infinite number of players, that world would have had infinite positive value.

Feldman might insist that acts of virtue are intrinsically good, and acts of vice are intrinsically bad. He would then be able to say that the world of a great many people with a small amount of welfare and a small amount of desert would be very bad because, while the desert and receipt of those people would be in line, they would have had to commit many morally bad actions to lower their desert so far. The intrinsic badness of those actions would outweigh the intrinsic goodness of lives with positive welfare proportional to their desert, leading to a very bad evaluation of the world. One response is to insist that the moral value of an agent's actions is just one

²² Fred Feldman, "Justice, desert, and the Repugnant Conclusion," 212-213

component of the value of his life. The all things considered value of a life would thus be obtained by adding the moral value of actions to the desert modified welfare value of a life. On such a view, the lives of the people who committed many morally bad actions would not be lives barely worth living: they would be substantially below the threshold at which a life becomes worth living, and thus the fact that the world of many such lives is bad would not be sufficient to show that world Z is bad because the people in the morally abhorrent world would not be living lives worth living.

Feldman, however, argues that intrinsic value for the person and intrinsic value for the world are divorced, and thus he would be able to assert that morally bad actions reduce the intrinsic value of the world but not the intrinsic value of any given life. If that is the case, a world Z in which everyone deserved little because of morally bad actions and received a corresponding amount would be very bad because while each life would be barely worth living, each life would contribute a net negative amount to the intrinsic value of the world. One problem with such a view is that it seems to imply that a world of many people who perform good actions but yet receive less than the amount required to live lives worth living might be a very good world. While the desert adjusted welfare of their lives would detract from the value of the world, their virtuous actions might contribute more to the value of the world. Feldman claims that as receipt increases beyond desert, the desert adjusted contribution to the world approaches an asymptote. It seems plausible that there would be a parallel occurrence as desert rises past receipt. The more good actions the people performed, the more valuable the world would become due to their virtue, while the amount which their net negative welfare contributed to the value of the world would eventually level out. Eventually, the virtue would outweigh the lack of welfare, with the bizarre result that a world with no enjoyment whatsoever, with pain but no pleasure, could be better than a world containing vast amounts of pleasure.

Section 1D-Acceptance

A final style of response is just to insist that we ought to accept the repugnant conclusion. Torbjorn Tannsjo, a proponent of this response, argues that the life barely worth living is akin to the life of an affluent person today.²³ He maintains that once we view the life barely worth living as not significantly different from the lives that we ourselves lead, we will acknowledge that a world of a great many people living such lives is in fact better than a world of many fewer people living much better lives. Such a view assumes that the repugnance of the repugnant conclusion inheres in the perceptions which people have of what the life barely worth living looks like. In particular, Tannsjo implicitly denies that people are repulsed by the theoretical principle that quality can be replaced by sufficient quantity, which suggests that most people are not convinced by Parfit's claim that the best things in life are lexically prior to many other good things. To the extent to which people feel the intuitive pull of Parfit's claim, then, they have reason to deny Tannsjo's assertion.

For the sake of argument, I will grant Tannsjo's claim that the life of an affluent person living in the west is a life barely worth living. Nevertheless, Tannsjo's view remains problematic because it exploits innate Pollyannaism, which David Benatar ascribes to the vast majority of individuals, in order to bias our view of the repugnant world.²⁴ Pollyannaism refers to the tendency to view our lives as better than they are. Tannsjo insists that our lives are lives barely worth living, which transfers our appraisal of our lives as well worth living to the life barely worth living so that we also view that life as well worth living. In actuality, however, both our lives and the life barely

²³ Torbjorn Tannsjo, "Why We Ought to Accept the Repugnant Conclusion" *Utilitas* vol. 14(2) (November 2002), 345

²⁴ David Benatar, *Better Never to Have Been*, (Oxford: Oxford University Press, 2006), 64-69

worth living are only barely worth living and thus our overly optimistic evaluation serves to make the repugnant conclusion appear far more attractive than it actually is. When we realize that our perceptions are biased, the repugnance returns in full force.

It will help to illustrate the point with an example. Suppose that Jake offers you two choices of future diets. The first choice is a diet of nothing but oatmeal for the rest of your life. Sufficiently sweetened, each bowl will bring you the slightest amount of pleasure, but not an iota more. The alternative is a revolutionary treatment which will remove both the need and the ability to eat without changing anything else about your life. If you choose this option, however, you will first be treated to ten incredibly delicious feasts which will produce a great deal of pleasure. Of course, the cumulative total of tens of thousands of small amounts of pleasure will exceed the pleasure from the great meals. Nevertheless, it seems plausible that you would choose the ten exceptionally good meals.

Now imagine that you defer the choice to the next day and go to lunch with a friend. The meal is expensive, and the food served is exactly the same oatmeal which you would have received if you selected the first diet, but you are having a great time with your friend and cognitive dissonance is also working to ensure that your evaluation of the meal is favorable. When you leave the restaurant, Jake reoffers the choice of diets, but now he presents the first diet as a lifetime of meals exactly as tasty as the one which you just ate. Under the circumstances, you would be much more likely to choose the lifetime of oatmeal because your innate distaste for that option has been camouflaged by other psychological factors, the association with seeing your friend and the cognitive dissonance that will result if you acknowledge that you wasted your money on the food. When Tannsjo suggests that the life of an affluent westerner today is the life barely worth living, he implicitly ties it to false judgments about its worth in the same way in which Jake does. We

should be wary of drawing strong conclusions from an evaluation as optimistic and vulnerable to bias as our evaluations of our own lives, and thus the observation that the life of an affluent westerner today is a life barely worth living should not change our judgments of the repugnant conclusion.

The significance of Tannsjo's claim would be easier to evaluate if it could be divorced from claims about our lives and lives like ours and thus made less susceptible to our innate biases. Perhaps the life of an affluent person in the west today might be comparable to the life of a dutchman of the mid-17th century when the quality of life in the Netherlands was very high relative to the rest of the world in the way that the western life is today. Imagining a world full of well-off dutchmen, however, might still seem quite repugnant. They lacked modern medicine, plumbing, and electricity, and were constantly at war with the English, French, and Spanish. My intuition that a world full of such lives is good is rather weaker than when I imagine lives like my own. Now it might just be the case that I am mistaken about the relative value of lives in the mid-17th century Netherlands and the modern west, but I nevertheless think there are grounds to be suspicious of our tendencies to overvalue the familiar and thus grounds to doubt the significance of Tannsjo's claim about what constitutes the life barely worth living even if it turns out to be true.

Even if Tannsjo is right that the life barely worth living looks like an affluent life today, and that that revelation substantially changes our intuition about the world full of such lives, it still seems that the repugnant conclusion remains repugnant because of the type of moral duties which it justifies. Faced with the unsavory ethical implications of the repugnant conclusion, however, we can turn to ethical egoism.

Section 2: Ethical Egoism offers a Solution to the Repugnant Conclusion

Ethical egoism, roughly speaking, is the view that people are morally required to take the available action which maximizes their welfare.²⁵ Even if people are willing to accept the viability of ethical egoism, it is not immediately clear how ethical egoism enables them to refute the Repugnant Conclusion. In one sense, ethical egoism does not. Ethical egoism, being axiology agnostic, does not take a stance on the value theoretic aspects of the Repugnant Conclusion; it does not stake a position on the value of worlds. It may just be the case that the world of a great many people living lives barely worth living actually is the best world in some impersonal moral sense,²⁶ but ethical egoism addresses the truly repugnant part of the problem by allowing people to deny that the repugnant conclusion gives rise to any substantial ethical implications.

What kinds of ethical implications might the repugnant conclusion give rise to? The main ethical implication of the repugnant conclusion lies in what world an agent would be obligated to create. If an agent had to choose to create a world, the agent would have to create a repugnant world, A, of a trillion people getting one unit of whatever makes life worth living over a non-repugnant world, B, of a billion people getting ten units of whatever makes life worth living. He would even have to choose world A if the trillion people were replaced by a trillion ladybugs getting one unit of whatever makes life worth living.²⁷ These implications are unpalatable to say the least because almost everyone believes that it would be at least permissible to choose world B regardless of how many myriads of ladybugs with lives barely worth living populate world A. This strong intuition suggests that either the value theoretic claims of the repugnant conclusion are

²⁵ I will offer and justify a more precise definition in Section 3

²⁶ Parfit describes the impersonal moral sense as applying to outcomes, as opposed to persons or acts. He writes that "it would be better, in this sense, if fewer people suffer from some crippling illness, or if the Lisbon earthquake had not occurred." (*Reasons and Persons*, 385) I believe that ethical egoism is compatible with an impersonal moral sense on which an outcome is better than another if it contains more total welfare.

²⁷ Parfit, "Can we avoid the Repugnant Conclusion?" 118

wrong or that we must adopt a moral theory which does not derive ethical duties from the claim made by the repugnant conclusion.

Utilitarianism and deontology are unable to deny all of the ethical implications which arise from the repugnant conclusion unless they embrace what I will call an *only real people matter* standpoint. So long as they believe that the interests of merely potential people matter, they cannot avoid repugnant implications. The utilitarian finds himself in the worst position because he must admit that total value is the only morally relevant criterion, and, given the massive sum of wellbeing in world A, it is morally required to choose it. The deontologist is on slightly better ground because he might be permitted to choose world B depending on the distribution of people to whom he has special duties. Nevertheless, if worlds A and B are equal in all other relevant attributes, he would have to choose world A because his duty of beneficence is stronger towards the people who will exist in world A than those in world B given the far greater total benefit in A than B. The deontologist may appeal to the fact that in any practical situation there will be some people to whom he has special duties, but nevertheless both the deontologist and the utilitarian have trouble convincingly denying both duties derived from the repugnant conclusion.²⁸

If ethical egoism is correct it allows us to dodge any ethical implications deriving from the repugnant conclusion. On ethical egoism it does not matter that the life of another person might increase the value of the world if creating that life is not also the available act which maximizes the welfare of the agent, and thus nobody would have a moral duty to procreate unless it were already in his best interests to do so. Thus, ethical egoism allows the agent to create any world if the value of the agent's life is not affected and is thus silent on the choice between worlds A and

²⁸ My conception of deontology is roughly modeled on the version espoused by W. D. Ross in *The Right and the Good* (Oxford: Oxford University Press, 1967), 20-22. Ross envisions a system of prima facie duties, or features of a case which entail moral obligation if no other competing features are present. He describes duties based on promises, reparations, gratitude, justice, beneficence, self-improvement, and non-maleficence.

B if his life is equally good in either. If, however, the agent will live the same life as the other people in the world he creates, then the egoist must choose world B because he will live a +10 life instead of a +1 life. Consequentially, if ethical egoism is the correct moral theory, the repugnant conclusion becomes an almost completely empty value theoretic conclusion which only matters to the extent to which agents care about and are made better off by improving the world.

Ethical egoism is able to deny the ethical implications of the repugnant conclusion because, since the agent is the only person of inherent moral relevance, ethical egoism is a theory with an only real people matter stipulation. No merely possible people matter to egoist moral choices except insofar as their anticipated lives will affect the value of the agent's life. Thus, when the agent considers a world which will have a vast number of people over the millennia to come, he need not consider those future people morally relevant. The world with ten future people who have no effect on the agent's life occasions no differences in ethical duties from the world of a googol future people who have no effect on the agent's life. Other theories with only real people matter constraints (henceforth ORP), however, are also able to deny the ethical implications of the repugnant conclusion. ORP utilitarianism, for example, does not support a duty to procreate when it would be to the disadvantage of existing people to do so because the positive value which the potential child would contribute to the value of the world is not morally relevant until the child is actual. Similarly, an ORP deontology denies the existence of any duties to future people, and thus there is no duty of beneficence to induce the ORP deontologist to give birth to a child, if her duties to existing people such as herself, her spouse, and her family are insufficient to induce her.

ORP utilitarianism and deontology are able to circumvent the requirement to choose world A over world B because when the agent chooses between world A and world B, neither world is yet actual. Therefore, all of the people in both worlds are merely possible people. The only real

people matters variants thus allow both the utilitarian and the deontologist to deny that the lives of any of the possible people in worlds A and B are morally relevant. The only morally relevant people are those, like the agent himself, who are already existing. The ORP utilitarian requires the agent to make the choice which will make him the most well-off. Ex hypothesi, the life of someone in world B is better than the life of someone in world A, so the agent is morally required to choose world B, and thus the utilitarian gets the result that he wants. Similarly, the ORP deontologist only has one duty, his duty to himself, and thus if the choice of world B will make him better off than the choice of world A, he also must choose world B. Notably, both ORP utilitarianism and ORP deontology work exactly like ethical egoism in this situation because only the agent himself matters.

All three ORP theories are able to deny the ethical implications of the repugnant conclusion, but ethical egoism is the most plausible of these three theories because it can get the right results in a couple of cases in which both ORP deontology and ORP utilitarianism give the wrong results. In *Train to Pusan*, James has promised to go to dinner with his family and friends. He will catch the last train from Seoul to meet them at Pusan. He gets to the train station and goes to buy the last ticket, but Emily runs up. Emily has an appointment with a renowned fertility specialist who is visiting Pusan. If she buys the last ticket, she will make it to Pusan in time for her appointment and will conceive Sarah, who will live a life of +100, which is quite good. If she gets the last ticket, however, James will miss the dinner, causing 10 relatives and friends each to have a life 1 worse than the lives that they would have had otherwise. James is more concerned that Emily makes it to her appointment than that he makes it to dinner, and thus his life is 1 better if Emily gets the last ticket. Emily herself will have an equally good life either way because the unique goods and ills of having a child will cancel out. James knows all of this information.

Both ORP utilitarianism and ORP deontology require James to take the last ticket. On ORP utilitarianism, the relevant moral factors are the +1 that James will get by handing over the ticket and the -1 that each of 10 of the other people at the dinner will get if he hands over the ticket. The morally relevant people, then, will lose 9 units of welfare if James hands over the ticket, so ORP utilitarianism requires him to keep the ticket for himself. The morally relevant factors look much the same for the ORP deontologist although there is some added weight which the disvalue to his friends and family acquires because James has special duties to those people and has promised to attend the dinner. The duty of beneficence which James might have to Emily is too weak to override his special duties of non-maleficence to his friends and family because beneficence is often considered to be the weakest duty, because there will be no benefit to Emily beyond the fulfillment of a preference, and because Emily is a stranger who cannot lay claim to any partiality. Thus, the ORP deontologist must also choose to take the last ticket.

Intuitively, however, James ought to give the ticket to Emily. The benefit to Sarah is so considerable relative to the harm to friends and family that James seems to have very good cause to give the ticket to Emily. Ethical egoism, unlike the other ORP theories, is able to agree with the intuition that James ought to give the ticket to Emily. James will get +1 if he gives the ticket to Emily, and thus the case is open and shut for the egoist: James is morally required to give over the ticket. Utilitarianism without the only real people matter stipulation backs up the intuitive judgment. Giving the ticket to Emily will create 91 net value because James gains 1 and Sarah gains 100, while 10 will be lost in the form of disappointed dinner guests. Deontology without the only real people matter stipulation might also return the judgment that James ought to give the ticket to Emily because the duty of beneficence becomes much weightier when the beneficent act will cause a very substantial positive change in someone's life, in this case Sarah's. With that said,

however, the deontologist has more room to deny that it is right to give the ticket to Emily because the duties which James has to his friends and family might still outweigh the stronger duty of beneficence that James has when Sarah's life is taken into account. Furthermore, the deontologist can maintain that the duty of beneficence only operates when there is no duty of non-maleficence, and that here James has a duty of non-maleficence to his family and friends.

A second case, *A Brother and a Friend*, reinforces the advantage which ethical egoism has over the other ORP theories. In *A Brother and a Friend*, Malcolm's mother, Edna, will become pregnant if she stays home tonight, eventually giving birth to Justin, who will have a life of +100. She will stay home tonight if Malcolm, a middle school student, decides not to go the school basketball game. Edna is a teacher, and one of Malcolm's friends, Samuel, is among her students. If Edna becomes pregnant, she will take months of leave towards the end of the school year. The substitute teacher will not be as good, and thus Samuel will not do as well in the class, with the result that his life gets -3. Malcolm, however, will have a +1 life if Edna becomes pregnant due to good times he will have with his brother. The value of Edna's life will not change because the positive value which she will get from having another child will be exactly equal to the disvalue of lost time at her job and the pain of pregnancy. Malcolm knows all of these facts.

Intuitively, Malcolm ought to stay home from the basketball game. While Samuel will be slightly worse off if he stays home, Justin will gain far more than Samuel will lose. Once again, ethical egoism is able to support this intuition because Malcolm will be better off if he stays home from the game, and thus, on ethical egoism, he is morally required to do so. The standard versions of utilitarianism and deontology also offer some support for this intuition. Normal utilitarianism requires Malcolm to stay home because the +100 which Justin gains and the +1 Malcolm gains decisively outweigh the -3 which Samuel receives. On normal deontology, the judgment is the

same because the special duties which Malcolm has towards Justin, his potential brother, are presumably at least as strong as the duties which he has towards his friend Samuel, and thus the morally relevant difference is not in the *nature* of the duties but rather in the amount of value or disvalue which the people to whom he has duties will receive. Once again, the deontologist could argue that the duty of non-maleficence towards Samuel outweighs any duty of beneficence, but when the differences in value are so great, while the partiality which Malcolm owes the two is comparable, that seems wildly implausible. To maintain that the harm accruing to Samuel outweighs the benefit to Justin is close to asserting that preventing harm is lexically prior to providing benefits, which seems clearly wrong in cases in which billions of value of benefit hinge on causing a single unit of harm.²⁹

The other ORP theories, however, once again return the wrong results. The ORP utilitarian is forced to admit that it would be better for Malcolm to attend the basketball game because, among morally relevant people, Malcolm will lose one value by going to the game while Samuel will gain three, for a net increase of two value. Thus, Malcolm is morally required to attend the game on ORP utilitarianism. On ORP deontology, the same result obtains because the special duties which Malcolm has to himself are comparable to the special duties which he has to Samuel in virtue of the fact that Samuel is his friend. The morally relevant difference, then, again comes down to relative values which each of them will receive, and the fact that Samuel will gain more than Malcolm loses determines the case in favor of Malcolm attending the game. If only ORP theories are able to deny the implications that agents must choose world A and that agents have a positive duty to procreate, and ethical egoism is able to return the intuitive result in cases in which the other

²⁹ It seems clearly acceptable to subject one person to a mild headache in order to secure bliss for a thousand

ORP theories come to the wrong conclusion, then there is good reason to think that ethical egoism is the ethical theory best able to solve the repugnant conclusion.

Someone will certainly object to the cases they have offered, pointing out that there are other cases which are not nearly as favorable for ethical egoism. First imagine a case which I will call *Train to Busan Redux*. It is just like Train to Busan except that instead of James getting +1 by giving his ticket to Emily, he gets -1 instead. Ethical egoism now requires James to take the ticket himself instead of giving it to Emily because his life will be slightly better if he takes the ticket himself. Unless that small change has changed the common intuition about the case, it seems that all three ORP theories now come to the wrong conclusion. A similar revision gives us *A Brother and a Friend Redux*, in which nothing changes except that Malcolm's life loses one value when he stays at home instead of going to the game. Once again ethical egoism now joins ORP utilitarianism and deontology in insisting that the morally correct action is for Malcolm to go to the game. These small changes seemingly alter neither the common intuition nor the judgments which normal deontology or utilitarianism come to. Ethical egoism, then, seems to be only marginally better off than either of the other ORP theories, and thus they seem to be almost equally (im)plausible ways to avoid the ethical implications of the repugnant conclusion.

Furthermore, there seem to be numerous cases in which ORP deontology and utilitarianism come closer to the common-sense intuitions than ethical egoism. In any case which does not involve any merely possible people, the ORP theories are identical to their normal counterparts, and thus far more widely accepted than any form of ethical egoism. Imagine a very simple case, *Backyard Baseball*, in which Robert, wants to play baseball in the backyard. If his brother, Oliver, agrees to play baseball, Robert will be hit in the face by the ball, making his life 10 worse than if he stays inside. On the other hand, Oliver will have so much fun prior to the injury that the

experience will be a net +1 for his life even once the worry about his brother's injury is subtracted. Oliver knows all of this information. Ethical egoism maintains that it is morally required for Oliver to agree to play baseball. On the other hand, the other ORP theories both maintain that it is impermissible for Oliver to agree to play baseball. The total value of the choice will be -9 once Robert's -10 is added to Oliver's +1, and thus the utilitarian is clearly against the choice. The deontologist is also against the choice because non-maleficence is a particularly strong duty and the relation of being brothers gives Oliver reason to be at least as partial towards Robert as towards himself. Thus, ethical egoism here seems to be the least plausible of the ORP theories.

Ultimately, I have to bite the bullet in these cases, but I do have a few caveats to offer in defense of my view. First, the common-sense intuition may be skewed by the fact that the average person is not considering the case fairly. It seems likely that when asked to consider one of these cases, people project themselves into the place of the decision maker, Oliver, Malcolm, or James, and then replace the result stipulated for the agent's life with the result which they expect they would experience. Thus, the person evaluating *Backyard Baseball* might be thinking about how worried he would be and how bad he would feel if he hit someone in the face with a baseball, and that might contaminate his intuition, given that it is stipulated that Oliver does not feel bad enough that the incident is a net negative for his life.³⁰ He might even side against the egoist intuition because he thinks that the agent must be evaluating the events from a misguided or malicious viewpoint if the agent somehow manages not to be as bothered by the event as the reader thinks he ought to be. In order to evaluate the case fairly, the evaluator must make sure that he imagines the agent as having received some good larger than the evaluator imagines that he would lose by

³⁰ Having once broken a guy's nose while playing catch, I know exactly how bad that feels, which is really bad.

choosing a seemingly repugnant decision. When the agent properly interprets the case, then I contend that he would be much more likely to endorse the egoist intuition.

Furthermore, ORP utilitarianism seems to completely miss the intuition that underlies normal utilitarianism, namely impartiality. One of the most compelling features of utilitarianism is that what makes the world good is not the fact that certain people have certain amounts of good, but the fact that there is a certain amount of good. Thus, it does not matter who gets a good so long as somebody does. I cannot be partial to my friends and family because the world is not better for them getting a certain good than for anyone else getting it. ORP utilitarianism, however, is partial in favor of presently existing people. On one hand, this does allow the utilitarian to address the complaint that it is usually better to make people happy than to make happy people,³¹ which is compelling so long as creating happy people lowers the average value of the world. Prioritizing existing people, however, leads to odd cases when the action which makes happy people also improves the average happiness of all people. Imagine a world of one person, Dave, who is living a life which will be worth +1. Dave can choose to create another person, Bob, who will live a life worth +99, or he can choose not to, in which case his life will be worth +2 instead. It seems clear that Dave ought to create Bob because both total and average utility are much higher if he does so, with the average going to +50 and the total to +100 if he creates Bob, while the average and total are both +2 if he does not. Thus, ORP utilitarianism seems not to incorporate the basic utilitarian intuitions that more of a good is better, and that we should be impartial in our distribution of goods. There is then little positive reason to support ORP utilitarianism except for the fact that it can deny the ethical implications of the repugnant conclusion.

³¹ Parfit, *Reasons and Persons*, 394

Section III: Ethical Egoism is a Plausible Moral Theory

Ethical egoism, however, is far more plausible than most people believe it to be. I will consider a number of objections to ethical egoism in an effort to show that it is sufficiently viable to offer a useful path to avoiding the ethical implications of the repugnant conclusion. First, in order to evaluate the plausibility of ethical egoism it is necessary to define it, the more so as one major objection to ethical egoism is that it cannot be coherently defined. One intuitive definition is that (i) an agent, A, ought to conclude that he should act to promote A's interest and (ii) A ought to judge that others should also act to promote A's interest.³² G E Moore and Brian Medlin have both objected that this definition leads to contradictions. If B acts in B's own interest and against A's interest, then B judges that he acts as he ought to while according to A he has chosen the wrong course. Medlin claims that ultimately everybody ought to do what is in the interest of each other person because, when A is the judge, A ought to do what is in A's interest by (i), as should B, C, etc. by (ii), but at the same time when B is the judge, B ought to do what is in B's interest by (i), and A, C, etc. also ought to do what is in B's interest by (ii) and so on.³³ Furthermore, Moore maintains that for a thing to be good for someone is for that thing to be good and in someone's possession. Therefore, the state of affairs that constitutes B's interest is the possession of something good simpliciter by B. Moore maintains that it is incoherent for anyone to try to attain good only for himself because it is the same good regardless of who possesses it.³⁴ On the basis of

³² Medlin, Brian, "Ultimate Principles and Ethical Egoism" in *Morality and Rational Self-Interest*, ed. by David P. Gauthier (Englewood Cliffs: Prentice-Hall, 1970), 60-61

³³ Medlin, "Ultimate Principles and Ethical Egoism", 60-61

³⁴ Moore, G.E., "Is Egoism Reasonable?" in *Morality and Rational Self-Interest*, ed. by David P. Gauthier (Englewood Cliffs: Prentice-Hall, 1970), 49-51

these contradictions Moore and Medlin maintain that the ethical egoist is simply advancing an incoherent theory: not merely not talking sense but not talking.³⁵³⁶

These objections may be answered by amending the definition. I follow Jesse Kalin's suggestion that the problem lies in the second component, and that to produce a coherent definition agent A must judge that B acts correctly when B acts in B's interest, not A's.³⁷ Kalin suggests that Medlin rejects this approach because Medlin assumes that for A to believe that an action, x, is morally right, A must approve of x, and that approving of x requires wanting x to occur. On egoism, Agent A, desiring the promotion of his own interests, could not approve of B's action when that action failed to promote his own interest, and thus he could not view that action as morally right. Kalin maintains that an analogy with competitive games suggests that people may believe that an action is right in some sense without approving of it in the sense of desiring it, as when the manager of a football team acknowledges that the other team ought to run down the clock by keeping the ball on the ground even as he hopes that they do not.³⁸ If, then, believing an action to be morally right does not require wanting it to occur, the path is clear for the ethical egoist to modify premise (ii).

Medlin may still argue that morality is a different beast, and that to believe an action is *morally* correct requires approval in a way that believing an action is pragmatically correct does not. It seems, however, that on common sense morality there might be cases in which someone believes an action to be morally correct even as he does not want that action to be performed. Suppose the manager of a store has been erroneously informed that a particular customer is

³⁵ Medlin, "Ultimate Principle and Ethical Egoism, 61-62

³⁶ Moore, "Is Egoism Reasonable", 54-55

³⁷ Kalin, Jesse, "In Defense of Egoism", in *Morality and Rational Self-Interest*, ed. by David P. Gauthier (Englewood Cliffs: Prentice-Hall, 1970), 69-70

³⁸ Kalin, "In Defense of Egoism", 73-74

shoplifting. The manager asks another nearby customer whether he has seen the suspect steal anything. That man could do the morally correct thing and clear the suspect, or he could lie. If he lies, however, the manager will, in a fit of compassion, pardon the suspect. The moral goodness of the act of compassion may be greater than the moral goodness of telling the truth, even when the moral wrongness of the lie is subtracted from it. Overall, then, a better outcome might come about if the customer lies than if he tells the truth, and thus an impartial observer might not want the customer to perform the morally correct action. If the description of the case is correct, then it seems that believing an action to be morally right requires only the sort of approval which one may grant without wanting the act to occur.

I will avoid the inconsistency of evaluation which Moore and Medlin object to by defining ethical egoism as the theory that an agent, A, performs the morally correct action if and only if he performs the available action which maximizes his well-being. Thus, when agent A evaluates his own actions, they will be morally correct if they maximize the value of his own life, but when he evaluates the actions of another agent, B, they will be morally correct if they maximize the value of B's life. A couple features of this definition ought to be called to attention. First, ethical egoism remains coherent regardless of whether the hedonist is right or whether an objective list theory is correct or whether some other description of the value of a life obtains. Second, my ethical egoism is a maximizing theory: an action is only morally correct if it *maximizes* the value of the agent's life. An action which benefits the agent but less than another available action is morally wrong. Suppose that it is the day after Halloween and a child trades his snickers bar for a kit-kat bar when he could have traded it for a Reese's cup which he would have enjoyed more. Provided that pleasure matters to the value of a life, according to my theory, the child has performed a morally

wrong action. Of course, so long as the difference in his enjoyment is miniscule, the action is only very slightly wrong, but wrong nonetheless.

Like all maximizing theories, ethical egoism is vulnerable to the objection that it is overly demanding. However, the problem looks somewhat different for ethical egoism than for a more conventional moral theory like maximizing act utilitarianism (MAU) which requires people to perform actions contrary to their self-interest most of the time. For the ethical egoist there is no conflict between morality and self-interest because acting morally just consists in acting in accord with self-interest. Consequentially, the egoist need not expect people to give moral reasons greater weight than self-interested reasons, and thus dodges much of the problem which MAU is vulnerable to. Instead, the problem seems to be that ethical egoism undermines autonomy. Suppose a man can become a butcher, a baker, or a candlestick maker. He will be happiest if he becomes a baker, but only slightly happier than if he becomes a candlestick maker. On the other hand, he will be seriously unhappy if he becomes a butcher. Ethical egoism is committed to the judgement that the man performs the morally correct action by becoming a baker, and that he acts wrongly if he chooses either of the other two jobs. It seems, then, that facts about the world, notably about the man's dispositions and talents, shoehorn him into a very limited moral path. Common-sense morality, on the other hand, allows the agent a much greater latitude of amoral action in his own life. According to common-sense morality, the man could choose any of the professions without committing a moral error, although prudential reasons would rule out the option of becoming a butcher.

The ethical egoist may respond that his theory only undermines autonomy in the same way in which all consequentialist theories seem to do so. MAU in particular is vulnerable to exactly the same objection and, given that most people are unwilling to discount utilitarianism out of hand,

the objection may not be that serious. Alternatively, the egoist could see the objection as particularly targeting theories which extend the range of moral judgments into the agent's personal life and thus circumscribe morally appropriate actions more strictly. Some consequentialist theories, such as rule utilitarianism, give everyone the same set of duties, and limit these duties to areas of choice which are usually seen as morally relevant. Both ethical egoism and MAU, however, make it a moral requirement to promote the agent's welfare, which both gives them the ability to extend moral judgments into the seemingly non-moral realm of exclusively self-affecting decisions and compels them to do so. Thus, ethical egoism and MAU are particularly vulnerable to the objection that they undermine autonomy insofar as they extend moral judgments throughout the entirety of the agent's life and extend a unique and very narrow set of permissible possibilities for each person. For the ethical egoist, there are only ever multiple permissible actions when there are two or more actions which contribute exactly the same amount of value to the agent's life and there is no other action which contributes more value, which, practically speaking, ordains a single moral path for each agent.

The ethical egoist might reply that agents may exert influence over which course is best for them. If somebody greatly desires to achieve something, accomplishing it will have a greater value than if he is indifferent. Thus, the man choosing a job can exercise his autonomy by choosing to desire to be a candlestick maker or otherwise causing that course of action to yield better results, perhaps by studying the craft in his spare time. Aristotle believes that people may, through repetition and self-conditioning, alter their dispositions and come to see as good actions which they initially were indifferent or opposed to.³⁹ The same tools are open to the egoist who may shape his preferences to any object, thereby making the pursuit of that object into the morally

³⁹ Aristotle, *Nicomachean Ethics*, trans. Martin Oswald (Upper Saddle River: Prentice Hall, 1999), 33-35

correct path for him. Now, the initial expenditure of effort on shaping preferences may run contrary to his interests when he could have spent the time pursuing the objectives which he already had, and may thus be morally wrong, but his subsequent actions taken in accord with these new dispositions will be free from moral blemishes. Shaping morally correct action in such a way is analogous to giving your implicitly trustworthy roommate ten dollars on the condition that if you take out the trash tomorrow he must give it back to you. While giving him the money today is not really in your interest, tomorrow it is certainly in your interest to take out the trash. Thus, the ethical egoist may respond that he exercises autonomy by maintaining the ability to shape his preferences and thus determine what actions will be morally correct tomorrow, although often at the expense of a slight moral wrong today.

A similar response from preference shifting is not nearly as open to the maximizing act utilitarian. The morally correct act for the utilitarian depends on the results for the world as a whole. In most cases, shifting the preferences of the agent do not change which action is best for the world as a whole, because the consequences of an action radiate out, affecting myriad people increasingly removed from the action. The effect of the action on the agent himself is a relatively minor portion of the overall moral evaluation under the vast majority of circumstances. Thus, a charitable agent might give a child a chocolate bar, causing him to go home in a better mood, forestalling a fight with his father, who might then choose not to go to the bar, avoiding a drinking bout which would have angered his wife, and saving some money which he might spend at the store, pleasing the manager, and so on, and so forth. Even if the agent caused himself not to have a preference for giving chocolate bars to children, even if he made himself hate such actions, the total consequences of giving the chocolate bar would still make it the morally required action. Certainly, it is true that there are exceptions, like the hermit who may be able to change his moral

requirements via preference shaping because the consequences of his actions do not fan out in the same way. Preference shifting, however, does not allow the maximizing act utilitarian to avoid undermining autonomy in the vast majority of circumstances, although it would change how good or bad various actions would be for the agent.

Maximizing act utilitarianism seems to undermine autonomy more seriously than ethical egoism, but deontological theories and common-sense morality also restrict the autonomy of the agent. Such theories usually confine their judgments to the sort of actions which we commonly think of as morally relevant, and thus do not intrude so deeply into the life of the agent, but where they do make judgments, those judgments tend to be immutable. Indeed, the stability and impartiality of deontological judgments is one of the main selling points of deontology. If an agent desires to do something forbidden by one of these moral theories, there is no way to influence the morally correct path through preference shifting, and thus there is a sense in which the agent's autonomy is greatly restricted. It may be the case that these theories are silent on the question of whether a man should become a baker or a candlestick maker, but if Kantian deontology is right that lying is forbidden, then it might be morally wrong to become a spy instead.⁴⁰ While, in practice, deontological theories will leave many more morally acceptable courses than ethical egoism, the single course that ethical egoism leaves open tends to be the one which the agent has the greatest desire to take. Thus, there is an interpretation on which the deontological theory is more autonomy undermining than ethical egoism: if I want to become a spy more than anything, the fact that deontology allows a thousand other morally acceptable careers does not prevent my autonomy from being restricted more by that theory than by ethical egoism. Non-agent relative

⁴⁰ Immanuel Kant, *Groundwork for the Metaphysics of Morals*, trans. Arnulf Zweig (Oxford: Oxford University Press, 2002) 230

moral theories, then, do not entirely escape the objection of undermining autonomy, even if it is usually more serious for agent relative theories such as ethical egoism.

Another fundamental objection to ethical egoism is that it fails the test of generalization. Many philosophers have argued that any adequate moral theory must extend similar duties to everyone who finds himself in a similar situation. Kant suggests that considering the consequences of everyone following a particular maxim is an effective test of whether that maxim is morally valid,⁴¹ which suggests that any morally acceptable principle must be applicable in all similar situations. In the same vein, Medlin suggests that individual egoism, the principle that I should look after myself, is not an ethical theory at all because it is not a general principle, and that it does not assert itself as a moral theory unless it manifests itself in the universal form that everyone ought to pursue his own good.⁴² The solution, then, is to put ethical egoism into a universal form, as I have done above.

Section Four: Ethical Egoism and Collective Action Problems

Ethical egoism, however, when stated in a universal form, as in the definition which I have offered, leads to collective action dilemmas. Two such notable dilemmas are the tragedy of the commons and the prisoners' dilemma. The tragedy of the commons describes a large scale collective action dilemma in which each person acting in his own interest leads to a worse result for everyone. Common examples of the tragedy of the commons are situations in which there is a communal resource, such as a lake full of fish or a common pasture, which will be depleted if people do not limit their consumption. Rush hour traffic provides a more modern example. Each person gets to work faster by driving than by taking mass transit, but if everybody takes mass

⁴¹ Kant, *Groundwork for the Metaphysics of Morals*, 222-224

⁴² Medlin, "Ultimate Principles and Ethical Egoism," 59-61

transit, then reduced congestion causes everyone to get to work faster than if everyone drives. One person taking mass transit, however, is insufficient to reduce congestion to the point that it is faster than driving, and thus driving has a substantial marginal advantage to over taking mass transit.⁴³ With no incentive to take mass transit, everyone continues to drive, and to suffer longer commutes as a result. In both incarnations of the tragedy of the commons, individually advantageous behavior leads to universally negative consequences.

The prisoners' dilemma, on the other hand, refers to a particular two-person situation which exemplifies the same general characteristics as the tragedy of the commons. If only one of two criminals testifies against the other, the criminal who testified will only get three months, although the other will receive a full ten years. In the event that both of the criminals testify, then each of them will receive eight years. The criminals, Jones and Smith, have no chance to confer and thus must decide independently. Regardless of what Jones does, Smith ought to testify against his partner. If Jones also testifies against him, then Smith will receive eight years instead of ten, while if Jones does not testify against him, then Smith will escape with a mere three-month sentence. The exact same logic applies to Jones. Thus, each criminal, acting rationally, will testify against the other and both will get eight years, whereas if both had kept silent, they would each have received a mere year. Once again, following the rational course seems to lead to a disastrous result.⁴⁴

Collective action dilemmas have always presented a problem for theories of rationality, but they seem to present a much greater problem for ethical egoism. The existence of collective action dilemmas shows that the actions which self-interest theories of rationality claim are justified lead to disastrous consequences when generalized, but the problem is worse for ethical egoism because

⁴³ Parfit, *Reasons and Persons*, 61-62, 85

⁴⁴ Parfit, *Reasons and Persons*, 56-62

ethical egoism claims that those same actions are not merely rationally justified but *morally* justified. Thus, ethical egoism seems to catch agents in a quandary. Either they can do the morally correct action and drive to work, making the situation marginally worse for everyone, or they can take a morally wrong course by taking the bus. Agents no longer even have the option to weaken the dilemma by weighting their moral reasons more heavily than their self-interested reasons because moral reasoning has been collapsed into self-interested reasoning. It seems, then, that ethical egoism gives rise to the most full-blooded endorsement of the type of self-serving actions which give rise to collective action dilemmas. In Parfit's terminology, ethical egoism has revealed itself as directly collectively self-defeating,⁴⁵ and thus has cast serious doubt on its claim to be an adequate moral theory.

In order to answer the objection, the ethical egoist must show that a good egoist is not never self-denying. To be never self-denying is to always take the action which produces the greatest benefit. In *Reasons and Persons*, Parfit identifies one case in which being never self-denying leads to a very bad outcome. Suppose Bill is driving in the desert without his wallet when his car breaks down. Another motorist, George, comes along and is willing to drive Bill home, but only if Bill agrees to pay him upon arrival. Bill knows that, once he is home, there will be no self-interested reason to pay George because Bill needs no additional services from him, and because Bill is never self-denying, Bill knows that he will then refuse to pay. With no contract and no witnesses, George will not be able to bring suit or otherwise penalize Bill for not paying. Unfortunately for Bill, he cannot lie convincingly, so, knowing that he will refuse to pay, he is unable to convince George that he will pay, with the result that George leaves him in the desert.⁴⁶ Kavka's toxin puzzle seems to be another case in which being never self-denying makes someone worse off. In the puzzle, an

⁴⁵ Parfit, *Reasons and Persons*, 55-56

⁴⁶ Parfit, *Reasons and Persons*, 7

eccentric billionaire offers to pay you a million dollars tomorrow morning for intending tonight to drink a toxin which will cause serious discomfort tomorrow afternoon. One who is never self-denying, however, knows that, once he has the money, there will be no reason for him to drink the toxin tomorrow, and thus he cannot intend to drink it today.⁴⁷

One intuitive way to get the money in the toxin puzzle is to trick yourself about the nature of the case. Perhaps you convince yourself that you will not deserve the money unless you actually drink the toxin so that getting the money and not drinking the poison becomes the second-best option instead of the best, or perhaps you deceive yourself into believing that you actually like drinking the toxin in order to get the money. If you can represent to yourself that the best path is a path in which you drink the toxin, then you can get the money. If it is the case that you cannot honestly intend to drink the toxin while fully understanding the deal, then it becomes the case that deceiving yourself actually is the best path. The mechanism used here is similar to the preference shifting offered as a response to the problem of infringement of autonomy. By shifting what you sincerely believe will get you what you want, new options become open to you in which what you want to do aligns with what will make you well-off.

Whenever a problem seems unavoidable, then the ethical egoist has cause to shift his preferences so that the problem becomes less serious. Suppose that Elizabeth has a fulfilling job in the city which is the best job available to her, but which requires her to commute a long distance each day. All of the commuters have reason to drive to work, because driving is faster than taking the bus, but because everybody is driving, the commute is longer for everybody than if everybody took the bus. It seems that the order of possible states of affairs, from best to worst for Elizabeth, is Elizabeth driving while everyone else takes the bus, everyone including Elizabeth taking the

⁴⁷ Kavka, Gregory S. "The Toxin Puzzle." *Analysis* 43, no. 1 (1983): 33-36. doi:10.2307/3327802.

bus, everyone including Elizabeth driving, and everyone else driving while Elizabeth takes the bus. The only available states of affairs, however, are the third and fourth worst ones because Elizabeth cannot control the actions of the other commuters, and thus Elizabeth is highly displeased with her commute because the better possibilities are not available options. If Elizabeth shifts her preferences so that an available option becomes the best possible option, however, then the gap between the best possible and the best available will no longer exist to trouble her with how much better things could be. Thus, it seems that Elizabeth will be better off if she shifts her preferences so that a longer commute is a good thing. Perhaps she might come to enjoy the time which she has to herself in the car, in which case Elizabeth driving while everybody else drives would be the best option because it would afford her the most time in the car. Alternatively, she might become an avowed environmentalist so that only options in which she rides the bus would be appealing to her, although there would still be a better possible option, namely everybody riding the bus.

While preference shifting seems to provide some sort of a response in the commuting collective action dilemma, it does not seem to be a very convincing response. Preference shifting doesn't truly solve the problem but merely claims that it is not really a problem. In fact, it seems to lead to the conclusion that no intractable problem is a problem because once it is clear that it is not solvable, the agents involved shift their preferences so that whatever was initially taken to be a problem becomes the best outcome. On one hand, there may be a grain of truth in the claim that the best path to happiness is aligning your desires and expectations with what will in fact result, but, on the other hand, the implications seem highly problematic in the case of the communal fishing pond.

In the fishing pond case, if all the villagers have agreed to take only a certain number of fish, there is a difference between the states of affairs which are best in the short-term and those that are best in the long term. Thus, in the short-term, the ranking of possibilities from best to worst in terms of value for an individual, Ronald, is Ronald cheating on the communal agreement while nobody else does, everyone including Ronald cheating, nobody cheating, and everybody but Ronald cheating. In the long-term, however, nobody cheating becomes better than everybody cheating because if everybody cheats, the resource is seriously diminished. In the short-term, there is little incentive to preference shift because the best available option, everybody cheating, is almost as good as only Ronald cheating. When everybody cheats, the decline in available resources is minimal in the short term and thus has only a slight impact on Ronald's short-term interests. In the short-term, Ronald is close to impartial between being the only one to cheat, and everybody cheating. When nobody ever has a short-term interest in preference shifting, however, the long-term result comes about and suddenly there are no fish in the pond. It then seems absurd to argue that Ronald shifting his preferences to hate fishing or hate having the extra resources leads to a satisfactory result. It might even be the case that that was the only food supply and now Ronald will starve. It would be ridiculous to insist that Ronald, having shifted his preferences to hating eating, is now well off on account of his imminent death. It cannot be satisfactory to insist that people may always shift their preferences to reimagine collective action dilemmas as the best possible outcome, although it may be the case that once Ronald is starving to death, it is good for him to shift his preferences to align with the inevitable outcome.

When preference shifting intersects with the problem of seeking happiness, however, a more compelling response emerges. It is widely agreed that nothing is so anathema to happiness as trying to be happy, and that fact in itself has been leveled as an objection against ethical

egoism.⁴⁸ In her paper “Moral Saints,” Susan Wolf argues that the utilitarian moral saint must always consider whether his current action is best promoting total welfare, and that doing so prevents him from ever really enjoying anything. His fixation on welfare value prevents his life from contributing much at all in the way of welfare to the global total.⁴⁹ It seems likely that a similar problem exists for the never self-denying ethical egoist. He must always consider whether his actions are best promoting his welfare, which consideration prevents him from ever living in the moment. In Wolf’s words, he has one thought too many, and this additional thought prevents him from living as good a life as he might be able to do otherwise.⁵⁰ If following ethical egoism leads to a worse life than following some other moral theory, then it seems that ethical egoism may be failing in its own terms.

If ethical egoism tells people to do that which will make their lives go best and trying to have a good life makes one’s life go worse, then ethical egoism tells people not to try to make their lives go best. Instead, once somebody has accepted the truth of ethical egoism, then the morally correct action for him to take is to forget that he is an ethical egoist. So long as it is before his mind that he is an egoist, he will be motivated to always take the action which most promotes his welfare, causing him to be never self-denying, but being never self-denying involves having one thought too many and being unable to secure certain benefits in cases such as the desert breakdown or the toxin puzzle. Instead, the egoist ought to efface egoism from his mind in favor of some other moral theory, the belief in which will allow him to live the best possible life. In practice that theory is probably a widely accepted and not overly demanding theory such as common-sense morality, although likely with some personal modifications. Thus, it might be reasonable for a compulsive

⁴⁸ Parfit, *Reasons and Persons*, 6

⁴⁹ Wolf, Susan. "Moral Saints." *The Journal of Philosophy* 79, no. 8 (1982): 424 doi:10.2307/2026228.

⁵⁰ Wolf, "Moral Saints," 429-430

liar to adopt a moral theory which does not hold that lying is morally wrong so that he does not have to condemn himself morally on a daily basis. While it may seem that ethical egoism has just disappeared at this point, it remains the case that ethical egoism would still be the epistemically justified moral theory, even if some other moral theory were the pragmatically justified theory.

If ethical egoism is a self-effacing theory, then collective action dilemmas are not nearly so great a problem as they initially seemed to be. Once people have effaced the never self-denying interpretation of egoism, they would be inclined to act in a conventionally moral fashion. It seems likely that the moral theory which egoism was effaced in favor of would give some weight to the interests of other people in order to avoid conflicts and their attendant risks, and thus it is likely that egoists would have moral seeming reasons to take community-oriented actions such as abiding by the agreement to only take so many fish from the communal pond. Furthermore, collective action dilemmas were only a particular problem for egoism because the never self-denying interpretation of egoism suggested that the ethical egoist was not merely permitted but morally required to take actions which exacerbate dilemmas. With the never self-denying interpretation having been effaced, ethical egoism no longer provides any additional support for patterns of action which lead to collective action dilemmas and is thus no more guilty of promoting such problems than any other common moral theory.

The self-effacing interpretation of ethical egoism has implications for another, weaker objection to ethical egoism, the objection that so few people historically have professed any sort of ethical egoism. If ethical egoism were a plausible theory, the objection maintains, then far more people would believe it. If ethical egoism is self-effacing, however, that objection seems far less plausible. If anyone who accepts ethical egoism has both a strong moral and a strong rational reason to forget that he accepts ethical egoism, then it should not be at all surprising that there are

no self-professed ethical egoists on the streets. Furthermore, even if ethical egoism were not self-effacing, it would not be that surprising to find few people claiming to be egoists because of the opprobrium with which such an announcement is usually greeted and the corresponding difficulty in securing future cooperation from other people. It would also often be the case that an egoist parent would be better off teaching his children some other moral theory lest they find themselves morally barred from supporting him in his old age. In fact, he would probably be morally required to inculcate in them such morals as would make them most inclined to treat him with respect and reverence. Thus, even if many people acknowledge ethical egoism as true at some point, regardless of whether or not it is self-effacing they would be disinclined to promulgate their doctrine because it might redound to their disadvantage.

Of course, the argument that there are few apparent egoists because ethical egoism may not be circulated abroad leaves ethical egoism vulnerable to yet another objection, that a true moral theory must be the kind of thing which it is good to make known. Here conceiving of egoism as a self-effacing theory is essential to the response. If egoism is a self-effacing theory which leads people to something approaching common-sense morality, then it is not disadvantageous to propagate it. In fact, it might turn out to be a pleasure of sorts to explain to others that the implications of ethical egoism are not the oft maligned ones that they are generally conceived to be but are in fact largely consistent with standard conceptions of morality. Now, somebody could only spread ethical egoism before it was completely effaced from his mind, but, in that interim period, spreading ethical egoism would not have any particularly deleterious effect. Of course, it may also prove morally acceptable just to spread the theory in favor of which one is effacing or has effaced ethical egoism, but the force of the objection that ethical egoism is unsuitable for propagation consisted in the fact that it was reprehensible and detrimental to the agent to do so,

not in the fact that one might prefer to spread something else. Rule utilitarians might also find it more efficacious just to spread the rules that their theory approves and not the principle underlying it. Once the ethical egoist may claim that there are some situations in which it is permitted to spread ethical egoism without reservation, he has successfully avoided the objection.

Section 5: Egoism and the One-Person Repugnant Conclusion

One consequence of the egoist response to the repugnant conclusion is a seeming acceptance of the one-person repugnant conclusion (OPRC). The OPRC moves the focus from the evaluation of worlds to the evaluation of lives, exchanging lives barely worth living for moments barely worth living. Simply put, the one-person variant is the conclusion that for any life of several dozen years of high quality life, there is another, better, life of far more years that are barely worth living.⁵¹ Each day will be barely worth living, or even a complete wash, but over the course of each year, there will be just enough days that are worth living, albeit barely, that the year as a whole represents a positive contribution to a life. On Parfit's formulation of the life barely worth living, instead of living a normal human life, the OPRC maintains that it is preferable to live for many hundreds or thousands of years during which you enjoy nothing but Muzak and potatoes.⁵² Many people view the one-person conclusion with as much repugnance as the general repugnant conclusion, and thus the egoist response must provide some answer to avoid the general disapprobation with which people regard theories which accept the one life repugnant conclusion.

Egoism seems to embrace the one-person repugnant conclusion and render it of practical relevance because it instructs people to act so that their lives are as valuable as possible. The common assumption is that egoism is totalist, that is to say that it evaluates a life based on the sum

⁵¹ Parfit, "Can we Avoid the Repugnant Conclusion," *Theoria* vol. 82, Issue 2, May 2016, 119

⁵² Parfit, "Can we Avoid the Repugnant Conclusion," *Theoria*, 118

of atomic units of good and bad experienced throughout the life. The life of very many years barely worth living is a life of very many atomic units of good, and thus egoism seems to adjudge that life as very good. It is not clear, however, that egoism is committed to a totalist account of the good life. The egoist could instead back an averagist account of the good life, in which the best life was the life with the highest average good per unit of time or some sort of hybrid model in which both the average and total welfare affect the ultimate evaluation of the life. Nevertheless, while these approaches have the potential to avoid the individual repugnant conclusion, they lead to other deeply counterintuitive conclusions, including that adding blissful lives to a world might make that world worse if the current average exceeded the value of a merely blissful life.⁵³ These approaches were also available as tools to undermine the general repugnant conclusion, but their innate implausibility renders them inadequate for either task.

Another approach for the egoist is to deny totalism by promoting a theory of organic unities. Instead of the value of a life being entirely dependent on the atomic units of value within that life, it might be the case that lives can only be evaluated as wholes.⁵⁴ Just as the beauty of a garden can exceed the sum of the beauty of the individual flowers, the value of a life might depend on context dependent factors such as the ratio of the various constituent elements of positive value or the exemplification of flourishing characterized by the receipt of diverse goods. Thus, a life containing a vast amount of the pleasure of eating potatoes might be worse than a life containing far less total pleasure, but which contains an admixture of pleasures from knowledge, friendship, love, and sensory perceptions because the diversity of pleasures multiplies the sum by a considerable factor. This move seems initially plausible because most people have the intuition that a life of a thousand years of eating potatoes with a total pleasure of perhaps (+1000) is worse

⁵³ Fred Feldman, "Justice, Desert, and the Repugnant Conclusion," 192-193

⁵⁴ Noah Lemos, *Intrinsic Value: Concept and Warrant*, (Cambridge: Cambridge University Press, 1994), 36

than a life of a thousand years in which there are a diversity of normal human pleasures but in which the sum of the pleasures is a mere (+800). It does not seem that repugnant to call an extremely long life containing a typical blend of human activity, but with exceedingly muted pleasures, quite good.

In order for a response to the OPRC to be adequate, it must be able to resist the arguments which justify the general repugnant conclusion. Parfit first supports the general repugnant conclusion with the impersonal total principle, which states that, *ceteris paribus*, the best world is the one which contains the most of whatever makes life worth living.⁵⁵ He later reinforces the conclusion with the benign addition argument, which maintains that adding new lives worth living does not reduce the value of a world, so that a world A+ which contains 10 people with lives of (+10) and 10 people with lives of (+1) is not worse than world A which only contains the 10 people with lives of (+10). He then proceeds to argue that leveling out the two groups of people in world A+ in such a manner that total utility also increases slightly yields a world, B, with 20 people living lives worth (+6) which is better than A+ because it is more equal, has greater utility, and makes the worst-off people better off. Repeated applications of adding more people and leveling out the world eventually leads to the repugnant world Z, with a vast number of people living lives barely worth living.⁵⁶

The organic unities approach is able to address both arguments for the general repugnant conclusion. The argument from the impersonal total principle is rejected by the theory of organic unities because a life could contain more atomic units of whatever makes a life worth living while yet scoring worse if a multiplier like the suggested diversity multiplier is in place. The impersonal total principle might still hold if the *ceteris paribus* clause is held to mean that the two lives are

⁵⁵ Parfit, *Reasons and Persons*, 387

⁵⁶ Parfit, *Reasons and Persons*, 419-420

identical in respect to whatever else is relevant to the evaluation of lives, so that, if diversity of pleasure is important, life A would be better than life B if it contained at least as much of each kind of good present in world B, but that is not a radical claim and leaves open the possibility that a normal human life could be preferable to a very long life which only contained the repeated, muted pleasures of eating potatoes.⁵⁷

The organic unities approach also decisively undermines the benign addition approach. The benign addition argument starts with the move from A to A+, which, in the context of the OPRC, involves the claim that taking a normal human life of several dozen years which are quite worth living, life A, and adding years of life barely worth living yields a life, A+, which is not worse than life A. This move, at least, seems justified. Once one has lived a normal human life, getting bonus time at any quality of life which is worth living seems more like an improvement than a detriment. The repugnant conclusion results if B is either at least as good as or better than A+ because people would then be required to choose world B over world A given a binary choice. The next move, however, to life B, does not seem nearly as plausible. The move to life B requires leveling out life A+ so that the years of the normal human life, A, are no longer present in their original form. In the context of the general repugnant conclusion, the value of equality and the plausibility of a maximin principle drive this move.⁵⁸ In the context of single lives, however, neither equality nor maximin makes much sense. While there seems to be something unfair in a world where some people live much better lives than others, there does not seem to be any corresponding unfairness in a life in which some moments are much better than others. Suppose

⁵⁷ I assume that increasing the amount of pleasure from one element without changing the amount from other elements will never reduce the overall total in order to avoid objections that a world could be made better by reducing the amount of certain pleasures without any corresponding increases in other areas.

⁵⁸ Parfit describes maximin as the principal that "The best outcome is the one in which the worse-off people are best off." (*Reasons and Persons*, 422) He then applies it to the move from A+ to B on page 426

that you are offered a job for the next year. You will, however, only be paid for one week's work, but the pay during that week is quite high and so you agree to take the job. The fact that you are only paid for one week is not intrinsically bad and you do not seem to have any grounds for grievance. Now imagine that someone hires 52 workers but pays one of them for the work of all of them. Clearly this does not seem to be fair, and the other 51 workers do have a legitimate cause for complaint. Equality, then, seems to matter between persons in a way that it does not matter between times. Furthermore, if equality between the moments of a life is not intrinsically valuable, there is no reason to believe that any sort of a maximin principle applies to lives. Improving the worst moments of a life at the cost of the best moments only seems justified in the context of a life when it improves the total value of a life, unlike in the context of worlds in which it may be permissible to lower the welfare of the best-off more than the welfare of the worst-off is raised.

In the original benign addition argument, the leveling-off move from world A+ to world B did in fact increase the total utility. As Larry Temkin notes, the move from A+ to B is supported by appeals to equality, maximin, and total utility, and only opposed by an appeal to perfectionism.⁵⁹ In the context of a single life, however, maximin and equality have ceased to provide any justificatory value, and thus total utility alone is pitted against perfectionism which, according to Temkin, is the belief that "A is better than B if some of A's members are better off, or live fuller, richer lives, than the members of B."⁶⁰⁶¹ Here the notion of organic unities returns to do valuable work. It seems plausible that some features of the normal human life are uniquely valuable in a way in which the moments of the leveled down life are not. So long as it preserved

⁵⁹ Larry Temkin, "Intransitivity and the Mere Addition Paradox," *Philosophy and Public Affairs* vol. 16(2) 1st April 1987, 155

⁶⁰ Temkin, "Intransitivity and the Mere Addition Paradox," 155

⁶¹ Not to be confused with Perfectionism as referring to the quality of the constituent moments of value in a life expressed by Parfit in "Can We Avoid the Repugnant Conclusion," 117. Parfit calls Temkin's concept Elitism. (*Reasons and Persons*, 427)

the entire life A intact, it seemed obvious that adding more time to A+ did not create a life at all worse than life A, but once leveling out begins, some essential features of a life might be imperiled.⁶² The leveled-out life is missing some of the best things in life. It might lack diversity of goods, or intensity of goods, or some other value which multiplies the atomic units of good in a life.⁶³ Thus, organic unities prevent the benign addition argument from proving decisive in lives because they make the move from A+ to B unconvincing on many reasonable descriptions of the factors which make a life valuable.

If lives are organic unities whose values are not the mere sum of atomic units of value, it seems that worlds might be the same way. If the value of worlds is not a mere sum of the value of individual lives, then the repugnant conclusion might fail in the context of worlds in addition to in the context of lives. There are, however, reasons for thinking that the value of worlds might be the sum of the values of individual lives even if the value of a life is not the sum of the individual moments contained within it. If both lives and worlds are the sum of the value of the individual lives or moments, respectively, contained within them, double counting might result. In the case in which a diversity of goods multiplies the base value of those goods, a life containing a diversity of goods is better than a life containing an equal amount of base goods but less diversity. Suppose that the life with a diversity of goods is twice as good as a mono good life. It seems plausible that the base good of worlds is determined by the values of the lives contained within them. Now

⁶² One objection is that the shape of a life might contribute to its value. A long period of life barely worth living might then lower the value of a life by changing its progression for the worse. Nevertheless, the move from A to A+ holds in cases in which the low value part of the life is put at the beginning or in the middle. Furthermore, this kind of assertion only strengthens the claim that the organic unity approach can undermine the OPRC, and thus does not constitute an objection. Feldman considers whether the shape of a life impacts its value in *Pleasure and the Good Life*, 125-141, and concludes that it does not do so intrinsically, although the shape of a life probably produces pleasures and pains which do have intrinsic value (Thank you to Professor Swenson for bringing this to my attention.)

⁶³ A B life might contain both great goods and substantial evils so that the net value at any moment would be close to zero. This version of the B life still contains the best things in life despite not containing the A life in its original form. Nevertheless, this life scores lower on perfectionism because its best moments are not as good as those in A or A+, and thus perfectionism may be sufficient to justify the preferability of lives A and A+.

imagine a world, Plurus, containing 10 diverse lives with 10 base good, 1 base good from each of ten different sources. Plurus has a base value of 200 because the value of each life once the multiplier has been imposed is 20. Now imagine another world, Solus, which contains 10 lives each with 10 base good solely from eating potatoes. The base good of Solus is only 100 because there is no multiplier for the 10 base value of each life. If Plurus and Solus are organic unities which obey the same multiplier rule as lives, the base value of Plurus would be multiplied by two again to 400 while that of Solus would be an unaltered 100. While it seemed plausible, ex hypothesi, that a diverse life was twice as good as a one pleasure life, once that increased value is compounded at the world level, it leads to the implausible conclusion that Plurus is four times better than Solus. Double counting has unreasonably inflated a real but exaggerated advantage.

We might try to solve the problem by making the base value of a world equal to the sum of the base values of its lives, instead of the sum of the total values of its lives. Nevertheless, such an approach leads to odd conclusions when individual lives are homogenous, but each life contains pleasures different from those of the other lives. Imagine two worlds: one of them is Plurus from the previous example, while the other, Solus2 contains ten lives, each of which has only one type of pleasure, but none of which contain the same pleasure as the other lives. One life contains the pleasure of eating potatoes, while another contains the pleasure of lasting friendships, and a third contains the joy of knowledge. Intuitively, Plurus is better than Solus2 because each of the lives in Plurus is well-rounded while each of the lives in Solus2 is seriously limited in the scope of its goods. Plurus and Solus2 have the same base value because they each contains 10 lives with 10 base value, and they both have the same level of diversity because they each have 10 units of 10 different goods, thus both Plurus and Solus2 have the same multiplier value and are purportedly

equally good. Once again, applying the particular conception of an organic unity based on diversity of pleasures to worlds has yielded unintuitive outcomes.

Furthermore, even if worlds are organic unities, it is not clear that that enables them to avoid the repugnant conclusion. Once again, we must test the arguments for the general repugnant conclusion in the context of organic unities. Like in the context of lives, organic unities cause the impersonal total principle to become almost meaningless because it only operates when all other relevant considerations are equal, and the *ceteris paribus* clause is forced to do so much work that the claim becomes very weak. The worlds are only equal in every other relevant factor when one world is equal or greater in every category of what makes a life good. Therefore, it would no longer be sufficient for a world A to have more of what makes a life good than world B, but it would also have to have at least as much pleasure from knowledge and pleasure from beauty etc. Otherwise, there could be a world which had a higher total of things which make a life good than a world A, but which was worse because it had only one category of pleasure while the lower base amount of pleasure in world A was multiplied by a greater factor because of its equal distribution of pleasures across several categories. World A might have +1000 pleasure from potatoes but might be worse than world B because world B has +300 pleasure each from friendship, knowledge, and potatoes, the diversity of which multiplies the total by a factor of 2. World A would only be guaranteed to be better than World B if it had +300 pleasure from both friendship and knowledge in addition to the +1000 pleasure from potatoes. Technically, the weakened impersonal total principle still supports a minimal version of the repugnant conclusion under which World C with +1001 pleasure from potatoes is better than world A with only +1000 from potatoes, but this weaker conclusion is not nearly as repugnant as the former conception.

With the impersonal total principle enfeebled, the benign addition argument emerges as the main support for the robust repugnant conclusion. The benign addition argument is predicated on the claim that adding additional lives worth living to a world does not make it worse, even if it does not necessarily make it any better. Even if worlds are organic unities, it seems plausible that this claim remains true. If adding more of a good without any corresponding bad could make a world worse, it would follow that removing some goods could make the world better. For those who subscribe to Epicurus' view that death is not intrinsically bad, that might result in the unfortunate implication that there are times when the world is made better by killing people with lives worth living.⁶⁴ The resulting problem would closely resemble the key objection to averagism, notably that it implies that killing all but the happiest people would make the world better. At the least, it would imply that it would have been better if those people had never lived. Thus, at the world level, as at the level of lives, the benign addition argument is able to justify the move from A to A+.

At the world level, however, the benign addition argument is still able to justify the move from A+ to B. In the context of lives, equality between the moments of a life did not seem to be the kind of thing which made a life better. In fact, if that equality meant that some of the best things in life were lost, it might even make a life worse. Now that we have returned to worlds, however, equality once again seems capable of justifying the move from A+ to B. The value of equality is entirely consistent with worlds as organic unities. In fact, it seems implausible that there are atomic units of equality that could even be totaled because equality is a relational property between different entities, as opposed to a property of a single entity like being pleased or performing a

⁶⁴ Konstan, David, "Epicurus", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2018/entries/epicurus/>>. (Section 4: Psychology and Ethics)

virtuous action.⁶⁵ Thus, envisioning worlds as organic unities does not undermine the general repugnant conclusion despite the fact that evaluating lives as organic unities offers a compelling solution to the OPRC.

In this thesis I have considered the problem posed by the repugnant conclusion. I have suggested that the ethical implications of the repugnant conclusion are the source of the repugnance which philosophers such as Parfit have felt when confronted with it. I have argued that ethical egoism allows us to circumvent these ethical implications, and that it is sufficiently viable to make it a promising avenue for avoiding a thorny problem.

Bibliography

- Aristotle, *Nicomachean Ethics*, trans. Martin Oswald (Upper Saddle River: Prentice Hall, 1999)
- Arrhenius, Gustaf “Critical-Level Theories” in *Future Generation: A Challenge for Moral Theory*, (uu-diva-portal.org, 2000),
https://www.academia.edu/3488242/Future_generations_A_challenge_for_moral_theory, 72-86
- Arrhenius, Gustaf, Ryberg, Jesper and Tännsjö, Torbjörn, "The Repugnant Conclusion", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2017/entries/repugnant-conclusion/>>.
- Benatar, David, *Better Never to Have Been* (Oxford: Oxford University Press, 2006)
- Boonin, David, “How to Solve the Non-Identity Problem,” *Public Affairs Quarterly*, Volume 22, Number 2, April 2008
- Cowen, Tyler “Resolving the Repugnant Conclusion” in *The Repugnant Conclusion: Essays on Population Ethics*, ed. by Torbjorn Tannsjo and Jesper Ryberg (Dordrecht: Kluwer Academic Publishers, 2004)
- Feldman, Fred, “Justice, Desert, and the Repugnant Conclusion” *Utilitas* vol. 7, Issue 2, (November 1995) 189-206.

⁶⁵ Temkin, “Intransitivity and the Mere Addition Paradox,” 148

- Feldman, Fred, *Pleasure and the Good Life*, (Oxford: Oxford University Press, 2004)
- Harman, Elizabeth, "Can we Harm and Benefit in Creating?" *Philosophical Perspectives* vol. 18 (2004)
- Heathwood, Chris "The Reduction of Sensory Pleasure to Desire," *Philosophical Studies* 133 (2007)
- Huemer, Michael, "In Defence of Repugnance," *Mind* vol. 117, (October 2008)
- Huemer, Michael, "Against Equality and Priority," *Utilitas* vol. 24, Issue 04, (December 2012)
- Lemos, Noah, *Intrinsic Value: Concept and Warrant*, (Cambridge: Cambridge University Press, 1994)
- Kalin, Jesse, "In Defense of Egoism", in *Morality and Rational Self-Interest*, ed. by David P. Gauthier (Englewood Cliffs: Prentice-Hall, 1970)
- Kant, Immanuel, *Groundwork for the Metaphysics of Morals*, trans. Arnulf Zweig (Oxford: Oxford University Press, 2002)
- Kavka, Gregory S. "The Toxin Puzzle." *Analysis* 43, no. 1 (1983): 33-36.
doi:10.2307/3327802.
- Konstan, David, "Epicurus", The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (ed.), (Section 4: Psychology and Ethics) URL = <<https://plato.stanford.edu/archives/sum2018/entries/epicurus/>>.
- Medlin, Brian, "Ultimate Principles and Ethical Egoism" in *Morality and Rational Self Interest*, ed. by David P. Gauthier (Englewood Cliffs: Prentice-Hall, 1970)
- Moore, G.E., Chapter VI: The Ideal in *Principia Ethica*, (Buffalo: Prometheus Books, 1988), accessed online at <http://fair-use.org/g-e-moore/principia-ethica/>
- Moore, G.E., "Is Egoism Reasonable?" in *Morality and Rational Self-Interest*, ed. by David P. Gauthier (Englewood Cliffs: Prentice-Hall, 1970)
- Mulgan, Tim, "Two Parfit Puzzles" in *The Repugnant Conclusion: Essays on Population Ethics*, ed. by Torbjorn Tannsjo and Jesper Ryberg (Dordrecht: Kluwer Academic Publishers, 2004)

- Parfit, Derek, "Future People, the Non-Identity Problem, and Person-affecting Principles," *Philosophy & Public Affairs*, March 2017, Vol.45(2), pp.118-157
- Parfit, Derek, "Can we avoid the Repugnant Conclusion?" *Theoria*, May 2016, Vol.82(2), pp. 110-127
- Parfit, Derek, *Reasons and Persons* (Oxford: Clarendon Press, 1984)
- Derek Parfit, "Overpopulation and the Quality of Life" in *The Repugnant Conclusion: Essays on Population Ethics*, ed. by Torbjorn Tannsjo and Jesper Ryberg (Dordrecht: Kluwer Academic Publishers, 2004)
- Petersen, Thomas Soberk, "On the Repugnance of the Repugnant Conclusion," *Theoria*, Vol.72(2) (June 2006) pp.126-137
- Raikka, Juha, "The Repugnant Conclusion and the Welfare of Actual People" *Theoria*, Vol.68(2) (August 2002) pp.162-169
- Ross, W. D. *The Right and the Good* (Oxford, Oxford University Press, 1967)
- Tannsjo, Torbjorn, *Understanding Ethics* (Edinburgh: Edinburgh University Press, 2013)
- Tannsjo, Torbjorn "Why We Ought to Accept the Repugnant Conclusion" *Utilitas* vol. 14(2) (November, 2002)
- Temkin, Larry, "Intransitivity and the Mere Addition Paradox," *Philosophy and Public Affairs* vol. 16(2) 1st April 1987
- Wolf, Clark, "O Repugnance, Where is Thy Sting?" in *The Repugnant Conclusion: Essays on Population Ethics*, ed. by Torbjorn Tannsjo and Jesper Ryberg (Dordrecht: Kluwer Academic Publishers, 2004)
- Wolf, Susan. "Moral Saints." *The Journal of Philosophy* 79, no. 8 (1982): 424
doi:10.2307/2026228.