NIKHIL VENKATESH

# Repugnance and Perfection

## I. INTRODUCING THE REPUGNANT CONCLUSION

The late Derek Parfit was crucial to the establishment of the field now called population ethics. A foundational problem in population ethics is his "repugnant conclusion." He introduced it in *Reasons and Persons*[1], in a formulation I will call $RC_0$:

> $RC_0$: For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living.[2]

Though $RC_0$ is the conclusion of several arguments that Parfit gives in his book, he immediately remarks that he finds it "very hard to accept."[3] Most people share his reaction. However, philosophers have so far failed to

This article grew out of a BA thesis written several years ago under the supervision of Teruji Thomas. Many thanks to him for his thorough advice on both that thesis and this article. My thanks also to Simon Beard for comments and discussion on that thesis, and to Joe Horton, James Wilson, and two anonymous reviewers for *Philosophy & Public Affairs* for comments and discussion of this article. I am also grateful to receptive and probing audiences at the UCL Graduate Conference and the London Graduate Moral and Political Philosophy Workshop. This article is dedicated to the memories of Derek Parfit and Josh Parsons, both of whom inspired me to study philosophy.

1. There are anticipations of the repugnant conclusion in Henry Sidgwick, *Methods of Ethics* (London: Macmillan, 1962 [1874]); John McTaggart Ellis McTaggart, *The Nature of Existence* (Cambridge University Press, 1900); and John Rawls, *A Theory of Justice* (Oxford University Press, 1999 [1971]). However, it is Parfit's name and formulation that have been followed by subsequent writers, and it was following Parfit's book that work in population ethics began in earnest.

2. Derek Parfit, *Reasons and Persons* (Oxford University Press, 1984), p. 388.

3. Ibid.

agree on a population axiology that avoids $RC_0$ without being seriously defective in other ways. Some even deny that such a theory is possible[4].

In this article, I will focus discussion on $RC_1$, my translation of $RC_0$.

> $RC_1$: For any possible population of at least ten billion lives of very high positive welfare, there is some larger possible population of lives of very low positive welfare whose existence would be better, if other things are equal.

The differences between $RC_0$ and $RC_1$ are presentational, not substantive. Talking of "lives of very high positive welfare" and "lives of very low positive welfare" rather than people with "a very high quality of life" and those with "lives barely worth living" makes clear that the claim concerns two distributions of the same thing (welfare).[5] This is no departure from Parfit's view: on the page where he introduces $RC_0$ he discusses differences in "the sum of happiness," and in "whatever makes life worth living." Gustaf Arrhenius[6] puts his influential formulation of the repugnant conclusion in terms of very high and very low positive welfare. I replace "people" with "lives" because there is no reason that the repugnant conclusion should apply only to humans, as "people" might suggest. I do not stipulate that the population should be "imaginable" but rather "possible." It would be strange for the limits of human imagination to decide claims in population ethics. If large populations are not imaginable, then $RC_0$ is false, strictly speaking, whatever is true about which populations are better than others. $RC_1$ does not share this implication. Parfit was not wedded to the wording of $RC_0$, using slightly different formulations of the repugnant conclusion in his last two published papers on the subject.[7]

4. Michael Huemer, "In Defence of Repugnance," *Mind* 117 (2008): 899; Yew-Kwang Ng, "What Should We Do About Future Generations?," *Economics and Philosophy* 5 (1989): 240–42; Gustaf Arrhenius, "Future Generations: A Challenge for Moral Theory" (F. D. Dissertation, University Printers, 2000).

5. "Welfare" here is not tightly defined. One may think of it, very roughly, as the property or properties of lives that make them good for those who live them. That a life has positive welfare means that it instantiates these good-making properties to a greater degree than it instantiates the properties that make lives bad.

6. Gustaf Arrhenius, "An Impossibility Theorem for Welfarist Axiologies," *Economics and Philosophy* 16 (2000): 248.

7. Derek Parfit, "Can We Avoid the Repugnant Conclusion?," *Theoria* 82 (2016): 110–27; Derek Parfit, "Future People, the Non-Identity Problem, and Person-Affecting Principles," *Philosophy and Public Affairs* 45 (2017): 118–57.

In this article, I argue that by carefully considering the nature and variety of possible lives of very low positive welfare, and distinguishing $RC_1$ from a related but stronger claim I call $RC_2$, we can show that $RC_1$ is more acceptable than it first appears. It has been claimed that perfectionism—that is, a special concern for the best things in life—is a reason to reject the repugnant conclusion. I argue that although perfectionism gives us a reason to reject $RC_2$, it does not give us any reason to reject, and might in fact support, $RC_1$.

In his last two papers on the subject, Parfit develops a strategy for avoiding the repugnant conclusion that appeals in part to perfectionism. In the final part of this article, I argue that Parfit's strategy can help us avoid $RC_2$ but not $RC_1$. If I am right that $RC_1$ is more acceptable than $RC_2$, this may not be an unwelcome result.

## II. THE UP DOWN ARGUMENT

If $RC_1$ is, as Parfit says, "very hard to accept," why is it problematic? Why do population ethicists not simply reject $RC_1$ and theories that imply it? The reason is that there are arguments for $RC_1$ that are based on very attractive principles. The most important of these is what Parfit called "the Up Down Argument."[8]

The Up Down Argument has three intuitive principles as premises. These are (for any three populations, $p_1$, $p_2$, and $p_3$):

(1) Benign addition: If $p_1$ and $p_2$ are so related that $p_1$ would be the result of increasing the welfare of every life in $p_2$ by some amount and adding some new lives of positive welfare, and holding all other things equal, then the existence of $p_1$ would be better than that of $p_2$.

(2) Non-anti-egalitarianism: If $p_1$ and $p_2$ are so related that $p_1$ has a higher average welfare, a higher total welfare, and a more equal distribution of welfare than $p_2$, and all other things are equal between $p_1$ and $p_2$, then the existence of $p_1$ would be better than that of $p_2$.

8. Parfit, "Can We Avoid the Repugnant Conclusion?," pp. 121–23. See also Huemer, "In Defence of Repugnance," pt. 1.3; Ng, "What Should We Do About Future Generations?," pp. 240–42.

(3) Transitivity: If the existence of $p_1$ is better than that of $p_2$ and the existence of $p_2$ is better than that of $p_3$ then the existence of $p_1$ would be better than that of $p_3$.

The argument runs thus. Call lives of very high welfare A-lives and lives of very low positive welfare Z-lives. Call any population of at least ten billion A-lives an A-population, and any population consisting solely in Z-lives a Z-population. $RC_1$ claims that for any A-population, there is a possible, larger Z-population, in which other things are equal, whose existence would be better. Take any A-population $p_A$. A possible population $p_+$ would be the result of increasing the welfare of everyone in $p_A$ and adding some number of Z-lives. Benign addition says that $p_+$ would be better than $p_A$. There is also some possible population $p_z$ which would be like $p_+$ except that the welfare of every life in $p_z$ would be slightly higher than the average welfare of a life in $p_+$. $p_z$ would be more equal, have higher total welfare and higher average welfare than $p_+$, and would therefore—by non-anti-egalitarianism—be better than $p_+$. Since $p_z$'s existence would be better than that of $p_+$ and $p_+$'s would be better than that of $p_A$, transitivity tells us that the existence of $p_z$ would be better than that of $p_A$.

This is $RC_1$, as $p_z$ could be a Z-population. Were the number of Z-lives in $p_+$ very great, their welfare very low, and the increase in welfare for the A-lives very small, the mean welfare of a life in $p_+$, and therefore the welfare of all lives in $p_z$, could be said to be "very low."[9]

9. Teruji Thomas, "Some Possibilities in Population Axiology," *Mind* 127 (2018): 822–24, argues that the step from $p_+$ to $p_z$, where $p_z$ is a Z-population, can be resisted without giving up non-anti-egalitarianism. According to the axiology he calls "total lexic utilitarianism," one population is better than another if and only if it has greater total welfare—so non-anti-egalitarianism is respected. The welfare of lives in this axiology is represented by ordered pairs of integers $(x, y)$. The total welfare of one population is greater than that of another if and only if the sum of $x$ across all the lives in the former is greater, or the sum of $x$ in both populations is equal and the sum of $y$ across all the lives in the former is greater. For Thomas, Z-lives are lives where $x = 0$, while in A-lives $x > 0$. In $p_+$, there are A-lives, and therefore $\Sigma x > 0$; in $p_z$, $\Sigma x = 0$. Therefore $p_+$ has greater total welfare and is better than $p_z$—according to an axiology that respects non-anti-egalitarianism.

The plausibility of total lexic utilitarianism depends, as Thomas acknowledges (pp. 826–27), on there being some discontinuity in the spectrum of lives between A-lives and Z-lives. He himself doubts this. Certain considerations below, particularly the description of what I call "chronically irritated lives," seem to raise further doubts. For the remainder of this article, I will set aside total lexic utilitarianism.

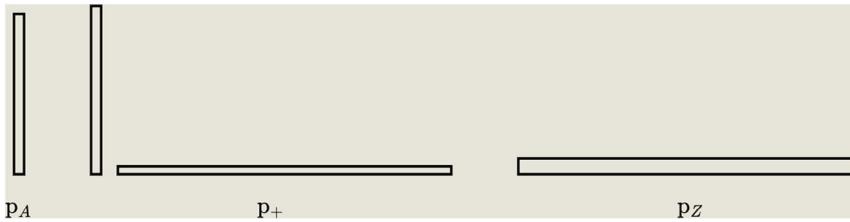


**Figure 1** The Up Down Argument

This diagram represents the three populations, with the width of a box indicating the number of lives in the population and the height indicating the welfare level per life. The Up Down Argument implies that each population is better than those to its left (Figure 1).

Since the three principles that serve as premises in the Up Down Argument have intuitive appeal, and $RC_1$ is counterintuitive, we have a paradox (or at least a puzzle). We must reject at least one intuitively plausible claim—transitivity, non-anti-egalitarianism, benign addition, or the falsehood of $RC_1$.

## III. WHAT ARE Z-LIVES LIKE?

The result of the Up Down Argument might lead us to reassess our original intuition about $RC_1$. Is it really repugnant? An important factor in this assessment is what Z-lives would be like. The more attractive Z-lives are, the less difficult it is to accept that for any A-population, there could be a better Z-population, other things being equal—that is, the less difficult it is to accept $RC_1$. Describing what lives of very low positive welfare would be like is therefore crucial to assessing $RC_1$. Here are some ways such lives could be.

### *III A. Drab Lives*

These lives have no moments of great pleasure, or satisfaction, or honor. They would also have no moments of great pain, or sadness, or dishonor. They would constantly have a barely positive balance of the things that make life worth living over those that make it worth ending. The things that make such lives worth living would be very basic sources of welfare—in Parfit's description "muzak and potatoes"[10] would be the only good things in such lives.

10. Parfit, "Can We Avoid the Repugnant Conclusion?," p. 118. We might think that Parfit wasn't cooking his potatoes correctly!

For those of us who have experienced pleasures greater than muzak and potatoes, drab lives seem like lives with the good things ripped out. But they are also lives with the bad things ripped out. In one respect, they would be better than any life that has hitherto been lived: they would contain no moments in which pain, sadness, or other sources of negative welfare were dominant.

In any case, drab lives are very difficult to imagine. In the lives that we can relate to, there are inevitable moments and sources of negative welfare: the occasional illness, bereavement, and fear of death. In a life of no pleasures beyond muzak and potatoes, someone with a normal human psychology would not have constantly positive welfare but would in fact experience great boredom, and thus negative welfare. This is why Parfit stipulates that those living drab lives must have a psychology different to ours[11]. It might just not be possible to imagine a life that is both stripped of very good moments and is of always positive welfare. When we imagine drab lives, therefore, we might imagine something worse than a life of constantly barely positive welfare. Since drab lives might not be as bad as we first think, we should not consider the fact that we intuitively find them unattractive to be evidence against $RC_1$.

*III B. Short-Lived Lives*

These Z-lives are like A-lives but much shorter. They are like A-lives in the sense that welfare-per-minute (i.e., welfare of the life divided by the number of minutes in it) is the same; however, they are of shorter duration and therefore have less welfare in total.

To meet the condition of being only just positive, such Z-lives would have to be very short indeed. This confuses our intuitions: it is difficult for us to understand lives much shorter than our own. We regard death after one year as a shame, because to us it is premature and does not match our expectations or biological norms. Moreover, many of the best things in our lives rely on a degree of longevity—the completion of our projects, or the birth and success of our children, for example. But we have stipulated that welfare-per-moment is the same in A-lives and Z-lives, so Z-lives must have experiences that generate the same welfare as these things (though fewer and/or shorter of them than in A-lives). Such experiences are difficult for us to imagine and would generate much of the good in

11. Ibid.

short-lived Z-lives. They are probably, therefore, better than we at first imagine. There would still be, in a short-lived Z-population, a lot of very happy people. Parfit himself says that imagining such a population makes $RC_1$ "significantly less repugnant."[12]

### *III C. Roller-Coaster Lives*

Parfit also considers what he calls "roller-coaster Z."[13] Lives in this population contain all the good aspects of a life of very high positive welfare, but also contain enough sources of negative welfare to give them barely positive welfare overall—we can think of them as having high and low points, like a roller-coaster. For any A-life, a roller-coaster Z-life could be created that had all of its good aspects, interspersed with additional periods of suffering. Such lives are more like ours than drab or short-lived lives are. Human lives tend to contain some bad things as a matter of course, and many of the things we most enjoy come with inevitable downsides (your joy at your football team and your despair when they are defeated, for example).

### *III D. Job Lives*

The biblical character Job lived a very enjoyable life until God decided to test his faith by making him suffer.[14] We could think of Job's life before God's decision as an A-life—had Job's life ended before his test, he would have lived a life of very high positive welfare. Job's life following God's decision had very negative welfare. We could extend the bad period for long enough that it almost canceled out the welfare of the good period, producing a life of barely positive welfare. Job lives could be identical with any A-lives—with a bad period tacked onto the end, which would be the reason for their difference in welfare. Like roller-coaster lives, they would differ in longevity and suffering; otherwise, all the aspects of A-lives would be there. To my mind, this makes the idea that a population of Job or roller-coaster lives could be better than a population of A-lives less

12. Ibid.

13. Ibid.

14. In the biblical story, Job is eventually restored to prosperity, but for my purposes we can ignore this and consider his life as having simply a good part followed by a bad part. Thanks to Simon Beard for suggesting the life of Job to me as an example.

preposterous: to the extent that we value the things in A-lives, we should value those same things in Job and roller-coaster Z-lives.

*III E. Cinderella Lives*

Cinderella had a dreary and oppressed life until her fateful meeting with the prince, after which she famously lived "happily ever after." Her life is a mirror-image of Job's: a period of negative welfare followed by a period of very high positive welfare. Like a Job life, a Cinderella life could be of very low positive welfare overall, and a population of Cinderella lives could instantiate all the aspects of some A-population, differing only in the presence of a period of negative welfare added to the beginning of each life. Cinderella lives seem at least as attractive as Job and roller-coaster lives—*Cinderella* is an uplifting story, after all.

*III F. Chronically Irritated Lives*

A Z-population consisting in these lives is, I think, one of the best possible Z-populations. Take any A-population. Now imagine that every being in that population is on a heavy dose of painkillers. These drugs have no side effects, biological or social, and are administered in the most innocuous way—perhaps, like fluoride, through the public water system. Imagine too that every being in the population is beset by some chronic illness that would give them a constant degree of pain, were the painkillers not administered. Now we can imagine a sequence of populations identical to this one, but with the dose of painkiller (and hence its efficacy) gradually reduced, and the population simultaneously increased. Eventually—I am assuming that chronic pain can be bad enough for this to be a possibility—we end up with a population in which every life is just like the lives in the A-population with which we started, but due to the chronic irritation the welfare of each life is very low (but still positive)—and there are more of them. This is a Z-population, and I think it is one that could be better than the A-population with which we started, were it large enough.

The chronic irritation here can be thought of as a background drain on welfare at each moment of a life. Those in this Z-population would do everything those in the Z-population do, enjoy the same art, relationships, nature, and so on. They would just enjoy each of these things to a lesser extent, because every moment of enjoyment would be accompanied by an

irritating pain. (Though not necessarily enough pain to make any moment a negative source of welfare—there could be chronically irritated lives which, like drab lives, have slightly positive welfare at each moment.) Everything about these lives would be the same as in our A-population (and we could start with any A-population), apart from this background pain.

The Z-lives described here are barely worth living, in the sense that they have very low positive welfare. But they seem decidedly better than we might first imagine a life described as "barely worth living" to be. Drab lives deprive us of pain as well as pleasure; short-lived lives are as good, at every moment of their existence, as an A-life. Both are in any case very difficult to imagine humans living, which may confuse our intuitional responses to them. The more realistic (or at least relatable) roller-coaster, Job, Cinderella, and chronically irritated lives have the same positive aspects as A-lives. Of course, they also contain a lot of negative aspects, such that they are of very low positive welfare overall. But we might still think that these lives are better for the good things in them—good things that we overlooked when initially considering "lives barely worth living." To the extent that these descriptions make Z-lives appear better than we first imagine them to be, they make $RC_1$ less repugnant.

## IV. ANOTHER REPUGNANT CONCLUSION

Recall $RC_1$:

> $RC_1$: For any possible population of at least ten billion lives of very high positive welfare, there is some larger possible population of lives of very low positive welfare whose existence would be better, if other things are equal.

$RC_1$ is subtly distinct from $RC_2$.

> $RC_2$: For any possible population consisting of at least ten billion lives of very high positive welfare, the existence of any possible population that consists of lives of very low positive welfare and is sufficiently large would be better, if other things are equal.

$RC_1$ holds only that for any A-population there is at least one Z-population which is better, other things being equal. $RC_2$ holds that any type of Z-population meeting this ceteris paribus clause, so long as it contains enough lives, would be better than any A-population. To accept $RC_1$ without accepting $RC_2$, one must think that lives of the same welfare level can have different contributive values to the population in which they occur. (As I have defined welfare, all Z-lives must have similar value *to the people living them*, but this may come apart from the impersonal value they contribute to states of affairs.) Say that one thinks that Cinderella lives are much better, in this sense, than drab lives. For any A-population, one might think, there is a larger population of Cinderella lives that is better. Then one accepts $RC_1$. But one might also think that some A-populations are such that a population consisting of drab lives could not be better, however large. Then one would reject $RC_2$, as it holds that for any A-population a population consisting of enough Z-lives of any type would be better. $RC_1$ therefore does not imply $RC_2$, since one could hold the former while rejecting the latter. One can accept $RC_1$ without accepting that (for example) a population of drab lives could be better than an A-population. This is often overlooked, and emphasizing it makes $RC_1$ less repugnant.

$RC_2$ entails $RC_1$. If, for any A-population, there is a population of any type of Z-life (in which all other things are equal) whose existence would be better, then for any A-population there is a population of some type of Z-life (in which all other things are equal) that would be better. This is equivalent to $RC_1$. Since $RC_2$ implies $RC_1$ but is not implied by it, $RC_2$ is logically stronger than $RC_1$.

$RC_2$ can also be drawn as a conclusion from the Up Down Argument. In the Up Down Argument to $RC_1$, we could start with any A-population. We then add some number of Z-lives, while increasing the welfare of A-lives. This creates a better population. Then we increase the sum and equalize the distribution of welfare between individuals, making the population even better. If we could end with any type of Z-lives, then it appears that we could end with any type of Z-population, so long as it was big enough. And the principles of benign addition and non-anti-egalitarianism place no limits, other than the ceteris paribus clause, on what kind of Z-lives can be added or constructed at either step. In other words, the Up Down argument implies that for any A-population, any possible Z-population in which all other things are equal and is sufficiently large is better—that is, $RC_2$.

How does the distinction between $RC_1$ and $RC_2$ relate to the discussion of different types of Z-lives above? One should accept $RC_1$ if the Z-life one thinks best is such that some very large number of them could be better, for any A-population. However, if $RC_2$ is true this would have to hold for all Z-lives, including those one thinks worst. $RC_2$ is therefore harder to accept than $RC_1$. Since $RC_2$ is harder to accept than $RC_1$ and is just as much implied by the Up Down Argument, $RC_2$ puts more pressure on benign addition, non-anti-egalitarianism and transitivity. Those who wish to preserve those three principles must accept not only $RC_1$ but the more repugnant $RC_2$.

## V. PERFECTIONISM

Perfectionists place value on certain goods, which Parfit refers to as "the best things in life."[15] An extreme perfectionist view holds that one population is better than another if and only if it contains more or better of these good things. More moderate perfectionist views hold that it counts in favor of a population to some extent that it contains more or better good things. These perfectionist goods might include love, art, virtue, great achievements, and so on. Perfectionism has been regarded as a reason to deny the repugnant conclusion. I believe that this is a mistake, or at least half a mistake: both moderate and extreme perfectionism *favor* $RC_1$, though they would—if true—give us reason to reject $RC_2$. (I do not take a stand on whether either of these perfectionist views are true—extreme perfectionism in fact strikes me as implausible, though worth discussing as a pure expression of perfectionism. My claim is a conditional one—if either of these perfectionist views were true, this would support $RC_1$ but undermine $RC_2$.)

The claim that perfectionism favors $RC_1$ might be surprising. James Griffin, perhaps the first to bring perfectionism into contact with the repugnant conclusion, suggested that $RC_1$ could be rejected on the grounds that A-populations contain perfectionist goods and Z-populations do not (or that they contain such goods to a lesser degree). He writes,

15. Parfit, "Can We Avoid the Repugnant Conclusion?," p. 123. See also Derek Parfit, *On What Matters: Volume One* (Oxford University Press, 2011), p. 389; T. M. Scanlon, *What We Owe to Each Other* (Belknap Press of Harvard University Press, 1998), p. 125.

> Perhaps it is better to have a certain number of people at a certain high level than a very much larger number at a level where life is just worth living. Then we might wish to stop the slide [from A-populations towards larger Z-populations] . . . at that point along the line where people's capacity to appreciate beauty, to form deep loving relationships, to accomplish something with their lives beyond just staying alive . . . all disappear.[16]

Griffin's perfectionist goods are the appreciation of beauty, deep loving relationships, and accomplishments. He seems to suggest that lives without these goods (even if they are of a positive welfare level) are such that any number of them makes for a worse population than some number of lives that involve these goods—and further, that this justifies rejecting the repugnant conclusion.[17]

Griffin did not distinguish $RC_1$ from $RC_2$. His argument could only justify the denial of $RC_1$ if all Z-lives lack the perfectionist goods he mentions. However, if Z-populations can be made up of Job, Cinderella, roller-coaster, or chronically irritated lives, it is not always true that these perfectionist goods "disappear" in them, or even that they are fewer or worse in Z-populations than in A-populations. As I described them above, such lives include all of the good things present in A-lives, plus sufficient sources of negative welfare to render them as having barely positive welfare overall. For any A-population, and whatever we take perfectionist goods to be, we can construct Z-populations which instantiate those goods to at least the same degree. We can do this by adding a period of very negative welfare to the beginning or end of each life in the A-population, creating a Z-population of Cinderella or Job lives, respectively. Or we can intersperse each A-life with occasional pains to create roller-coaster lives, or add a low-intensity background pain to create chronically irritated lives. Each Z-life thus created will mirror an A-life with respect to all the good things in life (whatever they are) and yet still be of very low positive

16. James Griffin, *Well-Being: Its Meaning, Measurement and Moral Importance* (Clarendon Press, 1986), p. 340.

17. "[Parfit's argument to RC1] treats well-being as measurable on a single continuous additive scale, where low numbers, if added to themselves often enough, must become larger than any initial, larger number. But this seems not true in prudential cases, and it would seem likely that this incommensurability in prudential values would get transferred to interpersonal calculation." Griffin, *Well-Being*, p. 340.

welfare. Z-populations need not, therefore, instantiate fewer or worse perfectionist goods than A-populations.

Furthermore, if such a Z-population is much larger than a comparator A-population—as the Z-populations involved in $RC_1$ and $RC_2$ are—then there is good reason to think that it would instantiate more and better perfectionist goods. In a larger population, there would be more people appreciating each beautiful thing, more people of virtue, more people accomplishing their projects. So there would likely be more perfectionist goods in a much larger population of, say, roller-coaster Z-lives, than in a smaller population of A-lives. There would also likely be better perfectionist goods. A much larger population means a greater range of possible relationships and interactions, new cultures, and new artists. Much great art stems from the interaction of different people: if there were one thousand times as many people with the gifts of Shakespeare or Bach, there would not merely be one thousand times as many *Hamlets* and fugues. Rather, there would be greater works building on conversations and exchanges between such people. Likewise, there would not simply be more people accomplishing their projects, but a greater range of projects to accomplish, with the help of more fellow-travelers and constructed from more diverse social forms. Assuming there is no upper limit to population size, then, for any A-population a much larger Z-population is possible, with each life instantiating all of the goods of the lives in the A-population and having more—and better—of the good things that come with population size.

The implication is this: far from being a reason to reject it, the extreme perfectionist view that one population is better than another if and only if it has more and better perfectionist goods, *entails* $RC_1$. For any possible population of at least ten billion lives of very high positive welfare, there is some larger possible population of lives of very low positive welfare which would be better with respect to the best things in life, and therefore whose existence would be better, other things equal, according to extreme perfectionism. The moderate perfectionist view, which holds only that instantiating more and better perfectionist goods counts in favor of a population, does not entail $RC_1$, as a moderate perfectionist could think that this consideration was outweighed by a concern for, say, average welfare or degree of suffering. However, moderate perfectionism would say that, for such Z-populations, there is something that counts in favor of them and against the A-population they are compared to. Furthermore,

since the size of a Z-population has no upper limit, it is plausible (given my arguments above) that the quality and quantity of perfectionist goods in a Z-population has no upper limit. Thus, a moderate perfectionist, to avoid $RC_1$, would have to hold that however much a Z-population bettered an A-population in terms of perfectionist goods, this would always be outweighed by other considerations which favored the A-population. Perhaps the additional amount of suffering in such Z-populations—which would increase along with their perfectionist goods—would always be more significant, or perhaps perfectionist concerns are lexically dominated by the value of average welfare.[18] These are not implausible positions, but they are positions whose perfectionism seems not just moderate, but very modest.

Griffin's argument works much better against $RC_2$. $RC_2$ is false if there is some kind of Z-population such that, however large, there is some A-population that it cannot be better than. Extreme perfectionism implies that $RC_2$ is false, and moderate perfectionism is a consideration against accepting it. Consider populations of drab lives. These lives instantiate no perfectionist goods. At least some, and plausibly all, A-lives do. So there are A-populations such that a Z-population—of any size—consisting in drab lives instantiates fewer (and no better) perfectionist goods—because it instantiates none. According to the extreme perfectionist view, this is sufficient to conclude that such Z-populations could not be better than such A-populations, and therefore that $RC_2$ is false. According to moderate perfectionism, their lack of perfectionist goods is a reason to think that drab Z-populations would be worse than at least some A-populations, and therefore is a reason to reject $RC_2$, though it could be outweighed by, say, a concern for total welfare.

There are two important objections to address here. The first is skepticism about my premise that for any A-population, we can construct Z-populations which instantiate the same goods by adding sources of negative welfare to each life. This skepticism might stem from the view that the welfare of lives is determined not additively, but narratively. That is to say, we cannot just add up the welfare in each part of a life to find its overall welfare; rather, the relations between the parts (the "shape of a life") matters. This casts doubt on the idea that we can take an A-life, add

18. I thank an anonymous reviewer for persuading me to clarify and soften this conclusion.

some parts of negative welfare to it—as we do in Cinderella, Job, chronically irritated, and roller-coaster lives—and come out with a Z-life.[19]

Nothing that I have said relies on a purely additive account of welfare. It is consistent with my argument that, for instance, a life of X years of low welfare followed by Y years of high welfare instantiates greater welfare, overall, than a life of X years of high welfare followed by Y years of low welfare. (This would suggest that for Cinderella lives and Job lives to both be of the same very low welfare the bad period of Cinderella's life would have to be longer or worse than that of Job's, or the good period shorter or not as good.) All that my argument needs is that adding some amount of badness to a life makes it worse, to the extent that what would be an A-life becomes a Z-life. The amount of badness to be added might be very big, and it might matter when in the life it occurs and how it relates to other parts of the life. The objector here must hold, not the weak and plausible claim that the shape of a life matters, but the stronger and less plausible one that some A-lives cannot be made into lives of very low positive welfare by *any* addition of negative welfare.

The second objection is that varying the quantity and quality of perfectionist goods between comparator A- and Z-populations violates the "other things equal" clause of $RC_1$, $RC_2$, and $RC_0$.[20] One thing to say is that if this were true, perfectionist concerns would be irrelevant to the Repugnant Conclusion either way. I would not be able to make the argument that perfectionism favors $RC_1$, but neither would Griffin be able to make the argument that it favors $RC_1$'s rejection.[21] It would also mean that Parfit misunderstood his own conclusion: he believed perfectionist concerns were relevant to it, as we will see below.

In my view, the "other things equal" clause serves to preclude variance between the A-populations and Z-populations being compared that does not stem from the facts about them that make them A-populations and larger Z-populations. For example, consider a population of trillions of very beautiful people leading drab lives. Their beauty does not affect their welfare; however, on some views it might increase the value of this population—perhaps to the level at which such a population could be better than any A-population of ten billion people. A view implying this

19. Thanks to James Wilson for pressing this objection.
20. Thanks to Teruji Thomas for pressing this objection.
21. Thanks to Bastian Steuwer for helping me to see this response.

would not thereby imply $RC_1$, because of the "other things equal" clause. This Z-population differs from A-populations in the degree to which it instantiates beauty, which is unconnected from the fact that it is a population of lives of very low positive welfare.

However, some things must vary between an A-population and a much larger Z-population, simply in virtue of them being an A-population and a much larger Z-population. Z-lives cannot be exactly like A-lives and simply receive a lower welfare score. They must have more sources of negative welfare, or fewer of positive welfare. Two populations cannot be of different sizes without there being more of the things that occur in each individual's life (more births, more deaths, more friendships, more lazy Sunday afternoons, and so on). So the "other things equal" clause should not exclude differences between populations that affect, or are affected by, the number or welfare levels of the lives in them. Consider how a larger Z-population of Cinderella lives differs from any comparator A-population. It is larger, and everyone's life involves a period of suffering at the start. These are the features that make it a larger Z-population rather than a smaller A-population. But these features are accompanied by other differences: one being that people in the Z-population know what it's like to endure a period of suffering, another being that there is a greater number of pancreases in the Z-population, another being—if the argument so far in this section is correct—that there are more and better of the best things in life. This last difference is not an additional variation between the two populations, as a difference in natural beauty would be. It is a consequence of the populations being as they are stipulated to be. Thus, the "other things equal" clause restricts what can vary between the populations in question, but does not rule out populations which are made up of the kinds of Z-lives I describe above, and are much bigger than their comparator A-populations.

We can overcome these objections and stick with our conclusion, then, that both extreme and moderate perfectionism favor $RC_1$ and disfavor $RC_2$. To the extent that we care about the best things in life, we should not find $RC_1$ so repugnant. It is $RC_2$ that perfectionism—if it is correct—supplies reasons against.[22]

22. For another recent discussion of perfectionism and the repugnant conclusion, see Simon Beard, "Perfectionism and the Repugnant Conclusion", *Journal of Value Inquiry* 54 (2020): 119–140. Beard focuses solely on $RC_1$ and uses slightly different arguments to draw the similar conclusion that we cannot avoid $RC_1$ by appealing to perfectionism.

## VI. PARFIT'S LATE STRATEGY

Parfit, in his later work, searched for a way of avoiding the repugnant conclusion. He did not explicitly endorse either extreme or moderate perfectionism as I have defined them. However, he did invoke perfectionism in the hope of resisting the Up Down Argument for $RC_1$. Recall that in $p_+$ there are over ten billion lives of very high welfare and many more of very low positive welfare, and in $p_z$ every life is of low positive welfare:

> [In the best lives in $p_+$] the best things in life would be very good, and lives . . . in [$p_z$], would not include any of these good things. There would be no art, or science, no deep loves or friendships, no other achievements, such as that of bringing up our children well, and no morally good people. [$P_z$] would be much worse than [$p_+$] in what we can call *qualitative* or *perfectionist* terms . . . This great qualitative loss would, I believe, make [$p_z$] in itself a worse world than [$p_+$], even though [$p_z$] would give, to the same number of people, a greater and more equally distributed sum of well-being.[23]

Parfit's idea is that the losses with respect to perfectionist goods incurred in switching from a population in which there are some lives of very high welfare to one in which there are none cannot always be compensated for by increases in total welfare and equality. More specifically, he seems to believe that "great" losses and total losses (such that there are no perfectionist goods) cannot be compensated for in this way. His view can be captured by the following principle:

> Parfit-perfectionism: A population with no or much fewer and worse perfectionist goods could not be better than a population with some sufficient level of them, if the only other differences between the two were in total and average welfare and equality.

This view does not imply extreme perfectionism because it does not rule out that such compensation could sometimes work—for example, for small non-total losses with respect to perfectionist goods. Nor does it

23. Parfit, "Can We Avoid the Repugnant Conclusion?," p. 123.

imply moderate perfectionism because it does not entail that having more and better perfectionist goods always counts in favor of a population.

Parfit-perfectionism is incompatible with non-anti-egalitarianism, however. Non-anti-egalitarianism implies that a population like $p_z$ is always better than one like $p_+$, as it has greater total and average welfare, distributed more equally. It implies this even if the former has no, or much fewer and worse, of the best things in life. Imagine that $p_z$ consisted in drab lives. It would then contain no perfectionist goods. The only good things in a drab $p_z$, by stipulation, are muzak and potatoes. A $p_+$ population, on the other hand, could—and perhaps must—contain very many perfectionist goods, as it has over ten billion lives of very high welfare. This version of $p_z$ would therefore contain no perfectionist goods when $p_+$ contained very many and hence would contain much fewer (and no better). According to both of the Parfit-perfectionist principles, such a $p_z$ could not be better than such a $p_+$, no matter its total and average welfare, or its equal distribution.

If we are convinced by Parfit-perfectionism, it therefore seems that we should reject non-anti-egalitarianism. It is not always the case, according to Parfit-perfectionism, that if two populations are so related that one has a higher average welfare, a higher total welfare, and a more equal distribution of welfare than the other, and all other things are equal between them, then the existence of the former would be better. For instance, it is not the case when the former is a drab $p_z$ and the latter a $p_+$ (that contains very many perfectionist goods, as perhaps any $p_+$ must).

Will rejecting non-anti-egalitarianism (and hence the Up Down Argument as formulated above) allow us to avoid the repugnant conclusions $RC_1$ and $RC_2$? No and yes, respectively. Parfit-perfectionism can be accommodated by transforming non-anti-egalitarianism into the following principle:

> Non-anti-egalitarianism*: If $p_1$ and $p_2$ are so related that $p_1$ has a higher average welfare, a higher total welfare, and a more equal distribution of welfare than $p_2$, and $p_1$ has at least as many and as good perfectionist goods as $p_2$, and all other things are equal between $p_1$ and $p_2$, then the existence of $p_1$ would be better than that of $p_2$.

Non-anti-egalitarianism* is consistent with Parfit's idea, embodied in the Parfit-perfectionism, that some losses in perfectionist goods cannot be

compensated for by welfare or equality: one could hold such a perfectionist view alongside it (non-anti-egalitarianism* is also consistent with the extreme perfectionist view that no such losses can be compensated for). $RC_2$ does not follow from benign addition, transitivity and non-anti-egalitarianism*. This is because the additional clause in the latter means that it does not entail, as non-anti-egalitarianism does, that a $p_z$ consisting of drab lives would be better than $p_+$, for any $p_+$ that contains perfectionist goods. For drab $p_z$ populations, this is true however large they are, because drab populations of any size instantiate none of the best things in life. Therefore, if we accept Parfit-perfectionism, we cannot use the Up Down Argument to conclude that, for any A-population, any possible population that consists of lives of very low positive welfare and is sufficiently large, and in which all other things are equal, would be better. Specifically, we cannot conclude this when the Z-population consists of drab lives. Parfit's perfectionist concerns help us to avoid $RC_2$.

However, non-anti-egalitarianism* can be substituted for non-anti-egalitarianism in a modified Up Down Argument to deliver $RC_1$. This is because it need not be the case that there is a loss with respect to perfectionist goods incurred in switching from $p_+$, in which there are some lives of very high welfare, to $p_z$, in which there are none. This depends on the make-up of the latter. If $p_z$ consisted of drab lives, as we have seen, there would be such a loss. But if $p_z$ contained many Job, Cinderella, roller-coaster, and chronically irritated lives, there could be just as much (or more) art, science, deep loves and friendships, well brought-up children, and moral virtue, as in $p_+$, though each individual's welfare level averaged out at barely positive. As we have seen, such lives can be constructed by taking A-lives (lives like the best lives in $p_+$) and adding some sources of negative welfare to them. Such lives would therefore instantiate all of the good things found in the best lives in $p_+$. The difference between $p_+$ and $p_z$ would therefore be that the best lives from $p_+$ were transformed into Cinderella, Job, roller-coaster, or chronically irritated lives instantiating exactly the same good things, and that the worst lives from $p_+$ were made slightly better off in $p_z$ with respect to welfare (which could but need not mean that these lives instantiate more perfectionist goods). Such a $p_z$ would not be worse with respect to perfectionist concerns than such a $p_+$.

In this case, for any $p_+$ population, we can make a $p_z$ such that it has all the good things present in $p_+$, so that there would be no perfectionist

loss, as well as a gain in equality, total, and average welfare in switching from $p_+$ to $p_z$. Together with benign addition and transitivity, non-anti-egalitarianism* entails $RC_1$, while respecting Parfit's perfectionist commitments. Those commitments, therefore, do not help us to avoid $RC_1$.

In a paper unfinished at his death and published posthumously, Parfit endorsed what he called the "wide dual person-affecting principle."

> WDPAP: One of two outcomes would be in one way better if this outcome would together benefit people more, and in another way better if this outcome would benefit each person more.[24]

The first part of this principle relates to a whole population considered collectively. If an outcome "together benefits people more" than some other, for Parfit, this means that the benefits instantiated for the people in that outcome are greater in value than the benefits instantiated in the other. Having a life of positive welfare, he argued, could count as a benefit; the value of the benefit depending on the level of welfare. Each life of positive welfare contributes positively to the value of the collective benefit of the outcome. A Z-population could collectively benefit people more than an A-population, if it had enough lives in it. So the first part of WDPAP, considered alone, leads to $RC_1$ and indeed $RC_2$.

The second part of the principle does not. If an outcome "benefits each person more" than some other, for Parfit, this means that for each person in that outcome, they receive more benefits than they do in the other. Each person in an A-population receives a very large benefit, namely, a life of very high welfare. Each person in a Z-population receives a very small benefit, namely, a life of barely positive welfare. So any A-population "benefits each person more" than any Z-population, and therefore is better—this is, of course, inconsistent with $RC_1$ and $RC_2$.

The first part of WDPAP implies that for any A-population, there is a Z-population (of any kind, so long as it is sufficiently large) in which other things are equal and is *in one way* better. The second part implies that *in another way* any A-population is better than any Z-population. To use WDPAP to avoid $RC_1$ and $RC_2$, Parfit needs the second part of the principle to not be outweighed by the first.

24. Parfit, "Future People, the Non-Identity Problem, and Person-Affecting Principles," p. 154.

Parfit recognized this, noting that WDPAP could be "one premise of more complicated and forceful arguments *for* the Repugnant Conclusion" (my emphasis) unless we also "justifiably believe that great losses in the quality of people's lives could not be outweighed by any increase in the sum of benefits, if these benefits came in the lives of people whose quality of life would be much lower."[25] Such a belief would mean that the first part of WDPAP could not outweigh the second when considering a Z-population and an A-population: though the former would produce more benefit, these benefits would be in lives of much lower welfare than in the latter. How could this belief be justified? Parfit hoped, I think, that perfectionism supplied the answer;[26] the argument would be that "great losses in the quality of people's lives" are associated with great losses in perfectionist goods, which according to principles like Parfit-perfectionism cannot be outweighed by gains in total welfare. If what I have said above is correct, this justification fails, at least for some Z-populations. For such welfare, losses could occur without any loss with respect to perfectionist goods—for example, in the comparison between an A-life and a Z-life that differed only in the background presence of chronic pain.[27]

I have said nothing about the plausibility of the Parfit-perfectionism itself, or of WDPAP. The upshot of my arguments is this if Parfit-perfectionism is true, then $RC_2$ is false, and the Up Down Argument to it should be resisted by rejecting non-anti-egalitarianism. However, those principles do not tell against a reformed version of the Up Down Argument which

25. Ibid., p. 157.

26. In the final sentence of "Future People, the Non-Identity Problem, and Person-Affecting Principles," Parfit writes "I have started to defend this belief elsewhere"—presumably gesturing at the discussion of perfectionism in "Can We Avoid the Repugnant Conclusion?."

27. Perhaps another way of justifying the belief Parfit mentions is suggested by John Broome, in *Weighing Lives* (Oxford University Press, 2004, chap 10; chap. 14). It might be that Z-lives, though worth living, fall below what Broome calls "the neutral level," and thus adding such lives would not make a population better. Thus, they could not outweigh any losses in value from lives above the neutral level. Broome's view nevertheless—as he acknowledges—implies a kind of repugnant conclusion: that for any possible population of at least 10 billion lives of very high positive welfare, there is some larger possible population of lives of welfare only marginally above the neutral level whose existence would be better, if other things are equal. This would be less repugnant than $RC_1$ insofar as the neutral level is higher than the welfare of a Z-life. However, a high neutral level exacerbates what Broome calls "the negative repugnant conclusion." Ultimately, Broome was not as worried about $RC_1$ as Parfit, and it seems that when Parfit set out to avoid the repugnant conclusion, he wished to avoid it by a greater distance than Broome's view does.

employs non-anti-egalitarianism* in place of non-anti-egalitarianism and concludes with $RC_1$. If we accept WDPAP, we can only avoid $RC_1$ and $RC_2$ if we can justify the primacy of benefits (in terms of welfare) to each person over collective benefits (that is, total welfare). Parfit-perfectionism could only play this justificatory role if welfare losses were associated with losses in perfectionist goods. However, I have argued that there can be losses, including very large ones, in welfare without losses (and even with gains) in perfectionist goods. There may be ways in which we can justify the primacy of the second part of WDPAP over the first, but perfectionism is not one.

## VII. CONCLUSIONS

Considering more detailed descriptions of lives of very low positive welfare makes the repugnant conclusion less difficult to swallow. Some Z-lives, such as Cinderella or chronically irritated lives, do not seem so very bad as to render the notion that populations made up of those lives could be better than A-populations unthinkable. Even drab lives have the very real advantage over our lives of the absence of moments dominated by pain, boredom, and other sources of negative welfare.

If for any A-population, there is one Z-population—a very large number of chronically irritated lives, for example—whose existence would be better (and is not excluded by the "other things equal" clause), then $RC_1$ is true. $RC_2$ is a stronger claim: for it to be true, for any A-population, every kind of Z-life must be such that some number of them could be better. For those who find some kinds of Z-life worse than others, $RC_2$ is less acceptable than $RC_1$ and a greater problem for the principles of benign addition, non-anti-egalitarianism and transitivity, which together imply it.

Perfectionism is no consideration against $RC_1$, since some Z-populations could instantiate at least as many and as good of the best things in life as any A-population. In fact, since some Z-populations could be better with respect to these goods, for any A-population, extreme perfectionism entails $RC_1$ while moderate perfectionism weighs in its favor. Perfectionism is a consideration against the stronger claim $RC_2$, since there are some Z-populations (drab ones) which are without the best things in life. Parfit's perfectionist strategy for avoiding the repugnant conclusion could only be partially successful. While his perfectionist principle undermines non-anti-egalitarianism and therefore (if justified) blocks the

argument to $RC_2$, it can be accommodated by non-anti-egalitarianism* in an argument that still implies $RC_1$. Since not all Z-populations are worse off with respect to perfectionist goods than A-populations, perfectionism cannot be used to motivate the primacy of the second part of the WDPAP, which Parfit would have to do to use WDPAP to avoid $RC_1$ and $RC_2$.

Those who advocate accepting $RC_1$ may be heartened by these findings. At least some lives of very low positive welfare are not as bad as we might at first imagine, and perfectionists cannot (and should not, by their own lights) resist $RC_1$. However, if one wishes to preserve benign addition, non-anti-egalitarianism, and transitivity, one must also accept a stronger claim ($RC_2$) which perfectionist principles supply arguments against, if they can be justified.