**MERE ADDITION AND THE SEPARATENESS OF PERSONS[*]**
**Matthew Rendall**
**Forthcoming in the *Journal of Philosophy***

**This is the post-peer review version of the paper, but it has not yet gone through copyediting. There may still be minor changes.**

Imagine a continent inhabited by ten billion people, all equally very well off, and call this state of affairs world A. Unknown to the people of the first continent, you may choose to populate a second continent with a group of the same size, each of whom will be worse off than the first group, but whose lives will still be worth living. This combined population, world A+, seems at least as good as A, if not better. Suppose that you can then redistribute so that you reduce each person's utility on the first continent, but raise the utility of each on the second continent by a greater amount. The resulting population, B, contains 20 billion people, each of whose lives is of equally high quality, but not quite as good as those of the people in A or the better-off group in A+. Again, this seems an improvement. Finally, imagine that you repeat the process twenty-four times. Each step seems to lead to a better outcome. The result is a world with a huge population, Z, each of whose lives is barely worth living. If the relation "better than" is transitive, then it seems Z must be the best of all these worlds: what Derek Parfit calls the repugnant conclusion.[1]

Most of us agree that this conclusion is repugnant. But Z undoubtedly contains more utility. If we want to hold that it is worse, then we had better have a good reason. Some writers have suggested that A is lexically superior to Z: though the Z lives have value, no number of Z lives could be worth as much as a single life in A. Even if we adopt this view, however, it is hard to see how it could make the world worse to move from A to A+. It seems that creating lives worth living should, if anything, be good, if it has no effect on others. As Michael Huemer says, "Worthwhile lives are good. More of a good thing is better."[2] Once we have added the new people, it is hard to say why we should not redistribute—which both increases utility and benefits the worse-off—and move to B. On reaching B, there seems no reason not to move to B+…and on down the alphabet, leading to the paradoxical conclusion that Z must be better after all. Accepting that A could be lexically better will not enable us to avoid the repugnant conclusion. We must also solve the mere addition paradox.

We might argue:

1. It would be preferable to live one wonderful life in A+ than any length of time in B or Z, or any number of lives in B or Z in sequence;

2. One superior A+-life is preferable to any number of B- or Z-lives;

3. The value of a world is the sum of the desirability of its individual lives;

4. ∴ A+ is better than either B or Z.

Even if we accepted premise 1, this inference would still be dubious. "In the interpersonal case," as Jeff McMahan pointed out in an early review of the population ethics literature, "there is nothing analogous to the preference for a single [superior good] over any number of [inferior goods] on which to base the claim that a single life might contain more of what is valuable than any number of lives of a lesser quality."[3]

McMahan presented this as a criticism of attempts to block the repugnant conclusion by appealing to lexical superiority. But in fact, the distinction between intrapersonal and interpersonal aggregation helps explain how the mere addition paradox arises. If it would be better to live any single life in one world than through any combination of lives in another, then we may safely infer that that the former world is lexically preferable, as if we were aggregating intrapersonally. Such lexical judgments involve the "suspension of addition": the utility in the latter world, however great in quantity, effectively does not count. Such reasoning is defensible so long as welfare levels are uniform in each world. When, however, we compare worlds containing better- and worse-off people, it is no longer true that any life in the first world is better than any combination of lives in the second. The condition that allowed us to infer that the first world was lexically better and suspend addition is no longer met. When one pairwise comparison suspends addition, and the other does not, the result is the mere addition paradox.

## I.  The Repugnant Conclusion and Intrapersonal Aggregation

Could A really be better than Z? The repugnant conclusion may seem arithmetically inescapable. Though the lives in A are much better than those in Z, the sheer number of lives in the latter will ensure that they add up to more utility. Something similar, however, could also be true within a single life. Parfit asks us to imagine choosing between a hundred wonderful years of life, or a Drab Eternity in a world whose only pleasures are muzak and potatoes.[4] Assume that each jingle and potato brings as much paltry pleasure as the first. Most of us would still pick the Century of Ecstasy. Why?

One might think that the pleasures of the Century of Ecstasy are lexically better, in line with Mill's claim that one good is superior if informed people prefer it to any amount of another. Parfit's discussion suggests that he takes this view.[5] But other writers lack this intuition, and

doubt that it is defensible. "Suppose we are offered the choice between living another minute in ecstasy and having another thirty drab years of muzak and potatoes," comments Noah Lemos. "I, for one, would choose the thirty drab years."[6] More plausible is the claim that no amount of potatoes could be worth trading away a satisfactory *life*. Human welfare, as Dale Dorsey notes, is plural in nature: discrete pleasures and pains are important, but "global plans and projects" play a crucial role. More pleasure is better, other things being equal, but we have an overriding interest in our lives being good enough, satisfactory, a success. Whether a life succeeds or fails is not simply a matter of the sum of utility: we could coherently judge the life in the Drab Eternity a failure *however* long it lasted.[7] Our judgments in such cases, as James Griffin writes, "take a global form: this way of living, all in all, is better than that."[8]

These claims are simplest to defend on a desire-utilitarian or objective list account of welfare. Hedonistic utilitarians may seem committed to summing pleasures and pains. Yet suppose you have a choice between the Century of Ecstasy and the Drab Eternity. You choose the Century of Ecstasy because it sounds like a greater pleasure. On your deathbed, you still think it was. Your guardian angel reveals that had you chosen the Drab Eternity, you would have regretted having done so. It seems that you could conclude that the Century of Ecstasy *was* a greater pleasure. On this *preference-hedonist* account, as Parfit puts it, "one of two experiences is more pleasant if it is preferred."[9]

Such a judgment involves what Griffin calls the "suspension of addition": "the relevant notion of aggregation cannot be simply that of summing up small utilities from local satisfactions; the structure of desires already incorporates, constitutes, aggregation." Griffin contrasts his approach with Parfit's,[10] but some of Parfit's remarks are very similar. He too holds that on a "summative" approach, the Drab Eternity offers more utility, but that it does not offer

"a better life." The Century of Ecstasy is superior "in an essentially qualitative way."[11] This view is controversial, but not obviously unreasonable.[12] Moreover, it seems that we must accept that some goods can be lexically superior to others if we are to reject the repugnant conclusion. If the value of lives is simply the sum of their utility from moment to moment, then Z must be superior to A as a matter of simple arithmetic. If we claim that A is better than Z when both are taken as wholes, then, we *necessarily* suspend addition.

Having argued that an individual could rationally choose the Century of Ecstasy, Parfit's next move is to suggest that that this approach can be carried over to comparing the value of different populations.[13] This argument is similar to C. I. Lewis's proposal that in comparing the value of two objects, we should judge superior the one that gives rise to the set of experiences which we would prefer to live through in sequence.[14] Rawls accused Lewis of "the conflation of all desires into one system of desire," thus ignoring the separateness of persons.[15] In a similar vein, Jesper Ryberg observes that

> [s]trictly speaking, Parfit's argument involves some kind of trick. It assumes that preferences with regard to comparison of life scenarios can be transferred into comparison of population scenarios. However, when we move from the intra-personal case to the inter-personal case there is nothing analogous to the preference of a single individual.[16]

It is true that we cannot assume that intrapersonal and interpersonal aggregation are equivalent. Nevertheless, in this case the trick seems justified. The lives in the Drab Eternity contain only the transitory pleasures of muzak and potatoes, not any "plans and projects" that attach to lives as a whole. If we stipulate that these mild pleasures never wear thin, but that they also never add up to any greater good, one could rationally be indifferent to experiencing them in

the course of one infinitely long life and a series of separate lives. And if anyone could rationally prefer a single life in A to eternal life— or any number of lives—in Z, then it would be very strange to claim that Z was the preferable world. We can plausibly hold that:

5. If it would be preferable to live any single life in one world than to live through any combination of lives in another, then the first world is preferable for people.

6. It would be preferable to live any single life in A than any combination of lives in Z.

7. ∴ A is preferable to Z for people.

8. *Ceteris paribus,* a world that is preferable for people is better *simpliciter.*

When we say a world is preferable for people, we mean that is preferable for people *de dicto.* We compare possible populations of various sizes and compositions, and ask which set of lives would be most desirable, all things considered, if one were to live them end to end. The world we choose will be the one that is most desirable *for people*–as opposed to what might be desirable from some impersonal standpoint, such as one concerned with the just distribution of welfare—but not for any *particular* people. We are not deciding what would be preferable for some previously defined population *de re.* Some writers, to be sure, have held that we should consider only the interests of people who will at some point exist. But it is not clear how we could use this criterion to decide the optimal population size: the very point at issue *is* who should exist.[17] Alternatively, we might take the group to be both actual and possible people. R. M. Hare held a version of this view.[18] Hare maintains that we should aggregate what would be the retrospective preferences of all actual and possible individuals for having come into existence. This view is coherent, but it leads to the repugnant conclusion: the sheer number of people in Z who would retrospectively have a preference for having been born would lead us to judge Z more desirable than A.[19] While to dismiss Hare's view on this ground would beg the

question, it is fair to note that it has found few supporters.[20] Since the very question we are deciding is which population would be best, the *de dicto* criterion seems appropriate, and has more intuitively plausible implications.[21]

Even if World A is preferable for people, this does not prove that it is better. It might be that impersonal ideals such as equality or perfection affect the goodness of outcomes.[22] But everyone in A is equally well off. Egalitarians can have no complaint. So far as perfection is concerned, World A is, of course, much better than Z. We can reasonably judge that

9.  Other considerations speak at least as much in favor of A as Z.

10.  ∴ A is better than Z.

This gives us prima facie grounds for rejecting the repugnant conclusion. But it does not solve the mere addition paradox. Let us see why.

## II.  The Mere Addition Paradox and Interpersonal Aggregation

Suppose we start at A, and try to decide which world would be best. Just as an individual may accept reductions in the quality of her life in order to increase its extent, we may improve the world by trading reductions of quality against increases in population. The place to stop, as Griffin suggests, may be part way through the alphabet.[23] First, we should ask whether each individual would rationally prefer her present life in A, or a proportionately larger number of lives in B. If each would find the latter preferable, then we may conclude that overall B would be better than A, and move down the alphabet. Eventually, we will reach a point—say, world J—in which each would prefer a single life to any number of lives in K. Surely J will be better for people *de dicto* than K, or any world that comes after it.

Life in J is good. Higher up the alphabet, life is better, but far fewer enjoy it. Further down the alphabet it becomes dreary, pinched, and at last hardly worth living. Now, suppose we can

add an equal number of new people, all with mean and impoverished lives. Because these lives are nevertheless (barely) worth living, and they harm no one, it seems that J+ should be at least as good, if not an improvement. We can plausibly claim that

11. If two worlds contain the same number of satisfactory and equally good lives, and the second world contains additional lives that are unsatisfactory, but nevertheless worth living, then the second world is at least as desirable for people *de dicto*.

12. J+ contains as many satisfactory lives as J, all of which are just as good, plus additional lives that are unsatisfactory but still worth living.

13. ∴ J+ is at least as desirable for people *de dicto* as J.

14. *Ceteris paribus*, a world that is at least as desirable for people *de dicto* is at least as good *simpliciter*.

Some may now question whether other things *are* equal. Though J+ is preferable for people, some might hold that its inequality makes it worse than J.[24] Others might hold that the poorer lives make J+ less perfect, and therefore inferior.[25] But both these views face compelling objections. Remember that mere addition makes nobody worse off. If we remain at J, the only effect on anyone's welfare will be that people who could have enjoyed impoverished, but nevertheless worthwhile lives, lives which they would have been glad were created,[26] will never come into existence. To judge J better than J+ on grounds of equality, when that equality benefits no one, seems pointless, if not downright perverse.[27] Nor is it easy to see how mere addition would make the world less perfect when the original and new people will live on separate continents and have no contact with one another at all. To say that it would is like claiming that the presence of a Norman Rockwell in the local art museum would diminish the value of the

Mona Lisa.[28] Why not hang the painting if people enjoy it? How could it reduce the world's value if no visitor to the Louvre ever sees it?

We can reasonably hold, then, that

15. Other considerations speak as much in favor of J+ as J.

16. ∴ J+ is at least as good as J.

Yet should we move to J+, we face a dilemma. If we redistribute, we will produce K. K is *not* better than J. Yet compared with J+, it "contains greater total happiness, greater average happiness, much more equality, and a much higher standard of living for the worst-off."[29] How could we decently refuse?

It might seem that we could refuse. The superior lives in J+ are, after all, lexically preferable. One might think that this makes that world more valuable. Parfit considered but ultimately rejected an "elitist" solution to the mere addition paradox. Suppose some Europeans have lives barely *well* worth living, and some African children have lives that are barely worth living at all. If the former lives are lexically more valuable, it would make the world worse, on this view, to redistribute from the former group to the latter. "It is hard," as Parfit observes, "to believe this."[30] Yet if we do not believe it, we face a paradox. We have judged one life in J more desirable—thus more valuable—than any number of lives in K. Hence J is a more valuable world. It seems that once we move to J+, these same lives should retain their value. The new lives in J+, while inferior, are worth living. If anything, they should increase J+'s value. Something here doesn't add up. Why isn't J+ also superior to K?

The answer is that we can no longer suspend addition. Suppose we were to argue:

17. If it would be preferable to live any single life in one world than to live any combination of lives in another, then the first world is preferable for people *de dicto*.

18. It would be preferable to live any single life in J+ than any combination of lives in K.

19. *Ceteris paribus,* a world that is preferable for people *de dicto* is better *simpliciter*.

20. Other considerations speak as much in favor of J+ as K.

21. ∴ J+ is better than K.

This inference would clearly be unsound. First, premise 18 is false. To be sure, it would be preferable to live any *privileged* life in J+ than any number of lives in K. But for half the population in J+, it would be desirable to move to K. Each new person will have a strong preference for redistribution—probably stronger than that of each original person for preserving J+. Indeed, arguably that is what it means to say that K has more utility. We can no longer claim, in this new pairwise comparison, that the world with the privileged lives is preferable for people *de dicto*. Moreover, other things are no longer equal. If we attach intrinsic importance to equality, or give greater weight to benefits that go to the worse off, then this is a further reason to consider K better. J+ is, to be sure, better in terms of perfection, but that can scarcely be decisive. Premise 20 is arguably false as well.

Indeed, it seems that we could defensibly conclude that

22. If in two worlds with populations of the same size, one world has greater utility, then that world is preferable for people *de dicto*.

23. J+ and K have populations of the same size, and K has greater utility.

24. ∴ K is preferable for people *de dicto*.

25. *Ceteris paribus,* a world that is preferable for people *de dicto* is better *simpliciter*.

26. Other considerations speak at least as much in favor of K as J+.

27. ∴ K is better than J+.

Choosing K will deprive each of the privileged people of a life she would value more than any number of K lives. Might it nevertheless be better to remain at J+? Even if we were tempted to believe this, we could modify the example. Taking inspiration from Parfit's Second Paradox,[31] suppose that we can add to J, not just an equally large population, but rather a million times as many people, and suppose that redistribution can still produce K. Since a J life is preferable to any number of K lives, J will still be better than K. Since the new lives are worth living, J+ will still be at least as good as J. But this time it is undeniable that when we compare J+ with K, the latter will be preferable for people *de dicto*. Though each person would rather have one life in J+ than any number of lives in K, the smaller, privileged group's preference for J+ could not be a million times as strong.

"With one person," Griffin notes, "the aggregation is to a large extent already incorporated in the individual's global desires, and there is nothing comparable to that framework in the many-person case."[32] This does not pose a great problem so long as everybody has the same preferences, and everyone is treated the same. Suppose that every Italian prefers a single opera performance to any amount of beer. We can safely infer that if the government wants to benefit Italians and it is choosing between these options, then it should buy each an opera ticket. Yet nobody thinks that it would be preferable for Italians if we took away everyone's beer money so that one more could attend the opera. Even if each has a lexical preference for opera, the collective desire of millions for beer will greatly outweigh one individual's desire for *Don Giovanni.*

Analogously, if a single life in J is preferable for any person to any number of lives in K, we can infer that J is preferable for people *de dicto*. We cannot say the same of J+ compared with

K. If we choose J+, half the population will gain and half will lose. Now, employing the Lewis method really would mean "the conflation of all desires into one system of desire."

### III. The Suspension of Addition

The foregoing argument rests on five plausible premises:

28. If it would be preferable to live any single life in one world than to live any combination of lives in another, then the first world is preferable for people *de dicto*.

29. If two worlds contain the same number of satisfactory and equally good lives, and the second world contains additional lives that are unsatisfactory, but still all worth living, then the second world is at least as desirable for people *de dicto*.

30. If in two worlds with the same population the second world has greater utility, then the latter world is preferable for people *de dicto*.

31. *Ceteris paribus,* a world that is as desirable for people *de dicto* is as good *simpliciter,* and a world that is preferable for people *de dicto* is better *simpliciter*.

32. Other considerations do not require us to revise our assessments of comparative goodness in the cases examined.

When we compare J with K, the conditions in premises 28, 31 and 32 are met: J is lexically better. When we compare J with J+, the conditions in premises 29, 31 and 32 obtain, and we judge J+ to be as good as or better than J. When we compare J+ with K, premises 30, 31 and 32 speak in favor of K. The result is an intransitive value judgment: $J \leq J+ <K <J$.

That transitivity has failed is not surprising. The judgment that J is lexically better than K involves the "suspension of addition." We hold that though the K-lives have value, J is preferable for people *de dicto,* because no number of K-lives could be as desirable as a single one in J. The utility in K effectively does not count at all. Comparing J+ and K, on the other

hand, we do not suspend addition. We can't say that J+ is preferable for people *de dicto,* because for half the population, J+ would be the less desirable choice. When one comparison factors in the utility of the K-lives, and the other suspends addition, it is natural that we get intransitive results.

Many philosophers believe that intransitivity of value is impossible, or an incoherent concept. They must challenge one of the foregoing premises, or argue that their conditions have not been met. One might deny that we can appreciate the value of a very long life or series of inferior lives,[33] or question whether a single J-life that is clearly satisfactory would be preferable to any number of K-lives that are dubiously so. On this view, it would be imprudent to choose J over K, and the condition in premise 28 has not been met. One might contest premise 31's claim that what is preferable for people *de dicto* is, *ceteris paribus,* better *simpliciter*. If, as R. M. Hare argues, we should consider what will be retrospectively desirable for all possible people *de re*, then K will be preferable to J. Finally, one might try to argue that premise 29, 30 or 32 is false, and that J is better than J+, or that J+ is better than K.

It is probably the first line of argument that has the best chance of success. The claim that J is preferable to K is based on the assumption that no amount of additional utility could be worth trading away a satisfactory life. The difference in the quality of the lives in J and K might, however, be slight. We could stipulate, following a suggestion by Gustaf Arrhenius, that the lives in each world are identical, except that each life in K contains an additional pricked thumb.[34] Could we seriously consider it rational to prefer a single life in J to *any* number of worthwhile lives in K simply in order to avoid a once-in-a-lifetime thumb prick? Dorsey notes that a pricked thumb could cost an agent a "deliberative project," and thus have a big impact on her life.[35] Certainly we can imagine thumb pricks that would do this: pricking one's thumb

while defusing a bomb could bring ruin. But to meet this objection we need only stipulate that the additional thumb prick in K has no effect beside brief and mild pain. How can we reasonably believe that such a thumb prick could make all the difference between a successful life and a failed one?

We can't believe it—but then, the boundary between satisfactory and unsatisfactory lives is vague. Given the imprecision of our concept of a "satisfactory life," it would be surprising if it were not. Surely there are clearly satisfactory lives, clearly unsatisfactory lives, and a vague zone in between. J might be the last world in the sequence in which lives were *unambiguously* satisfactory, while K might be the first of a number of worlds in which it was *unclear* whether life was satisfactory. How many worlds fall into this zone will depend on how fine we make the gradations between lives; if the difference is at the level of a single thumb prick, there might be very many. The result is a sorites paradox. It is hard to believe that one thumb prick could possibly make a significant difference, just as a dieter might doubt that a single chocolate bar could make a difference between being thin and being fat.[36] Nevertheless, if we continue to add and redistribute we will move from J, where life is clearly satisfactory, to a world further down the alphabet where it is clearly not.

Such cases involve one value—pleasure from chocolate, utility—that can be measured in discrete units, and another, lexically preferred value—being thin, having a satisfactory life—whose boundary is vague. The temptation, as Christopher Knapp has pointed out, is to focus on the fact that taking another chocolate or moving to K leads to measurable improvement on the first spectrum of value, while never leading to unambiguous deterioration on the second.[37] Nevertheless, someone who gives priority to preserving his figure would do well to put aside the chocolates at a point while he is unambiguously thin—thin, to put it in the language of

supervaluationism, on all sharpenings of thinness. For the same reason, a rational agent should prefer a single life in J that is clearly satisfactory to any number of lives in K that may or may not be so. Since it may be unclear exactly where the vague zone begins, and her overriding goal is ensure a good enough life, she might also be wise to leave herself a margin of safety—stopping, perhaps, at H.[38] The important point is that we can defensibly hold that lives in J are lexically preferable to lives further down the spectrum. An ideal observer seeking the world most valuable for people *de dicto* would stop at the latest by J, even if the distinction between J and K is slight.

Some will insist that transitivity cannot fail in this way, since the transitivity of betterness is analytic.[39] Suppose, however, that we understand the best outcome as "the outcome that, from an impartial point of view, everyone would have most reason to want, or to hope will come about."[40] If the alternative with which an option are being compared can justifiably influence our assessment of its value, then we may, as Stuart Rachels and Larry Temkin have argued, arrive at intransitive value judgments.[41] Rachels and Temkin maintain that this could be because different factors could be relevant for making different comparisons, or the same factors might matter in different ways. I too am arguing that the choice of alternatives can affect the goodness of outcomes, but in a different fashion: by determining the populations *de re* whose interests must be taken into account. If the comparison is between J and K, we are choosing between populations of different sizes, and the best choice for people *de dicto* will be the lexically superior J. The choice between J+ and K, in contrast, requires us to compare outcomes for the same population, and what is best for people *de dicto* is the utility-maximizing K.

If transitivity of value can fail here, then it might fail in other contexts. At the same time, the combination of factors through which intransitivity has arisen—lexical superiority and

comparisons between outcomes for people *de dicto* that involve different populations *de re*—is

unusual. It may be peculiar to cases involving the creation of people. Those who worry about

the implications of intransitivity can hope that the mere addition paradox is *sui generis,* and that

"better than" is transitive in most, or even in all other cases.

## IV. Conclusion

The foregoing analysis has offered a new explanation of how the mere addition paradox

arises. If we deny the repugnant conclusion, we must accept that some lives can be lexically

superior to others. The most plausible explanation is that the former are better *as a whole*,

however many discrete utility units the latter ones may contain. Such a judgment entails

"suspending addition": denying that inferior sources of utility, however numerous, can ever add

up to a greater value than superior ones. Such lexical judgments are defensible when weighing

sources of goodness within individual lives, or deciding the optimal number of people to create.

There is no corresponding justification for suspending addition when weighing the interests of

some people against those of *others.* Though creating people with poor lives has minimal value,

*helping* people with poor lives has great value indeed. This explains our compelling intuition that

it is better to redistribute to the worse-off group.

This paper has offered a new explanation for the mere addition paradox; it cannot claim

truly to have solved it. That the relation "better than" could be intransitive remains deeply

paradoxical. Yet the conclusion seems unavoidable so long as we think it justified to suspend

addition when comparing some worlds, but not when comparing others. Future efforts to solve

the paradox should start from a recognition of this point.

17

12 Derek Parfit, "Overpopulation and the Quality of Life," in Peter Singer, ed., *Applied Ethics* (Oxford: Oxford University Press: 1986), pp. 145-64; Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1987).

[2] Michael Huemer, "In Defence of Repugnance," *Mind,* CXVII, 468 (October 2008): 899-933, at p. 923.

[3] Jefferson McMahan, "Problems of Population Theory," *Ethics,* XCII, 1 (October 1981): 96-127, at pp. 121-22.

[4] Parfit, "Overpopulation," pp. 160-61; Parfit, *Reasons and Persons,* 498.

[5] Parfit also suggests that impersonal perfectionist considerations could lead us to judge Z worse ("Overpopulation," pp. 163-64).

[6] Noah M. Lemos, "Higher Goods and the Myth of Tithonus," THIS JOURNAL, XC, 9 (September 1993): 482-96, at p. 489.

[7] Mozaffar Qizilbash, "Transitivity and Vagueness," *Economics and Philosophy,* XXI, 1 (April 2005): 109-31, at 128-29; Dale Dorsey, "Headaches, Lives and Value," *Utilitas*, XXI, 1 (March 2009): 36-58.

[8] James Griffin, *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford: Clarendon Press, 1990), p. 35; see also Clarence Irving Lewis, *An Analysis of Knowledge and Valuation* (LaSalle, IL: Open Court, 1946), chapter 16.

[9] Parfit, *Reasons and Persons,* pp. 493-96, quoted passage at p. 493.

[10] Griffin, *Well-Being,* pp. 8, 15, 34-36, 85-89, 340 n. 27.

[11] Parfit, "Overpopulation," p. 161; Parfit, *Reasons and Persons,* 493-99.

[12] For a defense, see Lemos, "Higher Goods."

[13] Parfit, "Overpopulation," p. 161.

[14] Lewis, *Analysis,* 546-47, 550-51.

[15] John Rawls, *A Theory of Justice: Revised Edition* (Cambridge, MA: Belknap Press, 1999), 163-64, and 164 n. 37.

[16] Jesper Ryberg, "Parfit's Repugnant Conclusion," *Philosophical Quarterly,* XLVI, 183 (April 1996): 202-13, at p. 205.

[17] See Gustaf Arrhenius, 'Future Generations: A Challenge for Moral Theories' (Ph.D dissertation, Uppsala University, 2000), p. 132; John Broome, *Weighing Lives* (Oxford: Oxford University Press, 2004), p. 74.

[18] R. M. Hare, "Abortion and the Golden Rule," *Philosophy & Public Affairs*, IV, 3 (Spring 1975): 201-22; Hare, "Possible People," in Hare, *Essays on Bioethics* (Oxford: Clarendon Press, 1993), p. 67-83.

[19] See Bettina Schöne-Seifert, "Zum moralischen Status potentieller Personen," in Christoph Fehige and Georg Meggle, eds., *Zum moralischen Denken* (Frankfurt-am-Main, Suhrkamp, 1995), vol. 2, pp. 210-24, at p. 220. Hare argues that in reality creating and maintaining World Z would involve so much disutility as to make the choice suboptimal ("Possible People," p. 68ff.) In the mere addition paradox, however, it is *stipulated* that World Z has greater utility. Surely we would still think Z a poor choice even if we could create it and preserve it simply by waving a wand. Hare, who set little store by philosophical intuitions, might have accepted that Z would be best in this case. That is a bullet that I and many others are unwilling to bite.

[20] For criticism, see T. G. Roupas, "The Value of Life," *Philosophy & Public Affairs,* VII, 2 (Winter 1978): 154-83, and Michael Lockwood, "Hare on Potentiality: A Rejoinder," in Nick

Fotion and Jan C. Heller, eds., *Contingent Future Persons. On the Ethics of Deciding Who Will Live, or Not, in the Future* (Dordrecht: Kluwer, 1997), 19-26. Even a recent defense of Hare's position denies that possible people have claims on us to be brought into existence, instead attributing our obligation to create new people to our "general duty of beneficence" (Oliver Hallich, "Das Argument der Existenzverhinderung in der Abtreibungsdebatte," *Zeitschrift für philosophische Forschung,* CXIV, 2 [June 2010]: 216-38, at p. 219.)

[21] Cf. Caspar Hare, "Voices from Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?" *Ethics,* CXVII, 3 (April 2007): 498-523.

[22] Larry S. Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (New York: Oxford University Press, 2012). So might differences of welfare among animals. Here I assume that non-human welfare levels are equal in all the worlds under comparison.

[23] Griffin, *Well-Being,* p. 340 n. 27.

[24] Temkin, *Rethinking the Good,* pp. 384, 394-97.

[25] Thomas Hurka, "Value and Population Size," *Ethics,* XCIII, 3 (April 1983): 496-507, at p. 506.

[26] Huemer, "In Defence of Repugnance," p. 904.

[27] Parfit, *Reasons and Persons,* p. 425.

[28] Cf, David Boonin-Vail, "Don't Stop Thinking About Tomorrow: Two Paradoxes About Duties to Future Generations," *Philosophy & Public Affairs,* XXV, 4 (Autumn 1996): 267-307, at pp. 274-75.

[29] Tim Mulgan, *Future People: A Moderate Consequentialist Account of Our Obligations to Future Generations* (Oxford: Clarendon Press, 2006), p. 68.

[30] Parfit, *Reasons and Persons,* pp. 437-38.

[31] Parfit, "Overpopulation," pp. 156-60.

[32] Griffin, *Well-Being,* p. 148.

[33] Huemer, "In Defence of Repugnance," pp. 908-10.

[34] Gustaf Arrhenius, "Superiority in Value," *Philosophical Studies*, CXXIII, 1/2 (March 2005): 97-114, at p. 107.

[35] Dorsey, "Headaches," p. 52.

[36] Qizilbash, "Transitivity and Vagueness," pp. 128-29; Christopher Knapp, "Trading Quality for Quantity," *Journal of Philosophical Research*, XXXII (2007): 211-33, at pp. 228-29; A. Faik Kurtulmus, "Uncertainty behind the Veil of Ignorance," *Utilitas*, XXIV, 1 (March 2012): 41-62.

[37] Knapp, "Trading Quality for Quantity."

[38] Cf. Kurtulmus, "Uncertainty," p. 60.

[39] John Broome, "Reply to Vallentyne," *Philosophy and Phenomenological Research,* LXXVIII, 3 (May 2009): 747-52.

[40] Derek Parfit, *On What Matters* (Oxford: Oxford University Press, 2011), vol. 1, p. 372; also Temkin, *Rethinking the Good,* pp. 13-15.

[41] Stuart Rachels, "A Set of Solutions to Parfit's Problems," *Noûs*, XXXV, 2 (June 2001): 214-38; Temkin, *Rethinking the Good.*