Rethinking the Good, Moral Ideals and the Nature of Practical Reasoning

Larry S. Temkin

Introduction

In Part 4, ‘Future Generations’, Derek Parfit raises many perplexing questions. Unfortunately, some find Parfit’s ingenious arguments more irksome than illuminating, while others think them little more than delightful puzzles. Alas, it is not uncommon for Parfit’s harshest critics to ask dismissively, ‘What do boxes have to do with ethics?’ – as if Parfit’s results are mainly clever artefacts stemming from a peculiar use of abstract diagrams. On the other hand, even among those regarding Part 4 as genuinely insightful, ‘Future Generations’ is often regarded as the book’s least significant part, relevant mainly to consequentialists concerned with determining the best principle of beneficence – Parfit’s Theory X – and to those (few) environmentalists and others seriously worried about future generations.

Such reactions to Part 4 are thoroughly misguided. Parfit’s arguments are original, and strike at the core of some of our deepest beliefs. Indeed, I believe Part 4 is one of the richest, most profound works in contemporary philosophy. My reasons for this will become clear as the essay progresses. It is divided into two main parts. In Part I, I discuss the Repugnant Conclusion, in part II the Mere Addition Paradox. I will argue that the Repugnant Conclusion forces us to reject the most common view of utility and its relation to our all-things-considered judgements, to develop a new way of understanding the role that moral ideals play in relation to each other, and to recognize that certain moral ideals may have to share certain formal or structural features in ways that have significant and previously unrecognized implications. Next, I will argue that the Mere Addition Paradox raises profound questions about the nature of moral ideals, questions that ultimately challenge the consistency and intelligibility of some fundamental assumptions about moral and practical reasoning. If the arguments I present here are right, Part 4 will alter our understanding of the nature of moral ideals and their role in our all-things-considered judgements in ways that have far-reaching practical and theoretical implications.

First, a few preliminary comments. The reader will notice that I have little to say about future generations. This is no accident. I believe Part 4 is misleadingly titled. Though Parfit is surely right that we can now affect the course of future generations as never before, and that therefore consideration of future generations must now play a prominent role in our moral reflections, I think his arguments’ implications about future generations are secondary, in importance, to their fundamental implications for moral theory itself. Correspondingly, I focus on the latter rather than the former.

Analogously, at some points the reader may wonder whether I have been faithful to Parfit’s views. I believe I have, but I confess I am much less concerned about assessing Parfit’s arguments as he presents them than about critically assessing what we can learn from those arguments. This is one reason why, for the most part, I ignore Parfit’s errors. It is his insights that I am interested in.

This essay is long, but in many respects it is not long enough. Besides leaving many questions open, I ignore many important arguments. So, for example, time and space considerations prevent me from addressing the Non-Identity Problem, the Absurd Conclusion, the Asymmetry and Hell Three, among others. These are rich arguments that would amply repay careful study. Moreover, there are interesting connections between these arguments and those presented here. Unfortunately, I cannot pursue these issues in this essay.

Part I. The Repugnant Conclusion

A. In chapter 17 of Reasons and Persons Parfit presents

The Repugnant Conclusion: 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. (p. 388)

Figure 13.1 helps to illustrate the Repugnant Conclusion. According to the Repugnant Conclusion, if only there are enough people in Z, Z will
be better than A, and this is so no matter how large A is, and even if those in A are very well off while those in Z have lives barely worth living. As Parfit writes, 'this conclusion [is] very hard to accept' (p. 388).

The Repugnant Conclusion has significant implications. In supporting this claim, I shall argue that the Repugnant Conclusion challenges what is perhaps the standard, or naïve, way of thinking about utility and its relation to our all-things-considered judgements. I shall then indicate how our views about utility suggests a new model for understanding moral ideals. Finally, I shall illustrate that, depending on their role in relation to our all-things-considered judgements, some ideals must share certain formal, or structural, features, and that this has important, though largely unrecognized, implications.

My discussion assumes, with Parfit, that the Repugnant Conclusion is genuinely repugnant and should be rejected. Also, throughout this essay I shall assume that there are only four principles, or ideals, relevant to comparing Parfit’s alternatives: utility (U), equality (E), maximin (M), and perfectionism (P). So, for example, I assume there is nothing to choose between Parfit’s alternatives in terms of other moral factors such as virtue, duty, liberty or rights. This assumption is, of course, controversial, both for what it includes and what it excludes, but it greatly facilitates my presentation and does not affect my central results.

B. I believe that many implicitly accept four basic assumptions:

1. Utility is intrinsically valuable, even if it is not the only thing that is.
2. In so far as one cares about utility, one should care about total utility.
3. How good a situation is regarding utility is a simple additive function of how much utility sentient beings have in that situation.
4. How good a situation is, all things considered, is an additive function of how good it is regarding each ideal; so, in so far as a situation gets better regarding utility, it will, to that extent, be getting better, all things considered.

Not everyone accepts (1)–(4). Average utilitarians reject (2), strict Kantians reject (1) and (4), and an increasing number reject (3)’s and (4)’s additive assumptions. Nevertheless, (1)–(4) have great appeal. In fact, though rarely explicitly formulated, I believe that they represent perhaps the most natural and prevalent way of thinking about utility and its connection to our all-things-considered judgements.

Together (1)–(4) support the Repugnant Conclusion. After all, given (1)–(3), the addition of lives worth living will make an outcome better regarding utility. But then, if adding an extra person improves a situation at all – even in only one respect, and even if, because the person’s life is barely worth living, it is just a tiny bit – (3) and (4) imply that if only there are enough people in Z, it would be better than A. That is, if only Z is large enough, its all-things-considered score will be higher than A’s, as its score for utility will be sufficiently higher than A’s to outweigh the extent to which it has lower scores for perfectionism and maximin.*

The preceding argument makes two assumptions. First, it assumes that Z’s scores for maximin and perfectionism are not inversely related to Z’s size; that is, that merely increasing Z’s population does not worsen Z regarding M and P. Second, it assumes the possibility of something like a numerical model for judging outcomes, according to which each outcome will merit a (rough numerical) ‘score’ representing how good that situation is, all things considered, where that score will be an additive function of other (rough) ‘scores’ for each relevant factor or ideal (i.e. U, E, M and P). Though I think the second assumption’s details are enormously complicated, I think both assumptions are intuitively plausible, and likely to be granted by most advocates of (1)–(4).

The foregoing results are important. They reveal that, to avoid the Repugnant Conclusion, we need an alternative model to (1)–(4) for understanding utility and the role it plays in our all-things-considered judgements. Let us next consider one such model.
C. Reflecting on the Repugnant Conclusion, most of us are reminded that we are not strict utilitarians. While utility matters a lot, it is not all that matters. But we also learn a further, deeper lesson. We learn that mere increases in the quantity of utility are not sufficient (or so we think) to outweigh significant losses regarding other ideals. This is why Z is worse than A. Though better regarding utility, it is not better regarding equality, and it is much worse regarding other ideals, like perfectionism and maximin.

The importance of the further lesson cannot be exaggerated. It suggests that there may be an upper limit to how good a situation can be regarding utility. Moreover, if this is so for utility, it is probably also so for the other ideals, and hence there may also be an upper limit to how good a situation can be, all things considered.

Consider an analogy from sports. In a gymnastics competition the best all-round gymnast is decided by adding together each person’s score in each event. For women, there are four events, each with a maximum score of ten; so the maximum total score one can attain is 40. While theoretically one can approach, and even attain, a perfect performance, one can never exceed the maximum score. So, to be the best all-round gymnast, it is not enough to near, or even attain, perfection in the floor exercises. One must also excel on the balance beam, the vault, and the uneven parallel bars.

On reflection, it seems that something like this may also hold for our all-things-considered judgements. Perhaps in comparing situations, all things considered, we must compare them in terms of each ideal, where there is a maximum score they can get for each ideal, determined by how much the ideals matter vis-à-vis each other. Suppose, for example, that in assessing outcomes, U, E, M and P were equally important and all we cared about. Then we might assign numbers to situations such that the highest score a situation could get for each ideal would be, say, 100, and the ‘perfect’ situation would score 400. Similarly, to note a more complicated example, if we thought that P was twice as important as E and M, and half again as important as U, then we might assign numbers to situations such that the highest score a situation could get would be, say, 150 for P, 100 for U, and 75 each for E and M. Clearly, the highest score for each ideal sets an upper limit to the score that situations can receive for that ideal. Situations will receive the highest score for an ideal if they are perfect regarding that ideal; to the extent they are less than perfect, they will receive appropriately lower scores.

Let us call such a position the gymnastics model for moral ideals. The gymnastics model expresses the views, noted previously, that there is an upper limit to how good situations can be regarding each ideal, and also an upper limit to how good situations can be, all things considered.

Now a ‘perfect’ situation regarding U – one that could not be improved on and would merit the highest possible score for U – would have infinite utility. Correspondingly, no finite world will be perfect regarding utility. So, on the gymnastics model, a finite world may approach the maximum score for utility, but it can never attain that score or go beyond it. Note that, on the gymnastics model, there is no limit to how much utility can be added to a situation. In addition, for any finite situation, it will always be possible to improve that situation regarding utility, even if only slightly. However, as there is an upper limit on how good a situation can be regarding utility, there is also an upper limit on how much a situation can be improved – that is, be made normatively better – regarding utility. So, after a point, mere increases in the amount of utility will not substantially increase the value of a situation, even regarding utility.

It follows that, on the gymnastics model, mere increases in utility alone, however great, will not always be sufficient to outweigh significant losses in other respects.

To illustrate this position, let us again make the simplifying assumption that each ideal matters equally, and let us (arbitrarily) assign scores to A and Z of up to 100 for U, E, M and P. A is perfect regarding E, and (by hypothesis) very good for U, M and P. So let us suppose that E = 100, U = 80, M = 80 and P = 80. A would then have an all-things-considered score of 340. Z is perfect regarding E, and, we may suppose, nearly perfect regarding U, but it is much worse regarding M and P. For Z, then, let us suppose E = 100 and U = 99, but M = 20, and P = 20. Z would then have an all-things-considered score of 239 – 100 points lower than A’s on a scale of only 400! One can see, then, how such a model can account for the genuine repugnancy of the Repugnant Conclusion. Of course, the extent of the repugnancy will depend on how A and Z in fact compare for U, M and P – or any other relevant ideals – as well as how much those ideals matter to our all-things-considered judgements.

There is a logical limit to how much one can reduce a situation’s inequality, but no logical limit to how much utility obtains. Because of this, it is natural to suppose that there is a limit to how much one can improve a situation regarding equality, but none regarding utility. Nevertheless, this supposition seems mistaken. However much more we may care about one ideal relative to another, they play similar roles in our all-things-considered judgements. Thus, on reflection, most believe that, other things equal, transforming a situation into one perfect, or nearly perfect, regarding E, U or P would not be better all things considered if it involved significant losses regarding U and P, M and P, or E and M, respectively.

Let us contrast the gymnastics model for moral ideals with the ‘standard’ model noted in section B. In terms of (1)–(4), the gymnastics model
can capture (2) — in so far as one cares about utility, one should care about total utility. It can also capture (4) — how good a situation is, all things considered, is an additive function of how good it is regarding each ideal; so, in so far as a situation gets better regarding utility, it will, to that extent, be getting better, all things considered. But it denies (3) — how good a situation is regarding utility is a simple additive function of how much utility sentient beings have in that situation. Thus, while it grants that doubling the amount of utility in a situation improves it regarding utility, it denies that it necessarily does this by a factor of two. Correspondingly, it denies (1) — utility is intrinsically valuable — in the sense that each unit of utility has intrinsic value of a constant amount which increases a situation’s objective value in direct proportion to that amount. On the other hand, the model is compatible with utility being intrinsically valuable in the sense that it can regard utility as a significant factor in all-things-considered judgements beyond the extent it promotes other ideals.

Faced with paradoxes involving future generations, and hoping to avoid the Repugnant Conclusion, Parfit and others have suggested that there may be no respect in which the ‘mere addition’ of extra people to an already large and well-off population improves the situation. However, many find this position implausible. They think that the addition of extra people with lives worth living must always improve utility. The gymnastics model captures this belief, while avoiding the unpalatable conclusion that sheer increases in a situation’s amount of utility must always significantly improve it regarding utility and (hence) all things considered. To be sure, from the subjective perspective, each extra life worth living will be as important to its possessor as every other. But from the objective perspective, though each extra life will matter some, after a point, it will matter less and less.

There are problems with the gymnastics model, which I cannot pursue here. For example, while it may seem plausible to think that there is an upper limit to how good a situation can be regarding utility or equality, it seems much less plausible to think that there can be a lower limit to how bad a situation can be regarding utility or equality — as if, after a point, it does not really matter (much) any more whether extra people are excruciatingly tortured, or the gaps between the haves and have-nots increase still further. Yet, as Parfit has demonstrated in chapter 18, it is not clear that one can have upper limits without lower limits, at least without facing serious problems of coherence and consistency. Similarly, as we shall see in part II, if one adopts a gymnastics model, and combines it with what I call the intrinsic aspect view of moral ideals, one may have to accept the ‘absurd’ conclusion of Parfit’s ‘How Only France Survives’, according to which it might be better if everyone but the best-off died, even if all those who died had lives well worth living, and even if the lives of the best-off were worsened as a result. On the other hand, if, to avoid such implications, one combines the gymnastics model with what I have called the essentially comparative view of moral ideals, then, as we shall also see in part II, the threat of intransitivity looms for our all-things-considered judgements.

I mention the foregoing because I do not want to minimize the difficulties facing a gymnastics model for moral ideals. Still, the gymnastics model is hardly alone in facing grave difficulties, and, reflecting on the Repugnant Conclusion, it seems clear that what is perhaps the most natural and prevalent way of thinking about utility and its relation to our all-things-considered judgements must be rejected. Whether we should ultimately accept the gymnastics model, a variation of it (e.g. one that rejects (4)’s additive assumption as well as (3)’s), or pursue an entirely new alternative to (1)–(4), I must leave open for now.

In sum, the Repugnant Conclusion forces us to rethink the nature and role of our moral ideals in relation to each other and our all-things-considered judgements. I have suggested one direction our rethinking might go.

D. I have claimed that however much more we may care about one ideal relative to another, they play similar roles in our all-things-considered judgements. So, for example, just as there is an upper limit to how good a situation can be regarding equality, there may have to be an upper limit to how good a situation can be regarding utility. Let me next suggest further reason to hold that a plausible and coherent account of the role that moral ideals play in relation to each other and our all-things-considered judgements may require some ideals to share certain formal or structural features. Specifically, I shall argue that if numbers count regarding utility, they may also have to count regarding equality.

Consider figure 13.2. As drawn, A’s and B’s better- and worse-off
groups are at the same levels, but B's groups are twice as large as A's. Regarding inequality, many would judge A and B equivalent. As they might put it, since the pattern of inequality is identical in A and B, there is nothing to choose between them; the 'mere' fact that B is larger than A is irrelevant to how they compare regarding inequality.

The judgement that A and B are equivalent expresses the following view:

\[ PV: \text{Proportional variations in the number of better- and worse-off do not affect inequality.} \]

According to PV, size is not itself relevant to inequality. Size variations will only matter in so far as they affect inequality's pattern – for example, by altering the better- and worse-off's levels or the ratios between them.

PV is widely accepted, and has great plausibility. It is, for example, implied by each of the economists' statistical measures of inequality, and they have not, to my knowledge, been criticized for that feature. Yet, despite its appeal, and the fact that numerous considerations can be offered supporting it, ultimately PV may be incompatible with our views about other ideals and the role they play in our all-things-considered judgements.

Consider figure 13.3. As drawn, C represents an unequal society, D a perfectly equal society with less total utility. E and F are just like C and D, respectively, except they are twice as large. Looking at C and D, many would judge that D is better than C, all things considered. They would judge the (slight) loss of utility in moving from C to D regrettable, but outweighed by the gain in equality. Similarly, looking at E and F, many would judge that F is better than E, all things considered. Again, they would judge the loss of utility in moving from E to F regrettable, but outweighed by the gain in equality. More importantly, few, if any, would approve a redistribution between the better- and worse-off in C so as to bring about D, yet oppose a redistribution between the better- and worse-off in E so as to bring about F. That is, most would agree that if moving from C to D were desirable, then moving from E to F would also be desirable.

The foregoing claims may seem obvious and uninteresting, but they have important implications. Depending on how one thinks that moral ideals combine to yield all-things-considered judgements, they imply that if numbers count for utility, they must also count for equality.\(^1\)

Consider. Suppose one thinks that the move from C to D is only a slight improvement, all things considered, because the gain in equality is just barely enough to outweigh the attendant loss in utility. Then it looks as if F would be worse than E if numbers count for utility but not for equality. After all, E and F are twice as large as C and D. This means that the loss in utility in moving from E to F will be twice the loss in utility in moving from C to D, and hence, on the view in question, that the move from E to F will be worse than the move from C to D regarding utility. Yet, if numbers don’t count for equality, or, more specifically, if we accept PV, the view that proportional increases do not affect inequality, then the gain in equality in moving from F to E will be exactly the same as the gain in equality in moving from D to C. But then it looks as if the gain in equality in moving from F to E won’t be sufficient to outweigh the attendant loss in utility, since, by hypothesis, the gain in question is barely enough to outweigh a loss in utility which is only half as large. In sum, if moving from E to F is significantly worse than moving from C to D regarding utility, and no better regarding equality, then it is easy to see that, all things considered, moving from E to F could be undesirable even if moving from C to D were desirable. But this, of course, is contrary to the 'obvious and uninteresting' view noted earlier.\(^1\)

The foregoing result may be surprising. But, on reflection, it is not, I think, perplexing or disturbing. Suffice it to say, I accept the view that if moving from C to D were desirable, so too would be moving from E to F. I also accept the view that in an important sense the latter move would be worse than the former regarding utility. This leads me to believe that there is an important sense in which the latter move would be better than the former regarding equality. But, of course, this implies that E is worse than C regarding inequality, which in turn implies that there is reason to reject PV.\(^1\)

Some will insist that inequality is not bad in the same way disutility is. More generally, they will insist that equality and utility are different
kinds of moral ideals, and that the two ideals must be treated differently. So, for example, some will insist that while utility is essentially additive, equality is essentially distributive—that is, that while utilitarians are essentially concerned with how much there is, egalitarians are essentially concerned not with how much there is, but with the way, or pattern, in which what there is, is distributed. On this view, the fact that numbers count regarding utility should be completely irrelevant to whether numbers count regarding equality.

Such thinking is powerfully seductive, but as we have seen, it is not clear that our moral ideals can be fully and adequately characterized in isolation from each other. Specifically, a plausible and coherent account of the role that moral ideals play in relation to each other and our all-things-considered judgements may require at least some ideals to share certain formal or structural features. Thus, as the Repugnant Conclusion reveals, there may be a limit to how much improvement regarding one ideal can offset significant losses regarding other ideals, and this in turn suggests that, just as there may be an upper limit regarding how good a situation can be regarding equality, so there may be an upper limit regarding how good a situation can be regarding each of our moral ideals. So, too, if one holds that numbers count regarding utility, one may also have to hold that numbers count regarding equality.17

Part II. The Mere Addition Paradox

A. Arguably, ‘The Mere Addition Paradox’ is both the most ingenious and the most error-filled chapter in Reasons and Persons. Between them, the versions of the Mere Addition Paradox and the two examples, ‘How Only France Survives’ and ‘Hell Three’, raise a host of intriguing questions. Yet many of Parfit’s claims and arguments are at best controversial, and at worst mistaken. For example, against Parfit, it could be argued that in the second version of the Mere Addition Paradox, New A’s inequality is much worse than A+’s (see p. 434), and Parfit is surely wrong in contending that ‘all of the plausible views [of inequality] would agree that the inequality in New B is less bad than the inequality in New A’ (p. 435). Moreover, regarding the first version of the Mere Addition Paradox, I think Parfit is wrong in claiming that if one accepts that A+ is better than A, and B is better than A+, one ‘cannot avoid the Repugnant Conclusion’ (p. 430); and that he is also wrong in making similar claims regarding the second and third versions of the paradox (see pp. 436–7 and 441). The conclusions in question can be avoided if, for example, one adopts something like the gymnastics model for moral ideals suggested in part I.19 In addition, I think that sometimes Parfit’s reason-

ing is inconsistent, implicitly adopting one view of moral ideals in support of certain conclusions, but another, incompatible view of moral ideals in support of other conclusions.

I level the above charges, but shall not pursue them. This is not because I think Parfit’s mistakes—if they are mistakes—are uninteresting. To the contrary, in almost every case much can be gleaned from carefully considering where, and why, Parfit has gone wrong. Still, the shortcomings of Parfit’s arguments are insignificant compared to the importance of their insights and implications. So in the space remaining I shall focus on the latter.

Sadly, many have regarded the Mere Addition Paradox as an arcane esoteric puzzle involving hypothetical future generations and having little to do with the central problems of moral philosophy. Nothing could be further from the truth. Unfortunately, to do full justice to the issues raised by the Mere Addition Paradox would require at least a book. So my remarks must be sketchy, programmatic and open-ended. Still, I can perhaps give the reader some sense of why I think the Mere Addition Paradox is one of the most profound, though least well-understood, arguments in contemporary moral philosophy.

B. Though in some ways Parfit’s second and (especially) third versions of the paradox are more interesting and compelling, for our purposes it will be sufficient to consider a slightly simplified form of Parfit’s first version of the Mere Addition Paradox.20 Consider figure 13.4. Parfit claims that, all things considered, most would judge that B is worse than A. After all, by hypothesis, A is already a large population (say, ten billion), and everyone in A is better off than everyone in B. Parfit also