

Mere addition and the best of all possible worlds

STEPHEN GROVER

Department of Philosophy, Queens College, C.U.N.Y., Flushing, New York, NY 11367–1597

Abstract. The quantitative argument against the notion of a best possible world claims that, no matter how many worthwhile lives a world contains, another world contains more and is, other things being equal, better. Parfit's 'Mere Addition Paradox' suggests that defenders of this argument must accept his 'Repugnant Conclusion': that outcomes containing billions upon billions of lives barely worth living are better than outcomes containing fewer lives of higher quality. Several responses to the Paradox are discussed and rejected as either inadequate or unavailable in a theistic context. The quantitative argument fails if some world is such that addition to it is not possible, i.e., if it is intensively infinite, as Leibniz claimed. If the notion of such a world is incoherent, then no world is quantitatively best and the quantitative argument succeeds, but only at the cost of embracing the Repugnant Conclusion.

I

INTRODUCTION

In *The Existence of God*, Richard Swinburne neatly summarizes a simple argument against the coherence of the notion of a best among possible worlds:

... take any world *W*. Presumably the goodness of such a world ... will consist in part in it containing a finite or infinite number of conscious beings who will enjoy it. But if the enjoyment of the world by each is a valuable thing, surely a world with a few more conscious beings in it would be a yet more valuable world – for there would be no reason why the existence of the latter should detract from the enjoyment of the world by others – they could always be put some considerable distance away from others, so that there was no mutual interference. I conclude that it is not, for conceptual reasons, plausible to suppose that there could be a best of all possible worlds, and in consequence God could not have overriding reason to create one.¹

This quantitative argument should be distinguished from others that appeal to qualitative considerations. Perhaps any world containing conscious beings who enjoy it is bettered by a world containing the same beings enjoying their world even more. If there is no upper limit to enjoyment for these beings, it is implausible to suppose that there could be a best among possible worlds. If there is an upper limit to enjoyment for these beings, there may be worlds in which other kinds of beings are capable of higher levels of enjoyment, and still other worlds in which yet other beings enjoy themselves even more, and so on. Perhaps there is no limit to this way of improving on a given world.

¹ R. G. Swinburne *The Existence of God* (Oxford: Clarendon Press, 1979), 114.

These qualitative considerations are extremely interesting, but they are largely irrelevant to the quantitative argument except in so far as there is some level of enjoyment below which lives are not worth living at all. Adding beings whose lives fall below this level presumably makes a world worse rather than better, and God may have overriding reason not to create worlds that contain any such beings.

As presented by Swinburne, the quantitative argument against the coherence of the notion of a best possible world – I shall call it ‘the quantitative argument against BPW’ or just ‘the quantitative argument’ – applies whether the number of conscious beings in a world is finite or infinite. Infinite numbers are hard to think about; adding to such numbers is even harder to comprehend. So I shall begin with the finite case, leaving the infinite for later (section VI).

As it stands, the quantitative argument relies upon a claim that Derek Parfit has called into question, namely, that mere addition always makes an outcome better.² On Parfit’s definition there is ‘*Mere addition* when, in one of two outcomes, there exist extra people (1) who have lives worth living, (2) who affect no one else, and (3) whose existence does not involve social injustice’ (420). Parfit’s examples of mere addition typically involve outcomes in which the extra people are worse off than those in the outcome to which they are merely added. It is in this way that the ‘Mere Addition Paradox’ is generated. Parfit is also often concerned with the merits of the claim that mere addition does not make outcomes worse, rather than the stronger claim, endorsed by the quantitative argument against BPW, that it makes outcomes better. But there is enough overlap here for it to be worth asking whether Parfit’s arguments, and the criticisms and responses that they have generated, can help in assessing the strength of this simple and appealing argument against the coherence of the notion of a best of all possible worlds.

II

THE MERE ADDITION PARADOX

Chapter 19 of Parfit’s *Reasons and Persons*, in which the idea of mere addition is introduced and examined, has been described as ‘both the most ingenious and the most error-filled chapter’ in the book.³ All three versions of Parfit’s Mere Addition Paradox have been criticized. Some of these criticisms pro-

² D. Parfit *Reasons and Persons* (Oxford: Oxford University Press, 1984, reprinted with corrections, 1984, 1985). References in the text are to this work. The Mere Addition Paradox is first stated in Parfit ‘Future generations: further problems’, *Philosophy and Public Affairs*, 11 (1982). Parfit ‘Overpopulation and the quality of life’ in P. Singer (ed.) *Applied Ethics* (Oxford: Oxford University Press, 1986) contains a summary of the argument in *Reasons and Persons* and some additional discussion.

³ L. S. Temkin ‘Rethinking the good: moral ideals and the nature of practical reasoning’ in J. Dancy (ed.) *Reading Parfit* (Oxford: Blackwell, 1997), 300.

vide avenues of escape from the Paradox for those who hold that mere addition does not make outcomes worse, but offer no comfort to those who, in the manner of the quantitative argument against BPW, hold that it makes outcomes better. Other criticisms are more wide-ranging, and some call into question the whole enquiry in which Parfit is engaged. For example, Robert Adams thinks that ‘it is a very dubious enterprise to assign comparative values to outcomes in abstraction from our moral assessment of processes by which they might arise’.⁴ This may be so, but it tells as much against the quantitative argument as against Parfit. In what follows, I assume that assigning values to outcomes or possible worlds is a worthwhile enterprise, however dubious the intuitions on which it relies. On one model of creation, this is just what God does when She surveys the possible worlds and selects one from among them to create.

The first version of Parfit’s Mere Addition Paradox can be summarized as follows: suppose we merely add to the ten billion lives, all of very high quality, that are lived in outcome A, another ten billion lives that, whilst they are well worth living, are of much lower quality than all the lives lived in A. This is outcome A+. If mere addition never makes outcomes worse, A+ is no worse than A. Now imagine outcome B, in which twenty billion lives are lived, all at a level lower than the lives lived in A but higher than the average level of the lives lived in A+. B is in some ways better than A+ : in B (as in A) there is no inequality, and in a shift from A+ to B the worse-off group gains more than the better-off group loses. If B is better than A+, and A+ is no worse than A, then B is also no worse than A. But this implies that a significant lowering of the *quality* of all the lives that are lived does not make an outcome worse, provided that the increase in the *quantity* of lives is substantial enough. Many people reject this claim: they think B is worse than A. If we agree, but also think A+ is no worse than A, and B is better than A+, we face the Mere Addition Paradox.

The Paradox reiterates: merely add another twenty billion worthwhile lives to B, again stipulating that the extra people all live lives of lower quality than all the lives lived in B, and we get B+. If mere addition never makes outcomes worse, B+ is no worse than B. Now construct outcome C, in which forty billion lives are lived, all lower in quality than the lives lived in B, but higher than the average level of the lives lived in B+. If C is better than B+, C is no worse than B. The procedure repeats until we reach Z, an outcome in which billions upon billions of lives are lived, all barely worth living.

That Z is better than A is Parfit’s ‘Repugnant Conclusion’ (388), but it is not forced upon us by this version of the Paradox.⁵ Even if B is better than

⁴ R. M. Adams ‘Should ethics be more impersonal?’, *Philosophical Review*, 98 (1989), 484.

⁵ The ‘New Repugnant Conclusion’, in which Z is replaced by New Z – billions upon billions of lives not much above the Bad Level – is forced upon us by the second version of the Paradox, according to Parfit (433–438). I do not discuss this version of the Paradox. I mention the Bad Level in section VII.

A+, and A+ no worse than A, B may not be better than A: ‘not worse than’ may not imply ‘at least as good as’ (430–431). If B is no worse than A, and C no worse than B, and so on down the alphabet, it might seem that we are forced to the only marginally less repugnant conclusion that Z is no worse than A. But this presumes the transitivity of ‘not worse than’, which may be unjustified: perhaps Z is no worse than Y, and Y no worse than X, but Z is worse than X (431–432). However, these complexities are irrelevant if, as the quantitative argument suggests, mere addition makes outcomes better: if A+ is better than A, and B better than A+, then surely B is better than A (430). ‘Better than’ is transitive even if ‘not worse than’ is not.

To the extent that we are inclined to accept the quantitative argument against BPW, we must explain how we intend to deal with the Mere Addition Paradox. This is not because anything in the Paradox implies that there is a best possible world. If Z is better than A, it will still be capable of improvement by the mere addition of even more lives barely worth living. But the quantitative argument relies upon the claim that mere addition makes outcomes better. If defending this claim involves endorsing the Repugnant Conclusion, we may prefer to abandon the quantitative argument instead.

III

THE VALUELESS LEVEL

We escape the Mere Addition Paradox if we agree that B is better than A, and this makes a great deal of sense. If mere addition makes outcomes better then it ought also to compensate for other ways in which outcomes are worse.⁶ Why shouldn’t the fact that there are twice as many people living outweigh the fact that these people are worse off? But this starts the descent towards the Repugnant Conclusion, and unless we are willing to embrace that conclusion we will have to come up with a reason for halting the descent somewhere between B and Z.

One way of doing this is to qualify the claim that mere addition makes an outcome better so that it holds only when the quality of lives added is above a certain level. This is an appeal to what Parfit calls ‘the Valueless Level’ (412–413).⁷ It halts the descent towards the Repugnant Conclusion by preventing further trade-offs between the quantity of lives lived and the quality of those lives: no matter how many lives below the Valueless Level

⁶ Temkin, 312–313.

⁷ As stated by Parfit, appeal to the Valueless Level is more complicated than suggested here, for it also involves an upper limit to the value of quantity: when the number of lives is below the limit, every worthwhile life makes an outcome better; once the number of lives rises above the limit, only lives above the Valueless Level count. These complexities do not matter here: we can assume that all the outcomes under discussion are above the upper limit. I discuss the issue of an upper limit to the value of quantity in section V.

are merely added to an outcome, this cannot compensate for a decline in the quality of lives in the outcome to which they are merely added.

By appealing to the Valueless Level, we avoid the Repugnant Conclusion: reiteration of the Mere Addition Paradox merely drives us down to that outcome in which many billions of lives are lived, all at the level below which mere addition ceases to make outcomes better. If the Valueless Level is close to the level where lives cease to be worth living, the conclusion we are forced to is only marginally less repugnant. If the Valueless Level is much higher, however, the appeal of the quantitative argument against BPW is much diminished. The quantitative argument begins from the insight that the lives lived by conscious beings are part of what makes the world good, provided that each life is on the whole enjoyable. It is then argued that however many such lives a world contains, there is another and better world that contains a few more. But if a world contains very many lives that are worth living, and if many of these lives fall below the Valueless Level, many lives will not contribute to the goodness of the world in which they are lived. The defender of the quantitative argument must adopt an ambivalent attitude to worthwhile lives that fall below the Valueless Level: even though these lives are of value to the creatures who live them, more such lives would not make for a better world.

One way of seeing the difficulty here is by running Swinburne's argument in reverse: take some world *W*, the goodness of which consists in part in its containing a large number of conscious beings who enjoy it, and suppose that of these enjoyable lives some fall above and some below the Valueless Level. On the view we are considering, a world just like *W* but with more lives below the Valueless Level will be no better than *W*. Presumably, then, a world just like *W* but with fewer such lives will be no worse. So in what sense was the goodness of *W* constituted, in part, by the enjoyable lives these beings led? Only lives above the Valueless Level count. The higher the Valueless Level is set, the lower the proportion of lives that contribute to the goodness of the world in which they are lived. The lower it is set, the more repugnant the outcomes we are driven to acknowledge as better.

Appeal to the Valueless Level is problematic in another way. The point of the appeal is to prevent trade-offs between the quantity and the quality of lives lived. But it is easy to imagine scenarios in which this leads to absurd results. This is particularly so if there is no upper limit to the disvalue of suffering, for no increase in the number of lives that are just below the Valueless Level will outweigh even the smallest increase in the number of lives that contain so much suffering as to be not worth living.⁸ These absurdities diminish as the Valueless Level is lowered, but then the repugnance of the outcomes judged better intensifies.

⁸ I take this issue up again in section V.

These criticisms of the appeal to the Valueless Level are not decisive, but anyone making such an appeal must justify the claim that there is such a cut-off and give us some idea where it falls. The actual world contains lives that vary dramatically in quality. Some lives are of such poor quality and contain so much suffering that they are not worth living; others are of very high quality. Most lives lie between these extremes. If the Valueless Level is high, not many of the lives lived in the actual world are such that more of them would make the world better. This suggests that the goodness of the world is mainly a matter of it containing a few lives of high quality, with other lives not counting for much at all. This is not a view with which many theists could feel comfortable. But if the Valueless Level is set low enough to encompass most of the lives lived in the actual world, we face again a descent towards the Repugnant Conclusion. It seems worthwhile, therefore, to explore other avenues of escape from the Mere Addition Paradox.

IV

IS B BETTER THAN A + ?

The next most obvious strategy is to question whether B is better than A + . Comparing A + with A the inequality in A + seems irrelevant because the lives of the worse off group are merely added, and this can hardly make A + worse than A. But when A + is compared with B the inequality in A + is pointed to as one way in which it is worse. Equally, the causal distance that renders the addition of the worse off group in A + mere is ignored when it is argued that in a shift from A + to B the worse off group gains more than the better off group loses.

Parfit emphasizes that no deliberate redistribution is involved in a shift from A + to B. He offers as an aid to comparison another outcome, Divided B, in which a population the same size as B's is separated into two halves that, like the groups in A + , cannot communicate. A + is then imagined to change into Divided B as the result of natural events, perhaps over some considerable period. That this change is for the better Parfit thinks is clear: the worse off group gain more than the better off group lose, so the change satisfies maximin principles that disallow inequalities unless they benefit the worst off, and also satisfies straightforward principles of equality. Nor is this a case where equality conflicts with beneficence: both judge Divided B better than A + . Parfit also thinks that Divided B is clearly as good as B. Thus we get the following evaluations: A + is no worse than A; Divided B is better than A + ; B is as good as Divided B. It follows that B is better than A + , and hence no worse than A (425–426).

But here there are problems, pointed out by Robert Adams.⁹ A shift from A + to Divided B lowers the quality of life of the people in one group and

⁹ Adams, 478–80.

raises that of the people in the other group by a greater amount. Nothing similar is involved in a shift from $A+$ to B because of the separation of the two groups in the former: such a change could come about only by one group going extinct while the other doubled in size. This point also applies to a shift from Divided B to B , and gives rise to doubts about the comparison of those outcomes as well. Adams thinks the most plausible evaluation of Divided B and B is not that they are as good as each other, but that neither is worse than the other. ‘Not worse than’ may here be non-transitive, just as in the comparison of $A+$ with A . If so, we get the following evaluations: $A+$ is no worse than A ; Divided B is better than $A+$; B is no worse than Divided B . From these we can infer that Divided B is no worse than A , but not that B is no worse than A , as required by the Mere Addition Paradox.

This objection, though substantial, is not one to which proponents of the quantitative argument against BPW can appeal. Why not? If mere addition makes outcomes better, $A+$ is better than A . If Divided B is better than $A+$, and B no worse than Divided B , B is surely no worse than A . The quantitative argument judges $A+$ very much better than A . How could all this advantage be lost when Divided B is better than $A+$, and B no worse than Divided B ?

Adams also raises a more general worry about the role played by causal isolation in the evaluation of outcomes like $A+$ and Divided B . When extra people are merely added to A it is the isolation that underwrites the judgement that $A+$ is no worse than A . Without the isolation, Adams thinks that the addition to A of people who live lives of lower quality than any of the lives lived in A makes the outcome worse. Parfit might well agree.¹⁰ If A is compared with B through a series of intermediate outcomes involving isolation, this isolation will have first to be introduced and then expunged. But if the introduction of isolation is enough to alter evaluations that we would otherwise make, so perhaps is its removal. The lesson Adams draws is that comparisons between outcomes should take into account the processes by which they arise. As B could come about from $A+$ or Divided B only by the extinction of one of two isolated groups, and as the extinction of a population is pretty obviously a bad thing, B may be worse than A even if $A+$ is no worse than A , and Divided B better than $A+$. Adams sees the stipulation that $A+$ come about from A by mere addition as a tacit admission of the relevance of processes as well as outcomes. Mere addition requires that extra people affect no-one else, but this requirement is ignored when $A+$ is transmuted into Divided B and then into B .

However, no exponent of the quantitative argument against BPW can endorse this broad methodological criticism of the Mere Addition Paradox.

¹⁰ Adams doubts that $A+$ is no worse than A even given the isolation of the extra people in $A+$, so without the isolation $A+$ may well be worse than A (478). But removing the isolation may have less significant effects when $A+$ is judged much better than A : the shift from ‘no worse than’ to ‘worse than’ is less dramatic than that from ‘much better than’ to ‘worse than’.

The quantitative argument relies upon the idea that possible worlds can be directly compared in value as candidates for creation. If there is any reference to processes here, it is not one that can discriminate between possible worlds, for whichever world is selected for creation it will come about by the same process that would have been involved had another world been selected instead, i.e. by divine fiat. If there are possible worlds corresponding to outcomes A and A+, then, according to the quantitative argument, the second is better than the first. What about worlds corresponding to Divided B and B? It is hard to see, from a divine point of view, any difference between B World and a world in which its population is separated into isolated halves, though it is tempting to think that God would prefer the former as more likely to promote solidarity amongst creatures. But the weaker claim that Divided B World is as good as B World is at any rate highly plausible: in each world the same number of lives are lived and these lives are of the same quality.

What about the comparison of A+ World with Divided B World? Nothing in the quantitative argument itself suggests that weight be given to principles such as equality or maximin, so an exponent of that argument might refuse to recognize any morally relevant distinction between the two worlds. In defence of this refusal, it could be acknowledged that beneficence would require us to bring about Divided B from A+ if we could, but this is a point about outcomes, not possible worlds, and it is the morally good features of the redistributive process that carries weight here. No redistribution takes place when Divided B World is created instead of A+ World, so we cannot appeal to the relative sizes of gains and losses in order to prefer the former over the latter.

If we refused all consideration of equality, maximin and beneficence, we could deny that Divided B World is better than A+ World. But this will hardly help us avoid the Repugnant Conclusion, for what is espoused is the claim that it is only the quantity of worthwhile lives that matters. On this view, consideration of worlds intermediate between A World and B World is unnecessary, for B World is clearly twice as good as A World anyway. If this view is supplemented by an appeal to the Valueless Level, descent towards the Repugnant Conclusion will stop at that level, and we are back to the position discussed in section III. But it is hard to see how we could end up anywhere else, given the initial judgement that A+ World is better than A World, for all this advantage must be cancelled in the comparison of A+ World with Divided B World. The quantitative argument judges A+ World to be very much better than A World: it contains ten billion more worthwhile lives, all of reasonable quality. If B World is not better than A World, there must be some way in which Divided B World is much worse than A+ World, and it is unclear what this could be.

Appeal to the Valueless Level is difficult within traditional theism, at least

when that level is set high. But if vast increases in the quantity of lives lived can always compensate for decreases in the quality of lives, descent towards the Repugnant Conclusion is inevitable, and all that remains is to fix the rate at which quantity trades off against quality. With no upper limit to quantity's value, this rate determines the speed of descent, but cannot put a halt to it.

V

IS THERE AN UPPER LIMIT TO THE VALUE OF QUANTITY?

Could we set an upper limit to the value of quantity and still defend the quantitative argument against BPW? Suppose we deny that any world contains an infinite number of anything, limiting ourselves again to the finite case. Infinity is then a limit towards which worlds can only tend. We might claim that as the number of lives increases and so tends towards this limit, each subsequent increase is of less value, and hence less capable of compensating for qualitative decline. Larry Temkin proposes such a model and argues that it can offer a way out of the Mere Addition Paradox even for those who think mere addition makes outcomes better.¹¹ Somewhere in the descent from A to Z quantitative increases cease to count for anything much and other moral principles take over. Temkin thinks this is especially plausible because some principles, such as maximin and equality, matter more when lives are low in quality. Thus, even if someone judges A+ better than A, and B+ better than B, he may judge F+ worse than F. The Repugnant Conclusion is avoided.¹²

This imposition of an upper limit to the value of quantity, even if it is a limit to which outcomes only tend and which they never attain, is clearly at odds with the spirit of the quantitative argument. In order to be confident, solely on quantitative grounds, that the idea of a best possible world is incoherent, the mere addition of each worthwhile life must always make an outcome better. That is presumably because a worthwhile life is an intrinsically good thing. The intrinsic goodness of a life cannot depend on how many other worthwhile lives are also lived. Parfit distinguishes between the *personal* and the *moral* value of lives: 'What we call the value of a life is not its personal value – its value to the person whose life it is – but the value that this life contributes to the outcome' (408). Temkin employs a similar distinction, between the subjective and objective value of lives: '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'.¹³

¹¹ Temkin, 295.

¹² *Ibid.*, 339–340, n. 23.

¹³ *Ibid.*, 296.

Just as theism has difficulty accommodating the claim that some low quality lives are of value only to those who live them and do not contribute to the goodness of the world in which they are lived, so it has a hard time adapting to the claim that, as the total number of lives increases, each extra life matters less. The cases are, of course, somewhat different. The Valueless Level draws a qualitative distinction between worthwhile lives that seems both invidious and arbitrary when that level is set high enough to moderate the repugnance of the Repugnant Conclusion. Placing an upper limit on the value of quantity draws no qualitative distinction, and arbitrariness is apparent only in settling upon some number of lives below which personal and moral value coincide fully and in fixing the rate at which, above that number, these values diverge. But in both cases, the improvement effected by mere addition of worthwhile lives depends upon something other than the value of those lives to the people living them – either upon the quality of those lives relative to the Valueless Level or upon the number of worthwhile lives already being lived.

In Swinburne's formulation of the quantitative argument each worthwhile life is awarded its full personal value: we are invited to take any world whatsoever, and then asked to acknowledge that a world with just a few more worthwhile lives would, other things being equal, be better. Swinburne does not say that the addition of each extra life will always represent the same gain in value, so there is no formal incompatibility between the quantitative argument and the claim that the value of quantity has an upper limit towards which outcomes tend, each extra life being worth a little less. But the invitation to take any world whatsoever does suggest that we need not consult the character of the world to which lives are merely added, but only the character of these additional lives.

There is another difficulty here. If the moral value of extra worthwhile lives declines as the total number of lives increases, the quantitative argument may fail even in finite cases, for we must reconcile the claim that the positive value of quantity has an upper limit with the claim that the negative value of quantity has no limit at all. Even lives of very high quality contain some suffering, and with no limit to the disvalue of suffering, we are forced towards Parfit's 'Ridiculous Conclusion': as extra worthwhile lives are added, the personal value in them counts for less and less while the personal disvalue in them always counts the same (407). At some point, even if all these extra lives are of high quality, they will nevertheless make outcomes worse rather than better, and the quantitative argument will fail.

This difficulty shows that it is hard to limit the value of quantity without placing a limit on its disvalue as well. By distinguishing between *compensated* suffering that comes in lives that are worth living and *uncompensated* suffering that comes in lives not worth living, Parfit evades the Ridiculous Conclusion (408). There is no limit to the disvalue of quantity as uncompensated

suffering increases, but increases in compensated suffering due to extra lives being lived have no moral disvalue at all. The most natural way of understanding this asymmetry between compensated and uncompensated suffering is that the moral disvalue of the suffering in extra worthwhile lives is outweighed by the moral value that these lives also contain. But interpreting the asymmetry in this way removes the upper limit on the value of quantity. Instead, we must deny any moral disvalue to suffering that comes in extra worthwhile lives: this suffering has only personal disvalue. Because it is outweighed by personal value, it does not make an outcome worse (408–409).

Despite the success of this manoeuvre, Parfit thinks we must reject an upper limit to the value of quantity. If uncompensated suffering has unlimited moral disvalue, it can outweigh the limited moral value of extra worthwhile lives, and we are forced to accept the ‘Absurd Conclusion’ (409–412). Parfit’s argument is complex and subtle, but one point is easily grasped: once the limit to the value of quantity is reached, vast increases in the number of worthwhile lives cannot make outcomes better, or only by a little, whilst small increases in the number of lives not worth living makes outcomes worse. Consider a population that grows exponentially in size whilst the proportion of lives not worth living slowly falls. Once the upper limit to the value of quantity is reached, this process ceases to be one of improvement. As the population grows further, moral disvalue accumulates but moral value does not, and eventually we reach an outcome that is intrinsically bad, even though it contains vastly many worthwhile lives and proportionally very few lives not worth living.

The same point can be illustrated within the context of the quantitative argument against BPW. World W contains very many (n) worthwhile lives and only a few (m) lives not worth living. W_2 contains n^2 worthwhile lives and $2m$ lives not worth living. W_3 stands to W_2 as W_2 stands to W , and so on. Can we consistently maintain that at some point this process, sufficiently reiterated, will produce a world worse than W ? If not, there is no upper limit to the value of quantity, and appealing to such a limit in order to escape the Mere Addition Paradox will not succeed.¹⁴

The refusal to trade vast increases in the number of worthwhile lives for small increases in the number of lives not worth living is hardly ignoble; what seems odd is willingness to trade, but only up to a certain point. A world in which the upper limit to the value of quantity has been reached is

¹⁴ Adams suggests that an upper limit to the value of quantity can be preserved by appealing to the Negative Threshold Principle: ‘if the average levels and distribution of happiness, suffering and other goods and evils among the people living at any time are *not too bad*, the existence of a larger number of people with the same average levels and distribution of happiness, suffering and other goods would not be worse’ (475). This is plausible, but the most natural way of explaining why the principle holds is that increases in moral disvalue do not make outcomes worse when they are outweighed by increases in moral value. So understood, the Negative Threshold Principle is not compatible with an upper limit to the value of quantity.

not the same thing as a best possible world, for other kinds of improvement may still be possible. But other improvements – achieved by raising the average quality of lives or by reducing uncompensated suffering, for example – are conceptually unrelated to the quantitative argument, and these sorts of improvements may anyway have intrinsic maxima and so not call into question the coherence of the notion of a best among possible worlds.

The distinction between compensated and uncompensated suffering is a familiar one in theodicy, and putative solutions to the problem of evil often attempt to show that all suffering is – or will eventually be – compensated. Independently of particular eschatological doctrines concerning individual immortality and the like, the claim that there is no uncompensated suffering seems every bit as hard to believe as Leibniz's claim that we live in the best of all possible worlds. If those doctrines are taken into account, and if they really do serve to eliminate all uncompensated suffering, then nothing akin to Parfit's Absurd Conclusion can be forced upon someone who responds to the Mere Addition Paradox by endorsing an upper limit to the value of quantity. But the tension between the quantitative argument and the existence of such a limit is still unresolved. The quantitative argument is appealing precisely because it identifies personal value with moral value: worthwhile lives are good, and the more of them the better. If a world is the better for containing more worthwhile lives, it may still be better if those lives are of lower quality. The Repugnant Conclusion still threatens.

VI

MERE ADDITION AND INFINITE WORLDS

By this point, the defender of the quantitative argument against BPW may be feeling rather impatient. God is not restricted in Her choice of worlds to those corresponding to Parfit's outcomes A through Z. She can, if She chooses, create a world with as many people as Z and grant them all lives as good as the lives lived in A. But this is to miss the point at issue. The Repugnant Conclusion does not conclude that Z is the best outcome imaginable, but only that it is a better outcome than A. Defenders of the quantitative argument must agree, on pain of inconsistency with the core claim of their argument, that A+ World is much better than A World. Unless they can supply some good reason for thinking that B World is worse than A+ World – and worse by about as much as A+ World is better than A World – they must also agree that B World is better than A World. This starts the descent to the Repugnant Conclusion, and that descent is very difficult to stop, at least within a theistic context.

Together with most people, I find the Repugnant Conclusion hard to accept. But it is also hard to identify any flaw in the quantitative argument. The major unresolved question about that argument is whether it works in

the infinite case as in the finite. If a world contains an infinite number of creatures living enjoyable lives, is it obvious that another world containing a few more lives is thereby better? Before we can answer this question we must first make sense of it. Could any world contain an infinite number of lives worth living? Perhaps, if the world extends infinitely in time and space, there is no problem here. Could another world contain more lives than this first one? This is harder to understand. In the finite case, the natural way to think of mere addition is by adding extra lives to a world already given, carefully placing these extra lives where (or when) they will have no impact on lives already being lived. If the world is not big enough to do this comfortably, we can extend it by creating extra space (or time). But this way of thinking does not transfer easily to the infinite case, for the original problem about adding to an infinite number has now spawned a further problem about extending an infinite space (or time). Perhaps extension is not required: an infinite world ought to have room for a few extra people. Yet this in turn raises a question about the *density* of lives within infinite worlds. Can an infinite world have lots of unoccupied space (or time)? How so, when it contains an infinite number of lives?

Hilbert's hotel comes to mind. This establishment has infinitely many rooms, all occupied. It can nevertheless accommodate infinitely many new guests: the guest in room 1 moves to room 2, the guest in room 2 to room 4, the guest in room 3 to room 6, and so on. This leaves all the odd-numbered rooms vacant. Once infinitely many new guests have settled in, the process can be repeated, as often as required. But although this may help us understand the notion of adding to an infinite number, and even adding infinite numbers to each other, it does not help us settle the question whether we end up with more.¹⁵ In most hotels the occupancy rate goes up as each new guest arrives, but Hilbert's never gets any more crowded. Doubling the number of guests typically improves a hotel's profitability, but it is not obvious that it can do so when turnover is already infinite. It is also worth pointing out that relocating an infinite number of guests involves considerable disruption, so this way of adding is hardly 'mere addition'.

There are four different views we might take on these issues. We could deny that any world contains an infinite number of worthwhile lives. Mere addition will then always be possible, and will always result in a better world. Secondly, we could allow infinite worlds, allow that they can always be merely added to, and agree that such addition results in a better world. This is Swinburne's view. On both these views, the quantitative argument against BPW goes through. A third view allows infinite worlds and allows mere addition to them but denies that addition results in a better world. Or,

¹⁵ Mustn't we end up with more, if we have added? But this question can be turned around: how can we have added, if we didn't end up with more? 'Adding' and 'ending up with more' are internally related in finite cases, but this does not tell us how they relate in cases involving the infinite.

fourthly, we could allow infinite worlds but deny that they can always be merely added to. On these last two views the quantitative argument fails, and there is, in quantitative terms, a best possible world.

The second and third views allow mere addition to infinite worlds, but disagree over the consequences of this addition. The third view is puzzling: if mere addition to an infinite world does not make for a better world, does subtraction make for a worse one? The answer must be no; if we pick an infinite world at random and then pick another world with just a few more lives, the second is not better than the first, so the first is not worse than the second. But if this is right, we can subtract any finite number of worthwhile lives from an infinite world and still end up with a world as good as the one we started with. This implies that all infinite worlds are as good as each other in quantitative terms, despite the fact that some contain more worthwhile lives. Presumably these extra lives would be of value to the people who live them, but they would not make the worlds in which they are lived any better. Personal and moral value again diverge. This argument could be resisted by rejecting the claim that some infinite worlds contain more lives than others, but this empties the claim that it is possible to add to infinite worlds of all significance.

The fourth view does not face this problem. If some infinite world is such that mere addition to it is not possible, we cannot pick out a second world containing a few more lives that could be transformed into the first by subtraction. Of course, addition may be possible in this world without that addition being mere. In an infinitely extended world densely peopled with worthwhile lives there will be no way of adding lives whilst guaranteeing their causal isolation. This is what Hilbert's hotel is really like: once the hotel is full, new guests can be accommodated only by relocating guests already resident.

I have no idea whether the notion of a world that is infinite in this special way is coherent. But it does seem to have been what the most famous champion of the view that we live in the best of all possible worlds had in mind. In the *Theodicy* Leibniz writes:

... what can be said of a creature or of a particular substance, which can always be surpassed by another, is not to be applied to the universe, which, since it must extend through all future eternity, is an infinity. Moreover, there is an infinite number of creatures in the smallest particle of matter, because of the actual division of the *continuum* to infinity. And infinity, that is to say, the accumulation of an infinite number of substances, is, properly speaking, not a whole any more than the infinite number itself, whereof one cannot say whether it is even or uneven.¹⁶

Leibniz may have thought that every possible world was intensively infinite in this way, with the best possible world distinguished from all others on non-

¹⁶ Gottfried Wilhelm Leibniz *Theodicy* A. Farrar (ed.) trans. E. M. Huggard (London: Routledge & Kegan Paul, 1951, reprinted La Salle, IL: Open Court, 1985), s. 195.

quantitative grounds.¹⁷ But even if there is only one such world, that will be enough to show that the quantitative argument against BPW fails, if only in this special case.

This failure of the quantitative argument offers Leibniz, and anyone who wants to follow him, some scope for claiming that the notion of a best of all possible worlds is coherent. But it does not seem of any direct help in avoiding the Repugnant Conclusion. The quantitative argument fails when applied to worlds that are intensively infinite because mere addition to such worlds is impossible. This does not undermine the claim that where mere addition is possible – as with A and B and the rest of Parfit's alphabet – it makes outcomes better. Could the identification of an intensively infinite world as quantitatively best provide an upper limit to the value of quantity, and so halt the process of trading quantity for quality which forces us towards the Repugnant Conclusion? In Temkin's model, infinity is a limit to which outcomes tend but which they never attain. There is something mysterious about the idea of a series tending to a limit when any term in the series is infinitely distant from the limit to which the series tends.¹⁸ The mystery is all the greater when the limit of a series is not a countable infinity at all. But even if we are happy to live with these mysteries, we have no reason to think that extra lives decline in value as the limit is approached, and no reason to weaken our attachment to the claim that mere addition makes outcomes better until we arrive at that outcome in which mere addition ceases to be possible at all. What is true of an intensively infinite world, just because it is one to which mere addition is impossible, is that we cannot be forced by some variant of the Mere Addition Paradox to trade the quality of lives lived within this world for further gains in quantity, because all such gains will come about by addition that is not mere.

Are we then stuck with the Repugnant Conclusion? Defenders of the quantitative argument, contemplating worlds that correspond to Parfit's alphabet of outcomes, must judge Z-like worlds better than A-like worlds. How bad is this? Some people dismiss the Repugnant Conclusion as irrelevant because we could never in practice face a choice between outcomes like A and Z. Parfit argues that this is not an adequate response: unlike Robert Nozick's imaginary 'utility monsters' – beings who get 'enormously

¹⁷ *Theodicy*, s. 225: 'The wisdom of God... goes even beyond the finite combinations, it makes of them an infinity of infinities, that is to say, an infinity of possible sequences of the universe, each of which contains an infinity of creatures. By this means the divine Wisdom distributes all the possibles it had already contemplated separately, into so many universal systems which it further compares the one with the other. The result of all these comparisons and deliberations is the choice of the best among all these possible systems, which wisdom makes in order to satisfy goodness completely; and such is precisely the plan of the universe as it is'.

¹⁸ According to Temkin, infinite utility is a limit that finite worlds approach (297). On the simplest understanding, infinity is not a limit at all: the series 1, 2, 3, ... has no limit. Nor does the following series of value assignments to extra lives work: 1, 1/2, 1/3, ... As Parfit notes, the value of each extra life approaches zero, but the value of the sequence has no limit. For the sequence to have an upper limit, the value of extra lives must decline in a different way, e.g., 1, 1/2, 1/4, ... (402).

greater gains in utility from any sacrifice of others than these others lose¹⁹ – there is nothing deeply impossible about outcomes like Z. The comparison between A and Z is thus a fair test for our moral principles (388–90). I sympathize with Parfit here, but suppose for the moment that he is wrong, and the impossibility of bringing about Z makes it an inappropriate test of our principles. Does this mean the Repugnant Conclusion can be ignored by defenders of the quantitative argument against BPW? I do not think so. This way of dismissing the Repugnant Conclusion is not available in the context of creation. For us, bringing about Z is impossible, but God can create any world She chooses. Z World is no bare hypothesis but appears somewhere on the menu of creation. If the quantitative argument never fails, the repugnance of the Repugnant Conclusion is not hypothetical but real.

To see why, suppose that of the four views outlined above, either the first or second is true. The quantitative argument then goes through and God must have chosen to create a world to which mere addition is possible. Of the worlds better than the world that God chose to create, some are better in virtue of containing vastly more worthwhile lives, all barely worth living. There is something very disturbing about the idea that one way God could have done better was by creating a world that betters the actual world as Z betters A. When we think of ways the world could have been better we usually imagine qualitative improvements to the lives lived within it. When reasoning along with the quantitative argument, we imagine lives as good as or better than our own being added to the world. What we do not typically imagine are worlds that contain vastly many people whose lives are barely worth living.

Suppose instead that the fourth view is right, and there is, in quantitative terms, a best of all possible worlds. It remains true that, had God chosen to create a finite world, or an infinite world to which mere addition was possible, there would be a world that bettered the actual world as Z betters A. Contemplating this fact must still make us rather uncomfortable. But this discomfort is eased if God has chosen to create a world that is intensively infinite, for the only improvements that are possible are other than quantitative. The Repugnant Conclusion is still repugnant, but the repugnance is now more hypothetical than real.

Within the context of the quantitative argument, then, the Repugnant Conclusion ceases to be a threat if the actual world is intensively infinite and cannot be merely added to. But this does little to minimize the repugnance that we feel when we are coerced into preferring outcomes like Z over outcomes like A. If Parfit is right that the choice between A and Z is a fair test of our moral principles, we would like to find a way to modify our moral principles so that we can stop short of the Repugnant Conclusion.

¹⁹ Robert Nozick *Anarchy, State, and Utopia* (New York, NY: Basic Books, 1974), 41.

VII
CONCLUSION

I have argued that the defender of the quantitative argument against BPW cannot escape the Mere Addition Paradox by appealing to the moral relevance of processes (section IV) or by placing an upper limit on the value of quantity (section V). I also believe that these ways of evading the Paradox fail even when considered independently of the quantitative argument. Whether Z could ever come about from A, or could do so in a way that did not involve injustice or the breach of some other moral principle is, as Parfit says, ‘irrelevant to our question in its purest form’:

We are asking whether, if Z comes about, this would be better than if A comes about. We could imagine a history in which only Z-like outcomes occur. The people in Z would then be no worse off than anyone who ever lives. If we believe that Z would be worse than A, this would not here be because Z’s occurrence would involve injustice. (390)

Parfit finds the Repugnant Conclusion *intrinsically* repugnant. The impracticality of bringing about Z cannot counter this repugnance, for it remains true that, if only we could surmount the technical difficulties, we should do so. Equally, avoiding the Repugnant Conclusion by appealing to other parts of morality will not address this intrinsic repugnance, for then it is only because bringing about Z rather than A would violate other parts of our moral system that we are obliged to refrain from doing so.²⁰ What Parfit wants is a revision of our moral principles so that, independently of considerations deriving from elsewhere, we prefer A to Z. He fails to come up with the necessary revision, but hopes that others will succeed (443).

If mere addition makes outcomes better, as the quantitative argument claims, it is unlikely that the rest of our principles are capable of revision in ways that avoid any descent towards the Repugnant Conclusion. The best we can hope for is a softer landing. I argued above (section III) that theism cannot appeal to the Valueless Level when it is set higher than the level of many worthwhile lives that are actually lived. What if it is set lower, not far above the ‘Bad Level’ that Parfit introduces in the second version of the Mere Addition Paradox? Lives at the Bad Level are of value to those who live them but they lack many of the features that make lives worth living. They are ‘gravely deprived, crimped and mean’ (436). If we set the Valueless Level just higher than this, the Paradox drives us to prefer outcomes in which many billions of lives are lived, all not much above the Bad Level, over outcomes like A or B. Although less repugnant than the Repugnant Conclusion, this is still hard to stomach. U or V or W may be less repugnant

²⁰ For further discussion, see Parfit ‘Overpopulation’, 150, n. 6.

than Z, but when we compare them to A or B or C it is hard to accept that we should prefer any of the first group over any of the second.²¹

What if we compare U or V or W not with A or B or C but with the actual world? In U, V and W there is no inequality and every life is worth living, so these outcomes are in some ways better than the actual world. I also think the fact that vastly many worthwhile lives are lived in U, V and W is a way in which they are good. How repugnant is it to admit that, if we could bring about U or V or W, we should do so?

To those who judge A better than B, the answer is clear: any descent from A is unacceptable. I do not find this very plausible. But I do not know where the descent should stop. The central claim of the quantitative argument against BPW is hard to resist, and it drives us to prefer B over A, C over B, and so on. The case in which the quantitative argument fails – the Leibnizian case of a world that is infinite in all directions and on all scales – is one where mere addition is not possible, and this does not call the claim that mere addition makes outcomes better into question. This result is strange, for whilst it suggests that there is, in quantitative terms, a best of all possible worlds, the principle that identifies an intensively infinite world as best also forces the Repugnant Conclusion upon us. But perhaps this result is not intolerable; although outcomes close to Z must be judged better than outcomes close to A, theists who hold that this world is intensively infinite can at least claim that no worlds better ours in the disturbing way that Z betters A. If God could have done better, it will not be because She might have made many more of us live lives that were much less enjoyable.

However, if the notion of an intensively infinite world is incoherent, or if our world is not infinite in this special way, then the theist's position is less tolerable, for some of the worlds that are better than ours stand to it as Z stands to A. This conclusion is repugnant, but theists must either assent to it or reject the central claim of the quantitative argument against the coherence of the notion of a best possible world.

²¹ For further discussion, *ibid.*, 160–164.