CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Benatar and Beyond: Rethinking the Consequences of Asymmetry

Kaila Draper 🆔

Department of Philosophy, University of Delaware, Newark, DE, USA
Email: kai@udel.edu

Abstract

David Benatar's asymmetry argument in defense of anti-natalism is unconvincing, but not, as most of his critics would have it, because the alleged asymmetry on which it is based does not exist. Rather, the problem is that the existence of that asymmetry does not warrant the conclusion that it is better never to have been. This paper explains Benatar's mistake and identifies the correct conclusions to draw from the axiological asymmetry he identifies. It also sheds light on certain puzzles in population ethics.

With great ingenuity, David Benatar has defended anti-natalism, the view that it is morally wrong for human beings to procreate.[1] He has offered several arguments in its defense, but his "asymmetry argument" has received almost all the attention. That argument is unconvincing, but not, as most of his critics would have it, because the asymmetry on which it is based does not exist. Rather, the existence of that asymmetry does not warrant his intermediate conclusion that it is "better never to have been." My immediate objective in this article is to explain this mistake in Benatar's reasoning and reveal the correct implications of the axiological asymmetry that, in my view, he correctly identifies. My ultimate objective is to shed light on certain problems in population ethics.[2]

1. Benatar's asymmetry argument and its critics

The heart of Benatar's asymmetry argument can be found in four propositions. He writes:

---

[1]David Benatar, *Better Never to Have Been* (Oxford: Oxford University Press, 2006).

[2]Like Benatar, I rely on various "intuitions" to support my conclusions. This is because I hold the (controversial) view that axiological and moral judgments that, upon reflection, seem to be correct are presumptively correct. I recognize, of course, that a value judgment can seem to be correct without being correct. The presumption generated by an axiological or moral intuition can be defeated in a variety of ways. If my own intuitive judgment conflicts with those of my epistemic peers, for example, or is the product of some epistemic shortcoming, or conflicts with a well-supported axiological or moral principle, then the presumption that my own intuitive judgment is correct is defeated.

It is uncontroversial to say that
(1) the presence of pain is bad,
and that
(2) the presence of pleasure is good.
However, such a symmetrical evaluation does not seem to apply to the absence of pain and pleasure, for it strikes me as true that
(3) the absence of pain is good, even if that good is not enjoyed by anyone, whereas
(4) the absence of pleasure is not bad, unless there is someone for whom this absence is a deprivation.[3]

Relying on the asymmetry between (3) and (4), Benatar argues that, because (1), (3), and (4) are all true (and in spite of the truth of (2)), it is better for a human being to never exist.[4] From there he argues that procreation violates the duty not to inflict harm.

To appreciate and assess his reasoning, let us suppose that the following two outcomes are each an epistemic possibility:

Happy Life (HL): Jane Doe exists and, like all humans, experiences some pain and even misery over the course of her lifetime. But the disvalue for her of the suffering she endures is far outweighed by the value for her of the pleasure she experiences so that, on balance, her life is an exceptionally good one.

Never Exists (NE): Jane Doe never exists.

Because of outcomes like HL, it is natural to think that Benatar cannot be right that human existence is always a harm. If Jane does exist, then her life is a happy one: the value for her of the good things in her life far exceeds the disvalue for her of the bad things in her life. Why, then, should Benatar think that it would be better for Jane never to exist?

His reasoning is strikingly simple. Applied to HL and NE, it proceeds as follows: By (1) and (3), the absence of pain in NE is good: it is (intrinsically) better than the presence of pain in HL. Thus, NE is better than HL in at least one respect. But there is no respect in which HL is better than NE. HL is not better than NE in terms of pleasure because, by (4), the absence of pleasure in NE is not bad: it is not worse than the presence of pleasure in HL. All things considered, then, NE is better than HL.

Benatar's reasoning can also be expressed in terms of preferability:

(i) The pain Jane experiences in HL is worse than the absence of that pain in NE. (from Benatar's (1) and (3))
Thus: (ii) The fact that Jane experiences pain in HL but not in NE is a reason to prefer NE to HL. (from (i))
(iii) The pleasure Jane experiences in HL is not better than the absence of that

---

[3]Benatar, *Better Never to Have Been*, p. 30. Although he is not a hedonist, Benatar uses pleasure and pain as exemplars of things that have value and disvalue, respectively. That way, he can simplify the presentation of his argument by expressing it in purely hedonistic terms. He also assumes that all human lives contain some pain. I follow Benatar by expressing my objection to his argument in hedonistic terms and by assuming that all human lives contain some pain.

[4]Benatar, *Better Never to Have Been*, pp. 28–49.

pleasure in NE. (from Benatar's (4))
Thus: (iv) The fact that Jane experiences pleasure in HL but not in NE is *not* a reason to prefer HL to NE. (from (iii))
(v) Nor (we may assume) is any other fact a reason to prefer HL to NE.
Thus: (vi) There is no reason to prefer HL to NE. (from (iv) and (v))
Thus: (vii) NE is preferable to HL. (from (ii) and (vi))

My characterization of Benatar's reasoning requires one crucial qualification. Benatar makes it clear that his axiological ranking of outcomes is based only on the personal value of well-being, as opposed to also taking into account any impersonal value that well-being might have. As he puts it, "I am not making an impersonal evaluation…I am interested in whether coming into existence is better or worse *for that person* rather than with whether, for example, the world would be better if he exists."[5] Thus, Benatar does not deny (or affirm) that a life like Jane Doe's in HL, a life in which the benefits far exceed the harms, makes the world better. Perhaps by increasing the total or average well-being of the world's inhabitants, Jane's happy life would move our very imperfect world at least a small distance in the direction of perfection. Indeed, I believe that there are impersonal reasons of a perfectionist sort to favor HL over NE. But whether that is the case is not Benatar's concern. Thus, strictly speaking, his axiological reasoning leads only to the conclusion that, in terms of the personal value of well-being, NE is better than, and preferable to, HL.[6]

Subject to that qualification, I agree with Benatar that (ii) the fact that Jane experiences pain in HL but not in NE is a reason to prefer NE to HL and that (vi) there is no reason to prefer HL to NE. But I argue in section 2 that the truth of these claims does not justify the conclusion that (vii) NE is preferable to HL. The fact that, in HL, the value that Jane's pleasure has for her far exceeds the disvalue that her pain has for her is an *undercutting reason* – it undercuts the reason to prefer NE. In other words, even though the fact that, should she exist, Jane would experience pain is a reason to prefer that Jane never exists, the broader fact that, should Jane exist, she would experience pain that is well-compensated for by the pleasure she would also experience is not a reason to prefer that Jane never exists.

Returning to my initial characterization of Benatar's reasoning, I am willing to concede both that NE is better than HL in at least one respect and that (in terms of the personal value of well-being) HL is not better than NE in any respect. Nevertheless, I deny that it follows that NE is, all things considered, better than HL. That may seem incoherent, but I hope to show that it is not. The position I defend is that, in terms of the personal value of well-being, although there is at least one respect in which NE is better than HL and no respect in which HL is better than NE, HL and

---

[5]David Benatar, Still Better Never to Have Been: A Reply to (More of) My Critics, *Journal of Ethics*, 17 (2013), 121–51 (p. 125).

[6]I disagree with Benatar's characterization of the personal value of a person's well-being as concerning exclusively "what is better or worse for that person." But I do believe that it is possible to distinguish the personal value of well-being from the impersonal value of well-being. The former is the sort of value that the benevolent individual finds in well-being. The latter is the sort of value that, for example, a perfectionist might find in well-being. The benevolent person finds value in well-being because they value actual persons (and actual sentient nonpersons). The perfectionist finds value in well-being because they value perfection and so want the world to be as good as it can be.

NE are nevertheless evaluatively incomparable, which means that neither of them is better than or worse than the other, nor are they equally good.[7]

Before proceeding, I should say something about the criticism Benatar's asymmetry argument has already received. There is a lot of it, which precludes my explaining in any detail why I find most of it to be unpersuasive. What I can do is clearly distinguish that criticism from my own by briefly reviewing the kinds of objections that have been raised and pointing out that, whereas other critics have focused largely on raising doubts about the conjunction of Benatar's (1)–(4), I concede the truth of that conjunction but argue that it does not provide a justification for anti-natalism.

In evaluating certain objections to the asymmetry argument, it is crucial to recognize that the conjunction of (1)–(4) is not an assumption in Benatar's reasoning. He defends it in part by arguing that it enables us to account for a variety of intuitively appealing axiological and moral judgments. Benatar suggests, for example, that (1)–(4) provide an explanation of what he calls the "asymmetry of procreational duties": we have a duty not to bring unhappy people into existence, but no duty to bring happy people into existence. The wrongness of bringing an unhappy person into existence can be explained by the fact that, by (1) and (3), their suffering would be bad and its absence good (even if that good is not enjoyed by anyone). The absence of a duty to bring happy people into existence can be explained by the fact that, by (4), the absence of their pleasure would not be bad.[8]

It is not surprising, then, that some of Benatar's critics have sought to provide alternatives to (1)–(4) that can do at least some of the explanatory work that Benatar assigns to (1)–(4). Elizabeth Harman, for example, rejects the conjunction of (1)–(4) (which she calls "the Asymmetry") partly on the grounds that

> what's intuitively attractive about the Asymmetry can be explained in other ways. In particular, we clearly have strong reasons to avoid causing people to suffer, including to avoid doing what will bring someone into existence who will suffer horribly. Our positive reasons to cause people to have good experiences are much weaker. When we fail to bring someone into existence who would have suffered horribly, then we have failed to do something we should not have done.[9]

If Harman is right, then, because our reason not to inflict suffering is stronger than our reason to cause good experiences, we have a stronger duty not to bring into existence an unhappy person than to bring into existence a happy person. This explanation is unsatisfying, however, for it is by no means obvious that we have even a weak positive reason

---

[7]My use of the expression "(evaluatively) incomparable" follows Joseph Raz's usage. See his Incommensurability and Agency, in *Incommensurability, Incomparability, and Practical Reason*, ed. by Ruth Chang (Cambridge: Harvard University Press, 1997), pp. 110–28. Some use the expression "incommensurable" to mean what I mean by "incomparable."

Some value theorists argue that there are other ways two outcomes can be comparable in addition to one of them being better than, or worse than, or just as good as, the other. According to Derek Parfit, two outcomes can be *roughly* equal in value; and according to Ruth Chang, they can be "on a par." When I say, for example, that HL and NE are incomparable, I am saying not only that (in terms of the personal value of well-being) neither of them is better than or worse than or as good as the other, but also that they are neither roughly equal in value nor on a par. See Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), pp. 430–32. See also Ruth Chang, The Possibility of Parity, *Ethics*, 112 (2002), 659–88.

[8]Benatar, *Better Never to Have Been*, p. 32.

[9]Elizabeth Harman, Critical Study of Benatar, *Noûs*, 43 (2006), 776–85.

to cause good experiences by bringing people into existence. At a minimum, we have a much stronger positive reason to cause an existing person to have good experiences than to cause someone to have good experiences by bringing them into existence. We have a stronger reason, for example, to provide a happy future to someone who exists by adding, say, decades of happiness to their life, than to bring into existence a person who, if they do come into existence, will enjoy decades of happiness. Benatar can explain this asymmetry in reasons to benefit because (4) says that the absence of pleasure is not bad unless there is someone for whom this absence is a deprivation. Thus, Benatar can say that, in terms of the personal value of well-being, it is bad if an actual person is deprived of a good future by not being rescued from death, but not bad if a merely possible person is deprived of a good future by not being brought into existence. Harman offers no alternative explanation of this asymmetry as part of her critique of the asymmetry argument.

This example illustrates that a crucial component of Benatar's asymmetry argument is his claim that (1)–(4) provide an account of several asymmetries. Thus, he issues the following challenge: "If my critics think that there are better, or at least other, explanations for those asymmetries, they must show how *all* rather than only *some* of those asymmetries can be explained in other ways."[10] Moreover, he proposes that, even if that challenge is met by accounting for the relevant asymmetries in a piecemeal fashion, his account has the advantage of providing a *unified* account of all of the relevant asymmetries.[11]

A second group of critics have attacked the internal coherence of the conjunction of (1)–(4). Ben Bradley, for example, argues that, on either of two leading attempts to define "good," Benatar's claim that "(2) the presence of pleasure is good" is incompatible with his further claim that "(4) the absence of pleasure is not bad, unless there is someone for whom this absence is a deprivation."[12] On one of the two definitions, "p is good" is understood to mean "p is better than not-p."[13] Bradley suggests that, from this definition and Benatar's claim that presence of pleasure is good, it follows that the presence of pleasure is better than the absence of pleasure and so, contrary to (4), the absence of pleasure is bad (in the sense of being worse than the presence of pleasure). Benatar resists this reasoning partly on the grounds that, if p is the presence of pleasure, not-p can be understood as the absence of pleasure "when there is someone for whom this absence is a deprivation." Benatar does not deny that that sort of deprivation of pleasure is bad.[14]

On the other definition, "good" means "better than indifferent." Correspondingly, "bad" means "worse than indifferent."[15] Given these definitions, and assuming that the absence of pleasure is good, bad, or indifferent, if the absence of pleasure (due to nonexistence) is not bad, as Benatar claims, then it must be indifferent or good. Obviously, it is not good, and so it must be indifferent. But then because the presence

---

[10]Benatar, Still Better Never to Have Been, p. 127.

[11]Benatar, Still Better Never to Have Been, p. 127.

[12]Ben Bradley, Benatar and the Logic of Betterness, *Journal of Ethics & Social Philosophy*, 4 (2010), 1–5. For another challenge to the coherence of (1)–(4), see Campbell Brown, Better Never to Have Been Believed: Benatar on the Harm of Existence, *Economics and Philosophy*, 27 (2011), 45–52.

[13]Albert Brogan, The Fundamental Value Universal, *Journal of Philosophy, Psychology and Scientific Methods*, 16 (1919), 96–104.

[14]Benatar, Still Better Never to Have Been, pp. 135–36.

[15]Roderick Chisholm and Ernest Sosa, On the Logic of "Intrinsically Better", *American Philosophical Quarterly*, 3 (1966), 244–49.

of pleasure is good, and "good" means "better than indifferent," it appears to follow that the presence of pleasure is better than the absence of pleasure, contrary to what Benatar claims. Ultimately, I think that this argument can be rejected on the grounds that there is no single set of evaluative standards that one can use to rank the presence of pleasure and its absence, when the latter is due to never existing. As I argue in the next section, the presence of pleasure and the absence of pleasure due to never existing are evaluatively incomparable.

Finally, a third group of critics have sought to show that the conjunction of (1)–(4) has unacceptable moral consequences. Some argue that Benatar's endorsement of (1)–(4) commits him to recommending suicide.[16] Others charge him with a commitment to speciecide.[17] Benatar argues (persuasively I think) that he is committed to neither.[18] He does believe, of course, that he is committed to anti-natalism, and perhaps some would be inclined to turn Benatar's asymmetry argument on its head, arguing from the denial of anti-natalism to the negation of the conjunction of (1)–(4). My argument in section 2 is that, contrary to what Benatar believes and most of his critics do not contest, his endorsement of (1)–(4) does not commit him to anti-natalism because it does not commit him to his intermediate conclusion that it is better never to have been.

2. Conditional value and ranking outcomes

To explain my position, I need to say more about the asymmetry in Benatar's (1)–(4). That asymmetry is best understood, I propose, in terms of a deeper asymmetry that Benatar may recognize, but fails to appreciate:

> *The Fundamental Asymmetry*: For any possible subject S of pain and pleasure, S's pain has *unconditional* intrinsic disvalue; but S's pleasure has only *conditional* intrinsic value: it has intrinsic value only on the condition that there is a time at which S exists.[19]

Benatar's (1)–(4) follow from the Fundamental Asymmetry. The first clause of the Fundamental Asymmetry yields the conjunction of (2) and (4) (the presence of pain is bad, and the absence of pain is good, even if there is no one who enjoys that absence). Consider Jane, for example. Given the first clause of the Fundamental Asymmetry, her experiencing pain has unconditional (intrinsic) disvalue. Thus, Jane's experiencing pain is bad and her freedom from pain is good (i.e., better than her experiencing pain) even if she never exists. The second clause of the Fundamental Asymmetry yields the conjunction of (1) and (3) (the presence of pleasure is good, but the absence of pleasure is not bad unless there is someone for whom the absence is a deprivation). Consider

---

[16]Harman, Critical Study of Benatar, p. 274.

[17]Joseph Packer, Better Never to Have Been?: The Unseen Implications, *Philosophia*, 39 (2011), 225–35.

[18]Benatar, Still Better Never to Have Been, pp. 148–50.

[19]I use "value" and "disvalue" in their respective noncomparative senses. To say that pleasure has value (or is good in the noncomparative sense of "good") is to say more than that it is better than the relevant alternative. Nevertheless, some may argue that to be good in the sense that I regard as noncomparative is reducible to something comparative. It may be proposed, for example, that to be good in that sense is to be "better than nothing." My own view, which I do not defend here, is that such analyses ultimately rely on a noncomparative notion of neutral and so do not actually show that goodness is reducible to something comparative. I do not deny, of course, that what is good (in the noncomparative sense) is better than what is neutral and that what is neutral is better than what is bad.

Jane again. Given the second clause of the Fundamental Asymmetry, Jane's experiencing pleasure has only conditional value. It has value conditional on Jane's existence. Thus, given that Jane exists (at some time or other), her experiencing pleasure is good and her being deprived of pleasure is bad (i.e., worse than her experiencing pleasure). But given that Jane never exists, her not experiencing pleasure is not bad, because her experiencing pleasure has no value.

Just as Benatar does not regard the asymmetrical conjunction of (1)–(4) to be obviously true, I do not regard the Fundamental Asymmetry to be obviously true. As discussed in section 1, Benatar seeks to justify his "basic asymmetry" primarily by appealing to its explanatory value rather than its intuitive appeal, especially its usefulness in explaining a variety of other apparent asymmetries. My justification for the Fundamental Asymmetry is of the same sort. This will become more apparent in section 3. The point I want to make in this section, however, is that, if we ground the conjunction of (1)–(4) in the Fundamental Asymmetry, then we have a basis for attacking Benatar's reasoning from the conjunction of (1)–(4) to the conclusion that (even in a case like HL) it is better never to have been.

To see this, the first thing to notice is that conditional value (and conditional disvalue) can make ranking possible outcomes a messy affair. (Indeed, as will become apparent in section 3, I believe that some of the well-known puzzles one finds in population ethics are puzzling at least partly because of the confusing logic of conditional value.) Suppose that we want to rank two outcomes, each of them an epistemic possibility (and their disjunction an epistemic certainty). Suppose further that a state of affairs P has conditional value and that the condition that must be satisfied if P is to have value is satisfied in only one of the two outcomes. Then which possible states of affairs have value varies across the two outcomes to be evaluated and ranked. (Specifically, P has value in only one of the two outcomes.) To put it another way, the very *standards* for evaluating and ranking outcomes vary across the two outcomes to be evaluated and ranked. How, then, should we proceed?

One possible approach is the *shared-standards-only approach*. On this approach, if the standards by which the outcomes in a set of epistemically possible outcomes are to be evaluated and ranked vary across the outcomes in the set, then one ranks the outcomes by applying only the standards that are shared by the outcomes that are being ranked. This approach would yield Benatar's conclusion that, even though Jane would have a great life should she exist, it is better that she never exists. For although Jane would exist and so her pleasure would have value given that HL is the actual outcome, Jane's pleasure has no value if NE is the actual outcome. Hence, there is no shared standard concerning the value of Jane's pleasure. But Jane's pain has disvalue regardless of whether HL or NE is actual. Thus, it provides a shared standard that favors NE over HL. On the shared-standards-only approach, then, we reach the conclusion that NE is preferable to HL. Again, that is the conclusion that Benatar would endorse, but it is a highly counterintuitive one. It contradicts what I take to be the following common-sense judgment (CSJ):

> CSJ: If someone's life is a very good one in the sense that the value for them of the good things in their life far exceeds the disvalue for them of the bad things in their life, then it would *not* be better for them never to have been.

Thus, if there is an alternative to the shared-standards-only approach that does not contradict CSJ and would otherwise be no less plausible than the shared-standards-only approach, we should favor it.

The better approach I have in mind is the *all-standards-are-equal approach*. If you want to rank two epistemically possible outcomes, but the standards relevant to evaluating and so ranking those outcomes vary across the outcomes to be ranked, one must evaluate the outcomes by each of the two relevant sets of standards. If this leads to a uniform ranking, an all-things-considered ranking of the outcomes is thereby reached. But if the rankings vary depending on which set of standards is used, then no all-things-considered ranking is possible, and so we can say only that the outcomes in question are (evaluatively) incomparable.

On this approach, it is impossible to reach an all-things-considered ranking of HL and NE. Judged by the standards that would apply should Jane's pleasure have value because Jane exists, HL is, on balance, better than (and preferable to) NE. This is because, although HL is somewhat worse than NE with respect to pain, it is much better with respect to pleasure. But judged by the standards that would apply should Jane never exist, Jane's pleasure has no value and, therefore, because her pain would still have disvalue, NE is better than HL. Because one cannot rationally privilege one set of standards over the other so long as both sets of standards remain epistemically possible, it follows that, all things considered, one cannot say that HL is better or worse than NE (or even that the two are equally good). One can say, however, that NE is not better than or preferable to HL, which is the result one needs to accommodate CSJ.

Given the all-standards-are-equal approach, we can also justify the view that the reason to prefer NE to HL (namely, the fact that Jane experiences pain in HL but not in NE) is undercut by the fact that the value Jane's pleasure would have for her should she exist would far outweigh the disvalue her pain would have for her. The source of the undercutting reason is the conditional value of Jane's pleasure. If her pleasure had unconditional value, then it would provide an *overriding* reason to prefer HL to NE. But because her pleasure has only conditional value, the fact that, should Jane exist, the value her pleasure would have for her far exceeds the disvalue her pain would have for her merely *undercuts* the reason to prefer NE. It undercuts that reason by rendering the two outcomes incomparable.

We can now see why Benatar's asymmetry argument fails. He assumes that, because never existing is better than existing in one respect, and existing is not better than never existing in any respect, it follows that never existing is better than existing. But that conclusion does not follow. For if the Fundamental Asymmetry is true and the all-standards-are-equal approach is correct, then, even though NE is better than HL in one respect and HL is not better than NE in any respect, NE and HL are evaluatively incomparable, and so neither is better than the other. Benatar could try to repair his argument by showing that the shared-standards-only approach is superior to the all-standards-are-equal approach, but, unlike the latter approach, the shared-standards-only approach contradicts CSJ. Thus, there is at least one good reason to favor the all-standards-are-equal approach. Benatar could also try to repair his argument by showing that the Fundamental Asymmetry is false. I suspect that it would prove difficult to formulate an argument that threatens the Fundamental Asymmetry without equally threatening the conjunction of (1)–(4), but it remains to be seen whether Benatar (or someone arguing on his behalf) can do so.

3. Parfit-testing the competing positions

Perhaps there is another way that Benatar could establish that his position is superior to my own. For he defends his position in part by appealing to its advantages in terms of

yielding plausible treatments of some of the problems in population ethics that were famously introduced by Derek Parfit in *Reasons and Persons*.[20] Thus, if my position is less successful in terms of its implications for those problems, then Benatar's position would have an advantage. On the other hand, to show that my position is presumptively the superior one, it will suffice to show that its implications for those problems are no less plausible than the implications of Benatar's position. I do not need to show that my position does better than Benatar's with respect to these problems – although I believe it does – because my position has the initial advantage of being consistent with CSJ.

3.1. Parfit's "Asymmetry" and other asymmetries

Benatar identifies four puzzling asymmetries:

(i) *The asymmetry of procreational duties*
While we have a duty to avoid bringing into existence people who would lead miserable lives, we have no duty to bring into existence those who would lead happy lives.

(ii) *The prospective beneficence asymmetry*
It is strange to cite as a reason for having a child that that child will thereby be benefited. It is not similarly strange to cite as a reason for not having a child that that child will suffer.

(iii) *The retrospective beneficence asymmetry*
When one has brought a suffering child into existence, it makes sense to regret having brought that child into existence – and to regret it for the sake of that child. By contrast, when one fails to bring a happy child into existence, one cannot regret that failure for the sake of the person.

(iv) *The asymmetry of distant suffering and absent happy people*
We are rightly sad for distant people who suffer. By contrast we need not shed any tears for absent happy people on uninhabited planets, or uninhabited islands or other regions on our own planet.[21]

He explains all four of these asymmetries in terms of (1)–(4). Because I have acknowledged the truth of (1)–(4) and have explained their truth in terms of the Fundamental Asymmetry, my position is no less successful than Benatar's in terms of explaining the four asymmetries.

My position is also neither more nor less successful than his in terms of providing a satisfying justification for Parfit's "Asymmetry," which is an instance of what Benatar calls the "asymmetry of procreational duties."[22] As discussed in section 1, Benatar explains that asymmetry in terms of the asymmetry between (3) and (4). By identifying an asymmetry that underlies even the asymmetry between (3) and (4), I move the asymmetrical bump in the rug a little further than Benatar does, but even the "Fundamental Asymmetry" to which I appeal is extremely puzzling. Thus, I do not take myself to have provided a fully satisfying justification for Parfit's Asymmetry.

---

[20]Parfit, *Reasons and Persons*, pp. 351–441.
[21]Benatar, Still Better Never to Have Been, p. 123.
[22]Parfit, *Reasons and Persons*, p. 391.

This is not to deny that both my position and Benatar's have the virtue of providing an axiological basis for the asymmetry of procreational duties. To see clearly that my position has that virtue, consider another possible outcome:

> *Miserable Life* (ML): Jane Doe exists and experiences much pain over the course of her lifetime. She also experiences some pleasure in her lifetime, but the disvalue her pain has for her far exceeds the value her pleasure has for her. In short, she has a miserable life.

The fact that Jane would experience pain in ML but not in NE is a reason to prefer NE to ML. Moreover, the pleasure that Jane experiences in ML weakens, but does not completely undercut, the reason to prefer NE to ML. Indeed, judged by the standards that would exist should Jane's pleasure have value because Jane exists, ML is, on balance, worse than NE. Judged instead by the standards that would exist should Jane never exist, Jane's pleasure has no value and, therefore, because her pain would still have disvalue, ML is once again worse than NE. Because the ranking is the same regardless of which of the relevant sets of standards is applied, the all-standards-are-equal approach yields the conclusion that, all things considered, ML is worse than NE. That result, coupled with the result that, all things considered, HL is not better than NE, shows that the Fundamental Asymmetry can provide an axiological basis for the asymmetry of procreational duties.

It is easy to see that Benatar's other three asymmetries can also be explained in terms of the Fundamental Asymmetry. Consider the prospective beneficence asymmetry. Given the Fundamental Asymmetry, it is indeed "strange to cite as a reason for having a child that that child will thereby be benefited" because the outcome in which the child receives those benefits is evaluatively incomparable to, and hence not better than, the alternative outcome in which the child never exists. Furthermore, it is not "similarly strange to cite as a reason for not having a child that that child will suffer" because, given the unconditional disvalue of suffering, the absence of that suffering is better than its presence regardless of whether the child exists.

Moving to the retrospective beneficence asymmetry, it makes sense to regret having brought an unhappy child into existence – and to regret it for the sake of that child – because, given the unconditional disvalue of suffering, preventing that child's suffering by not bringing the child into existence would have been a much better choice. Asymmetrically, it makes no sense to regret not having a happy child because, given that the child never exists, the happiness they might have enjoyed has no value.

As for Benatar's fourth asymmetry, one can sensibly feel sad for those whose lives are miserable, regardless of how far away they might be, because their suffering has unconditional disvalue. But due to the conditional value of happiness, the absence of happy people in faraway places is not worse for the people who might have inhabited those places because, given that they never exist, their happiness has no value.

3.2. Mere addition and the repugnant conclusion

Let us proceed to Parfit's discussion of what Benatar calls the "mere addition problem."[23] In response to the view that the addition of lives well worth living can, by reducing average utility, make an outcome worse, Parfit asks us to imagine two alternative

---

[23]Benatar, *Better Never to Have Been*, p. 174.

worlds. One of them is A, a world in which ten billion people have equally blissful lives. A+ is a world in which ten billion people all enjoy that same level of bliss, and an additional ten billion people enjoy lives that, on balance, are quite good, although not nearly as good as the lives of the other ten billion people. How should we rank these two worlds in terms of well-being (as opposed to equality or fairness or other values)? Parfit is inclined to say that the mere addition of lives that are well worth living cannot make an outcome worse.[24] Thus, he is inclined to say that A+ is not worse than A even though the average level of well-being is higher in A than in A+. In defense of this conclusion, he asks us to imagine a world in which there are only two people, Adam and Eve, each of whom has a blissful life. Would that world become worse if billions of billions of lives only slightly less blissful than the lives of Adam and Eve were added to it, thereby reducing its average level of well-being? Parfit proposes that saying yes would be "absurd."[25]

Benatar disagrees. Because he thinks that existence is always a harm, he insists that "it would have indeed been better if no lives had been added to the Edenic lives of Adam and Eve."[26] Perhaps intuitions vary on the question of whether it would make things worse to lower the average level of well-being by adding lives to the Edenic lives of Adam and Eve. The problem for Benatar, however, is that on his view, it would even make things worse to *raise* the average level of well-being by adding to the Edenic lives of Adam and Eve lives that are even more blissful than theirs.

Benatar does make an attempt to mitigate the damage this sort of case does to his position by way of suggesting that Parfit's use of the expression "worth living" is ambiguous between "worth continuing" and "worth starting." He proposes that, on the one hand, if Parfit means "worth continuing" by "worth living," then an outcome can be made worse by the addition of lives that, once started, are well worth continuing. Specifically, such an outcome can be made worse by the addition of lives that, although well worth continuing, are not worth starting. On the other hand, if Parfit wants to claim that an outcome cannot be made worse by the addition of lives that are well worth starting, then Benatar can agree that Parfit's claim is true, but insist that his asymmetry argument shows that there are no such lives.[27] Either way, A+ can reasonably be judged to be worse than A, and even the addition of blissful lives to the Edenic lives of Adam and Eve can make things worse, or so Benatar claims.[28]

I find Benatar's distinction between lives worth continuing and lives worth starting to be a mere distraction. When Parfit describes a life as "well worth living," or "blissful," he can be understood to mean that the disvalue (for its subject) of the bad things in that life is far outweighed by the value of the good things in it. In hedonistic terms, this would be a life in which the value of that life's pleasures far exceeds the disvalue of its pains. The crucial judgment, then, is that an outcome cannot be made worse by the mere addition of lives in which the value of the intrinsic goods in each of those lives far exceeds the disvalue of the intrinsic evils in each of those lives. For all that Benatar says about distinguishing "worth continuing" from "worth starting," that

---

[24]Parfit stipulates that 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." (p. 420)

[25]Parfit, *Reasons and Persons*, p. 420.

[26]Benatar, *Better Never to Have Been*, p. 175.

[27]Benatar, *Better Never to Have Been*, pp. 174–75.

[28]Benatar, *Better Never to Have Been*, pp. 22–27.

judgment retains its intuitive appeal. Furthermore, restricted to cases (such as my modified version of Parfit's Adam and Eve example) in which the added lives are not only blissful, but also raise the average level of well-being, Benatar's claim that adding lives always makes an outcome worse does indeed seem rather absurd. Thus, I think it is fair to conclude that Benatar's treatment of the mere addition problem requires him to bite an intuitive bullet. The best he can do is to insist, on the basis of the strength of his asymmetry argument, that some bullets are worth biting.

How does my position fare with respect to mere addition? Quite well, I think. If we imagine that A+ and A exhaust our epistemic possibilities, then by the standards that exist if A+ is actual, A+ is better than A. However, by the standards that exist if A is actual, A and A+ are equally good. Because the two standards yield incompatible judgments, no all-things-considered judgment is possible. Thus, one cannot say that A+ is worse than A. Nor can one say that A+ is better than A, or even that A and A+ are equally good. But one can still agree with Parfit that, at least in terms of the personal value of well-being, the mere addition of lives that are well worth living (in the sense of having great net value for their subjects) cannot make an outcome worse. More importantly, in my modified version of Parfit's Adam and Eve example, it is not worse to add blissful lives to the lives of Adam and Eve.

But perhaps this is too hasty. Parfit complicates matters by introducing not just a mere addition problem but also a "mere addition paradox."[29] He asks us to suppose that, in addition to A and A+, we have the third possibility of B, a world in which 20 billion people all enjoy lives much better than the worst lives in A+ but also slightly worse than the best lives in A+. Diagrammed in the usual Parfit-skyscraper fashion so that height corresponds to level of well-being and width to population size, we then have the following three worlds:
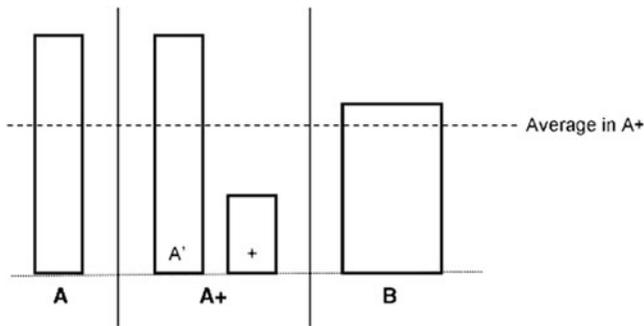
If we consult only our intuitions, how will we rank these worlds in terms of well-being (as opposed to equality or fairness or other values)? Parfit suggests that the judgment that B is better than A+ is the intuitively correct one. But he points out that, if we also judge that A is at least as good as B, then we seem to commit ourselves to the further judgment that A is better than A+, which contradicts his view that the mere addition of lives that are well worth living cannot make an outcome worse. If we instead affirm that A+ is at least as good as A, while continuing to hold that A is as at least as good as B and that B is better than A+, then it appears that we must reject the transitivity of the "at least as good as" relation. Transitivity can be preserved if we revise our original ranking by affirming that B is

---

[29]Parfit, *Reasons and Persons*, pp. 419–33.

better than A, but then Parfit's worry is that we may have taken a first step onto a slippery slope to his "repugnant conclusion", the conclusion that a world Z that contains zillions of people (including the 20 billion people who exist in B), all of whom have lives that just barely have a net positive value for them, would be better than A.[30]

Benatar's position avoids intransitivity because, on his view, A+ is worse than both A and B.[31] A+ is worse than A, according to Benatar, because more people are harmed (by being brought into existence) if A+ is the actual outcome than if A is the actual outcome; and A+ is worse than B because more harm is suffered if A+ is the actual outcome than if B is the actual outcome. On the assumption that the Z people experience at least as much pain as the A people, his position also avoids the repugnant conclusion, because the far greater number of lives in Z means that far more people are harmed if Z is the outcome than if A is the outcome.

His view does not, however, avoid the repugnant conclusion if we stipulate that, although the A people experience some amount of pain in addition to a huge amount of pleasure, the Z people experience very little pleasure and absolutely no pain. Lives that contain nothing bad for their subjects do not harm their subjects, and so Benatar must say that any life that contains no pain or other positive evil must be better than any life that contains some pain. Benatar concedes this point but maintains that lives that are barely worth living but contain no pain or other evil are not a genuine possibility. He claims that, if the Z people have long lives, then either their lives have a lot of pleasure in them and so are much better than barely worth living, or else they contain long periods of tedium and so actually do contain suffering. On the other hand, if the Z people have very short lives, then, according to Benatar, their lives do contain something bad because "the shortening of life is another bad."[32]

I find this argument unconvincing. The relevant outcomes are, of course, purely hypothetical. But if we are to consider hypothetical beings who never experience pain, then it is odd that Benatar wants to insist that such beings cannot possibly spend much of their lives in an affectively neutral state with no experience of pleasure or pain. Moreover, Benatar's suggestion that in a short life, the shortening of life is itself an evil seems to confuse negative evils – the mere absence of positive goods – with positive evils.[33]

I believe that my position holds up under the pressure of Parfit's "mere addition paradox" at least as well as Benatar's. If we imagine that A+ and B exhaust our epistemic possibilities, then if the same people exist in A+ and B, the same standards (for ranking outcomes in terms of the personal value of well-being) exist regardless of whether A+ or B is the actual outcome. By those standards, B is clearly better than A+, which is the intuitively correct judgment.[34] As we have seen, no all-things-considered ranking of

---

[30]Parfit, *Reasons and Persons*, p. 388.

[31]Benatar, *Better Never to Have Been*, n. 27 (p. 174), and pp. 176–78.

[32]Benatar, *Better Never to Have Been*, pp. 62, 176.

[33]Benatar could appeal to the view that death is intrinsically bad to support his view that a life cannot be barely worth living and yet free of any positive evils. As I have argued elsewhere, however, death is not intrinsically bad. See Death and Rational Emotion, in *The Oxford Handbook of Philosophy of Death*, ed. by Ben Bradley, Fred Feldman and Jens Johannson (New York: Oxford University Press, 2013), pp. 297–316.

[34]Although he is not explicit about the matter in his chapter on the mere addition paradox, perhaps Parfit wants us to imagine that there is no overlap between the B people and the A+ people. On that

A and A+ is possible if we assume that A and A+ are the two epistemic possibilities, and for the same reason no all-things-considered ranking of A and B is possible. I can at least agree with Parfit, however, that A is not worse than B and, more importantly, that A is not worse than Z. Thus, the repugnant conclusion is avoided. Moreover, intransitivity is avoided because the three relevant judgments are: B is better than A+, B and A are incomparable, and A and A+ are incomparable.

It might be objected that the all-standards-are-equal approach to ranking outcomes yields something close to the repugnant conclusion, because it has the consequence that a world Z in which zillions of people have lives that are barely worth living is incomparable to, and hence not worse than, A. My reply is that, although A and Z are incomparable if ranked *only* in terms of the personal value of well-being, the value of well-being also has an impersonal, "perfectionist" dimension.[35] Thus, because A is (ceteris paribus) an excellent world and Z is quite unimpressive by comparison, Z can reasonably be judged to be worse than A even if the standards that exist should Z be the actual world are the basis of the ranking.

3.3. The non-identity problem

Finally, let us consider Parfit's non-identity problem.[36] This problem arises because, in various kinds of cases, it seems wrong to bring some persons into existence even though, on balance, those brought into existence have lives that are worth living (in the sense that the disvalue of the bad things they suffer in virtue of having been brought into existence is exceeded by the value of the good things they enjoy). Parfit introduces the problem partly by way of the following example:

> *The 14-Year-Old Girl.* This girl chooses to have a child. Because she is so young, she gives her child a bad start in life. Though this will have bad effects throughout this child's life, his life will, predictably, be worth living. If this girl had waited for several years, she would have had a different child, to whom she would have given a better start in life.[37]

Parfit points out that we are inclined to think that the girl made the wrong choice: she should have waited to have her first child. The problem is that her choice appears to be victimless. Her actual first child – call him Joe – would not have existed had she waited. Because Joe's life is worth living, it seems that his mother's choice did not wrong him, nor did it wrong the possible child who never exists. Thus, the choice wrongs no one, and hence it is difficult to see why it is wrong. Yet Parfit and many others feel strongly that it is wrong.

The problem is complicated by the fact that intuitions about non-identity cases vary widely. Assuming that Joe's life is not just barely worth living but well worth living,

---

understanding of the paradox, my view implies that the two outcomes are incomparable in terms of the personal value of well-being. In terms of the *impersonal* value of well-being, however, B can still be regarded as better than A+ because B is a better world than A+. Therefore, my position is consistent with the judgment that B is all-things-considered better than A+ even if the respective populations of A+ and B do not overlap.

[35]I develop this view in The Personal and Impersonal Dimensions of Benevolence, *Noûs*, 36 (2002), 201–27.

[36]Parfit, *Reasons and Persons*, pp. 351–79.

some are inclined to deny that it would have been better for the much happier possible child to exist. Others (including myself) have the intuition that, although the existence of the happier possible child would have been better, it does not matter *that much* that Joe exists instead (so long as Joe's life is well worth living). Nevertheless, my position is likely to be attacked here. For the all-standards-are-equal approach does not offer any explanation for the alleged wrongness of not waiting. The two possible outcomes – the girl does not wait and so Joe (the less happy possible child) exists, and the girl waits and so the happier possible child exists – are evaluatively incomparable. Some will find that result unsatisfying.

I can respond, however, by proposing that the two possible outcomes are evaluatively incomparable only in terms of the personal value of well-being. In terms of the impersonal value of well-being, waiting has the better outcome: the world is a better world if the happier possible child exists. That may be an insufficient basis for saying that it would be wrong for the girl not to wait, but it would allow us to say that it would be better for her to wait. Still, it is arguable that Benatar's position has an advantage with respect to the non-identity problem, because his position accommodates the intuition that it would be wrong for the girl to bring Joe into existence. On his view, however, it would also be wrong for the girl to have the happier possible child, a result that Benatar reaches at the cost of denying CSJ. On balance, I am comfortable with my rejection of the assumption that the girl's choice is wrong as opposed to Benatar's rejection of the assumption that the girl's choice is victimless.[38]

4. Conclusion

I think it is fair to conclude that Benatar's attempt to justify his anti-natalist position by using it to shed light on various problems in population ethics does not rescue that position from my objection to it. My alternative position does at least as well as his position in terms of its implications for those problems, and it has the distinct advantage of accommodating the common-sense judgment that, if the good things in life far outweigh the bad, then it would not be better never to have been. To be fair, I should emphasize that Benatar's asymmetry argument is not his only argument in defense of anti-natalism. Indeed, he goes on to argue that we humans tend to over-estimate the value our lives have for us.[39] If he is right about that, then it may turn out that HL is not just a hypothetical example, but also an unrealistic example. Perhaps most or even all human lives are not worth the trouble. I do not think that is true, but I am an optimist by nature, and my worry is that Benatar is correct to suspect that people like me are optimistic to an epistemic fault.

---

[38]For an excellent defense of the position that the best response to the non-identity problem is to reject the wrongness of the relevant victimless choices, see David Boonin, How to Solve the Non-Identity Problem, *Public Affairs Quarterly*, 22 (2008), 129–59.

[39]Benatar, *Better Never to Have Been*, pp. 60–92.