THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Better never to have been believed: Benatar on the harm of existence

BETTER NEVER TO HAVE BEEN BELIEVED: BENATAR ON THE HARM OF EXISTENCE

CAMPBELL BROWN

University of Edinburgh

In *Better Never to Have Been*, David Benatar argues that existence is always a harm (Benatar 2006: 18–59). His argument, in brief, is that this follows from a theory of personal good which we ought to accept because it best explains several 'asymmetries'. I shall argue here (a) that Benatar's theory suffers from a defect which was already widely known to afflict similar theories, and (b) that the main asymmetry he discusses is better explained in a way which allows that existence is often not a harm.

1. BENATAR'S THEORY OF PERSONAL GOOD

First, some preliminaries. Existence is a benefit when it is better than non-existence, and a harm when it is worse.[1] We may think of this using possible worlds. Most of us, since we are mere contingent beings, exist in some but not all possible worlds. A person's existence in a world is a benefit to her if, and only if, this world is better for her than a world she doesn't exist in; and a harm if, and only if, it is worse.[2] A theory of what

---

For helpful discussions and comments on previous drafts I would like to thank Daniel Cohen, Toby Handfield and Yujin Nagasawa.

[1] This leaves the possibility that existence is neither benefit nor harm, for it might be neither better nor worse than non-existence. Existence and non-existence could be equally good, or they could be incommensurable.

[2] *Which* world she doesn't exist in? This doesn't matter, since all worlds she doesn't exist in are equally good for her; or at least, any world that is better than one of them is better than all (*mutatis mutandis* for 'worse'). An anonymous referee suggests that it is nonsensical to say things like 'a world in which this person doesn't exist is worse for her than one in which she does exist'. Presumably, the worry is that a world in which a person doesn't exist cannot be compared in this way with other worlds. However, while I agree that

45

makes worlds better or worse for a person we may call a theory of *personal good*.[3] Such a theory, on this picture, may have implications concerning whether existence is a benefit or harm.

Benatar's theory is a variant of familiar hedonism: a person's good is taken to supervene on the pleasure and pain she experiences. The innovation is that how these considerations are weighed in pairwise comparisons of worlds depends on the person's existence: if she exists in both worlds under comparison, then both her pleasure and pain are taken into account; otherwise only her pain is.

More exactly, Benatar's view may be stated as the conjunction of two principles:[4]

(P1)   If a person exists in both of two worlds, then which of these worlds is better for her depends on both the pleasure and pain she experiences in these worlds. Other things being equal, the more pleasure she experiences in a world, the better it is for her; and the more pain, the worse.

(P2)   If a person exists in at most one of two worlds, then which of these worlds is better for her depends solely on the pain she experiences in these worlds. The more pain she experiences in a world, the worse it is for her.

A person cannot experience any pain without existing, and therefore (P2) entails that if a person experiences any pain at all in a world, then this world is worse for her than a world she doesn't exist in. Benatar also believes that, as a matter of contingent fact, every person who ever lives will experience at least some pain. It follows that existence is always a harm.

2. WHAT'S WRONG WITH IT

Benatar's theory has a familiar problem – familiar, at least, to those who know the relevant literature on population ethics, especially the work of

---

such comparisons are dubious, I think this implies that statements like the one above are false, not that they are meaningless. As I explain below, my view is that, with respect to a person's good, worlds in which this person doesn't exist are incommensurable with other worlds, i.e. non-existence worlds are neither better, worse, nor equal in value with other worlds. This view is consistent with defining benefit and harm in terms of possible worlds as I have done here.

[3] Cf. Broome (2004: 63–68).

[4] Benatar does not explicitly present his view in precisely this way. His own presentation is stated mostly in non-comparative terms: 'good', 'bad', 'not good', and 'not bad' (see e.g. Benatar 2006: 30–31). What ultimately matters here, however, is how the view *compares* worlds (or 'scenarios', in Benatar's terms). We need to know under which conditions one world is judged *better* than another. So far as I can tell, what I've attributed to Benatar here is essentially his view expressed explicitly in comparative terms.

John Broome, which Benatar, inexplicably, ignores.[5] To illustrate, consider three worlds, *A*, *B* and *C*, and a person, Jemima, such that:

in *A*, Jemima doesn't exist;

in *B*, Jemima exists but experiences neither pleasure nor pain;

and in *C*, Jemima exists and experiences only pleasure.

It follows from (P2) that *A* and *B* are equally good for Jemima, and so are *A* and *C*.[6] But it follows from (P1) that *C* is better for Jemima than *B*. This seems simply incoherent. How can one thing be equal in value with both of two things one of which is better than the other? That seems too much like saying 'Claire is taller than Bob, and they are both the same height as Anne'.

Sometimes the relations 'better' and 'equal in value' are each defined in terms of a single primitive relation 'at least as good', as follows (where '$\succ$', '$\sim$', and '$\succsim$' stand for respectively 'better', 'equal in value', and 'at least as good'):

- $x \succ y$ iff ($x \succsim y$ and $y \not\succsim x$);
- $x \sim y$ iff ($x \succsim y$ and $y \succsim x$).

Given this approach, we may say that what the above example shows is that Benatar's view is incompatible with the transitivity of 'at least as good'. It implies that $B \succsim A$ and $A \succsim C$ (because $B \sim A$ and $A \sim C$), but that $B \not\succsim C$ (because $C \succ B$).[7]

Purported counterexamples to transitivity are hardly unheard of. Here's one.[8] Imagine a sequence of painful experiences: the first is two years of excruciating torture; the last is an incredibly long period with a mildly painful 'hangnail'; and each experience in the sequence is slightly less painful but much longer than its predecessor. You might think that a slight reduction in intensity of pain cannot compensate for a very large increase in duration, and so you judge that each experience in the sequence is at least as good as the next. Given transitivity, then, the first should be at least as good as the last. But you might doubt this; you might

---

[5] See e.g. the argument against the 'constituency principle' in Broome (1994: 168–170).

[6] Benatar comes at least close to stating explicitly that this is an implication of his view. Discussing a possible life which contains 'only good and no bad', he writes: 'About such an existence I say that it is neither a harm nor a benefit and we should be indifferent between such an existence and never existing' (Benatar 2006: 29). True, Jemima's existence in *B* is not exactly like this, since it contains no good (as well as no bad). But we could easily change the example so that she experiences some pleasure in *B* but more in *C*.

[7] Strictly, we should interpret '$\succsim$' here as meaning 'at least as good *for Jemima*'.

[8] The example is due to Temkin (1996). See also Rachels (1998).

Worlds	People		
	Jemima	Kieren	Leanne
A	–	0	1
B	0	1	–
C	1	–	0

think that two years of torture is always worse than a hangnail, no matter how long the latter lasts.

Examples like this, however, differ in an important way from the one given above. The pain example depends on the idea that sufficiently many non-compensating differences can combine to make a compensating one. The small reductions in intensity of pain between adjacent experiences in the sequence do not compensate for the corresponding increases in duration; yet the large reduction between the first and last does compensate for the corresponding increase in duration (or so it is claimed). But the example above is not much like this. It doesn't involve a long sequence of options with some parameter being decreased (or increased) in small increments from one to the next. Even if we were to accept intransitivity in these small-increment cases, we should remain reluctant to accept the quite different sort of intransitivity implied by Benatar's theory.

Further problems arise when we move from personal to general good – to what is good overall, taking into account the good of everyone. The following seems a plausible principle linking personal and general good.

(P3)   If a world is at least as good as another for everyone, and it is better for someone, then it is better.[9]

Now suppose we add two people, Kieren and Leanne, to our previous example, as shown in the table above.

As before, no person experiences pain in any of these worlds. The numbers represent quantities of pleasure experienced, and a dash indicates non-existence. For example, in world *A*, Jemima doesn't exist, Kieren exists but experiences no pleasure, and Leanne exists and experiences some pleasure. Compare first *A* and *B*. (P2) entails that *B* is at least as good as *A* for both Jemima and Leanne, and (P1) entails that *B* is better than *A* for Kieren. (P3) then entails that *B* is better than *A*, because it is better for someone and at least as good for everyone. By

---

[9] Cf. Broome (1991: 165).

parallel reasoning, however, it follows that C is better than B, and that A is better than C. So we have a cycle: $A \succ C \succ B \succ A$. Again, this is hard to make sense of.

Anyway, as is well known, such cycles present difficulties for rational choice. How might a person who subscribes to Benatar's theory make a rational choice between the three worlds, A, B and C? Suppose, for example, she would choose A. Then, assuming she is rational, we would expect her to stick with this choice if we were merely to remove one of the forgone options, C. But in fact she wouldn't: given a choice between only A and B, she would choose what she judges to be the better option, B. We could run the same argument no matter which option she would initially choose (e.g. if she'd choose B, we could ask why she wouldn't stick with this on removal of A). So it seems impossible for her to make a rational choice.[10] In technical terms, the problem is that Benatar's theory entails a 'choice function' that violates what Amartya Sen calls 'Property α' (a.k.a. 'Independence of Irrelevant Alternatives').[11]

3. THE ASYMMETRIES

Is there anything to be said in favour of Benatar's theory? The advantage he claims is that it gives the best explanation of several asymmetries (Benatar 2006: 31–36). The main one, which I'll call 'the obligation asymmetry' may be stated as follows:

> (Obligation Asymmetry) Bringing a person into existence is never morally obligatory, but it may sometimes be forbidden.[12]

This seems intuitively plausible. Suppose a man and woman know that, were they to have a child, it would have a very happy life. Nonetheless, we might think, they would do nothing wrong by deciding not to have the child. On the other hand, if they know the child would have a miserable life, they would do something wrong by having the child. This is the asymmetry that Benatar seeks to explain.

Here's how he explains it. First, his theory of personal good entails another asymmetry:

> (Benefit Asymmetry) Existence is never a benefit, but it may sometimes be a harm.

He then assumes (a) that it is obligatory to bring a person into existence only if this would benefit her and (b) that it is permitted (i.e. not forbidden) to bring a person into existence only if this would not harm

---

[10] Cf. Binmore and Voorhoeve (2003: 8).

[11] See e.g. Sen (1997: 122).

[12] The label 'obligation asymmetry' is mine, not Benatar's, as is the particular wording I've used to express the asymmetry here. For Benatar's formulation see Benatar (2006: 32).

her. Given these assumptions, the obligation asymmetry follows from the benefit asymmetry.

An explanation of the obligation asymmetry would, I grant, be a helpful thing. I want to suggest, however, that it is not well explained in the way Benatar tries to explain it, that is, via the benefit asymmetry. As I shall argue, we have good reasons to reject the latter asymmetry.

Let a J-world be any world in which Jemima exists, and let w_0 be a non-J-world. Then consider the following 'hexalemma':

(1) Some J-world is worse than w_0.
(2) Not every J-world is worse than w_0.
(3) If some but not every J-world is worse than w_0, then some J-world is at least as good as w_0.
(4) There is no best J-world.[13]
(5) If one world is better than another, and this second is at least as good as a third, then the first is better than the third.
(6) No J-world is better than w_0.

(For neatness, here I've abbreviated 'better for Jemima' as simply 'better', and so on.)

These six statements are jointly inconsistent. First, (1), (2) and (3) together imply that some J-world is at least as good as w_0. Call one such world w_1. Then, (4) implies that some other J-world is better than w_1. Call one such world w_2. Then, (5) implies that w_2, a J-world, is better than w_0. But this contradicts (6).

Statements (1) and (6) are the two sides of the benefit asymmetry, and so if we want to keep that asymmetry, we must sacrifice one of the other four, (2)–(5).

As we've seen, Benatar is forced to give up (5). This seems the worst option.

Another philsopher, Christoph Fehige, has proposed a theory – before Benatar did so – which supports the benefit asymmetry, but gives up (4) instead of (5) (Fehige 1998). According to Fehige's 'antifrustrationist' proposal, a world in which a person has no frustrated preferences is as least as good for her as any world.[14] This may be preferable to Benatar's view, but it still seems implausible. Consider a world where Jemima has no desires, or only very few or very trivial ones which are easily satisfied. Can we really not imagine a better world for Jemima than this? More

---

[13] For all x, if x is a J-world then for some y, y is a J-world and y is better than x.
[14] Fehige writes: 'everybody whose preferences are all satisfied is assigned a maximum level of utility'. This includes, he explains, several cases: 'the individual exists and has preferences and they are all satisfied, or it exists but has no preferences (say, it's a stone or a chair), or it does not exist' (Fehige 1998: 524).

generally, it seems plausible that for any really good J-world, we can conceive an even better one. Just imagine Jemima having a longer life, or having more friends, or being more successful, or whatever.

The hedonistic analogue of Fehige's theory is the view that avoiding pain is *all* that matters; pleasure never adds any value to a life. According to this view, there is a best possible J-world, indeed many of them. Any world in which Jemima experiences no pain, e.g. any world in which she doesn't exist, is at least as good for her as every world. Adopting this view would enable Benatar to retain the benefit asymmetry (and perhaps also the thesis that existence is always a harm, though this depends on the claim that life always holds at least some pain) without denying transitivity. But the claim that pleasure never matters just seems grossly implausible.

In support of (2), imagine a world in which Jemima has what we would normally regard as a very good life, virtually overflowing with pleasure, knowledge, beauty, and all that good stuff. Again, it seems implausible to claim that Jemima's existence in *this* world is worse for her than non-existence.[15]

Finally, we have (3). It may be helpful to explain this as follows. Say two worlds are *commensurable*, with respect to a particular person's good, if either (a) one is better for this person than the other, or (b) they are equally good. Then the one possibility ruled out by (3) is this: some J-worlds are worse for Jemima than w_0 and all the rest are incommensurable with w_0. You might think (as I do, incidentally) that no J-worlds are commensurable with w_0. How good a world is for a person, you might think, depends on how things are for her in this world. But in w_0, there is no way things are for Jemima, because she doesn't exist, and so there is no basis for comparing w_0 with other worlds with respect to Jemima's good. Suppose, however, you accept the antecedent of (3): you think some but not all J-worlds are worse than w_0. Then, since you think *some* J-worlds are commensurable with w_0, viz. the worse ones, it is unclear what reason you could have to deny that *all* are. In any case, it seems arbitrary to claim that the only J-worlds which are commensurable with w_0 are the ones which are worse than w_0 (unless you think every J-world is worse, which by hypothesis you don't).

So there is a significant cost to retaining the benefit asymmetry, a cost which in my estimation exceeds any gain. I favour instead rejecting this asymmetry and finding some other explanation of the obligation asymmetry. The symmetrical view I prefer (for reasons given briefly above) is that existence is neither benefit nor harm. This position is, I concede, not entirely comfortable. It deprives one of perhaps the most

---

[15] This doesn't beg the question against Benatar, since Jemima might have no pain in the imagined world. Benatar allows that this is at least conceivable.

natural explanation of so-called 'wrongful life' cases, viz. that the person wrongfully given life is harmed by her existence.[16] Instead, one must say that in these cases procreation is wrongful, though no one is wronged; the world is made worse, though not worse for anyone. I do not have an alternative explanation to offer here. My point is just that it would be worthwhile to seek out such an explanation, since the alternatives, especially denying transitivity, do not look good.

REFERENCES

Benatar, D. 2006. *Better Never to Have Been: The Harm of Coming into Existence*. Oxford: Oxford University Press.

Binmore, K. and A. Voorhoeve. 2003. Defending transitivity against Zeno's paradox. *Philosophy and Public Affairs* 31: 272–279.

Broome, J. 1991. *Weighing Goods*. Oxford: Blackwell.

Broome, J. 1994. The value of a person. Proceedings of the Aristotelian Society, Supplementary Volumes 68: 167–185.

Broome, J. 2004. *Weighing Lives*. Oxford: Oxford University Press.

Fehige, C. 1998. A pareto principle for possible people. In *Preferences*, ed. C. Fehige and U. Wessels, 508–543. Berlin: de Gruyter.

Parsons, J. 2003. Why the handicapped child case is hard. *Philosophical Studies* 112: 147–162.

Rachels, S. 1998. Counterexamples to the transitivity of better than. *Australasian Journal of Philosophy* 76(1): 71–83.

Sen, A. 1997. Quasi-transitivity, rational choice, and collective decisions. In *Choice, Welfare, and Measurement*, 118–134. Harvard: Harvard University Press.

Temkin, L. S. 1996. A continuum argument for intransitivity. *Philosophy and Public Affairs* 25(3): 175–210.

---

[16] Cf. Parsons (2003).