Can human extinction be morally desirable?
A critique of David Benatar’s anti-natalism

Anton A Van Niekerk
Stellenbosch University, South Africa
aavn@sun.ac.za
https://orcid.org/0000-0003-3362-3434

Abstract
According to David Benatar’s anti-natalism, the extinction of all sentient life, including that of humans, is morally desirable. That claim is contested in this article. The author commences by showing parallels between Benatar’s position and that of the existentialist Albert Camus. He points out that, also for Camus, the claim that “death is better” is meant in a moral sense. This is followed by a general exposition of Benatar’s anti-natalism. Particular attention is paid to Benatar’s argument about the “asymmetry of pleasure and pain”: the presence of pain is bad, whereas the presence of pleasure is good. However, whereas the absence of pain is good, it cannot necessarily be argued that the absence of pleasure is bad; what can, at most be claimed, is that the absence of pleasure is not bad. The author develops three arguments against Benatar’s position. He firstly (drawing on Dawkins) points out that Benatar’s position contests the most primordial of all instincts that characterize the phenomenon of life, i.e. reproduction. Secondly, drawing on Irenaeus and Hick, the author argues that the idea of a painless sentient existence is untenable. Thirdly, as suggested by the article’s title, the author argues that it does not make sense to promote a moral argument for the desirability of the non-existence of human life, since such an argument presupposes, for its very validity, due recognition as a moral argument or claim in an environment of morally responsible human interactions – that which Benatar’s argument suggests should disappear completely in the universe that he proposes.

Keywords
Albert Camus; anti-natalism; argument; asymmetry; David Benatar
Introduction: Camus and Benatar

The opening lines of Albert Camus’s famous book, *The Myth of Sisyphus*, are as follows: “There is but one truly serious philosophical problem and that is suicide. Judging whether life is or is not worth living amounts to answering the fundamental question of philosophy. All the rest – whether or not the world has three dimensions, whether the mind has nine or twelve categories – comes afterwards” (Camus 1955:11). Although David Benatar never, as far as I could find, refers to Camus, his book *Better never to have been* (Benatar 2006) seems to be a direct response to Camus’s challenge.

For Camus, life is absurd; it is analogous to the plight of the tragic mythological figure of Sisyphus who, as punishment by the gods, is compelled to roll a huge rock up a hill until, almost at the end, it breaks loose, rolls back and the process is to be repeated, *ad infinitum*. Camus’s book nevertheless develops into the advocacy of a heroic and sustained quest for meaningfulness in life in spite of the acknowledgement that absurdity is the primal feeling or affect generated in us by the fact of our existence in the world.¹ The challenge for men and women is to fight evil and promote justice even though there is no metaphysically justifiable hope for the world, and the experience of absurdity has the last word in any effort to construe such a hope.²

Benatar’s answer to Camus’s challenge is much more radical. For him, the reality of human life is, in the final instance, not absurd, but immoral. Therefore, it is “better” never to have been. His book is a systematic defence of the melancholic and desperate wish expressed by Heinrich Heine’s poem³ *Morphine*:

¹ See, besides Camus 1955 and 1961, also an extensive discussion of this topic by Louw, 1972:54–63.
² In an address to the Dominican monastery at Latour-Maubourg in 1948, Camus identified his own lack of hope as the most important difference between himself and his audience. “I share with you the same revulsion from evil. But I do not share your hope and I continue to struggle against this universe in which children suffer and die” (Camus 1961:50).
³ Benatar himself quotes poetry of Sophocles in the latter’s *Oedipus at Colonus* as the inspiration for the title of his book:

“Never to have been born is best
But if we must see the light, the next best
Gut is der Schlaf

Der Tod ist besser – freilich

Das best wäre, nie geboren sein

The “better” and “best” in these phrases are expressions, not of quality or of aesthetic desirability, but of 
morality. Drawing on a radicalized utilitarianism,\(^5\) Benatar concludes that the utilitarian calculus, which 
demands the optimization of pleasure and the minimization of pain, 
yields no other conclusion than the desirability of eradicating all human 
life. Thus this provocative and disturbing book goes way beyond Camus’s 
own project, and in fact challenges one of the seemingly most self-evident 
assumptions held by the overwhelming majority of the human race (i.e. 
that part of humanity who do not opt for suicide): that life is a benefit, a 
desideratum that is to be valued, cherished, protected, and prolonged as far 
as possible. At the same time, Benatar is not wilfully promoting individual 
or mass suicide (cf. 2006:211–221); once life occurs and is valued, it is 
morally impermissible to deprive anyone of such a benefit. To quote 
Benatar himself as to the core idea that he wants to promote: “The central 
idea of this book is that coming into existence is always a serious harm … 
Although the good things in one’s life make it go better than it otherwise 
would have gone, one could not have been deprived by their absence if one 
had not existed. Those who never exist cannot be deprived. However, by 
coming into existence one does suffer quite serious harms that could not 
have befallen one had one not come into existence” (Benatar 2006:1).

I shall, in this article, briefly deal with the main claims of the book, followed 
by a critical assessment of the central argument.

Is quickly returning whence we came.

When youth depart, with all its follies,

Who does not stagger under evils? Who escapes them?” (Benatar 2006:212)

4  Translation:

Sleep is good

Death is better

But the best is to have never been born

5  Benatar is not a traditional utilitarian. His specific brand of utilitarianism will be 
**Benatar’s anti-natalism**

Benatar argues against the common assumption that we are not doing something wrong when we reproduce, or, put differently, the assumption that to exist is, for the person who exists, more beneficial than not to exist. His argument is not that existence is necessarily harmful for those who do exist. A life exclusively characterized by good and in which evil, pain or suffering are never encountered, is, as far as Benatar is concerned, neither harmful nor advantageous. But, of course, of no life can it be claimed that it knows or experiences only good and nothing bad. All instances of sentient life of which we are aware contain both good and bad.

If, consequently, we wish to be consistent about the desirability of avoiding harm or suffering, we must admit that the only sure way of achieving such a state is to not exist at all. Only existing beings can experience harm or evil. This he motivates with reference to what he calls the “asymmetry of pleasure and pain” (Benatar 2006:30–39). We could, according to this utilitarian inspired view, agree that the presence of pain is bad, and the presence of pleasure is good. That symmetrical relationship does, however, not obtain when we consider situations that are characterized by the absence of pain and pleasure. When considering the latter, we must agree that the absence of pain is good – even if that good is not experienced by anyone actually alive. At the same time it cannot be claimed that the absence of pleasure is bad, as would be the case if the relationship between the absence and presence of pain and pleasure were symmetrical. What has to be acknowledged about the absence of pleasure, is *that it is not bad*, unless there is someone for which this absence would be a deprivation – i.e. the situation where pleasure is taken away from a being who already has access to that pleasure. Hence the *asymmetry between the presence and absence of pain and pleasure*: the presence of pain is bad, whereas the presence of pleasure is good. However, whereas the absence of pain is good, it cannot necessarily be argued that the absence of pleasure is bad; what can, at most be claimed, is that the absence of pleasure is not bad.

This asymmetry can diagrammatically be conceived as follows:
Pain = Bad
Absent pain = Good

Pleasure = Good
Absent pleasure = Not bad

If we are serious in our striving for the good, we are compelled to avoid pain at all costs. The *best possible world* then, according to Benatar, is not a world in which pain is absent and pleasure is present, but rather a world where both pain and pleasure are absent. The reason for this is that we need not say of the absence of pleasure that it is bad; the most we can say of such a state, is that it is not bad – which, of course, is not to say that it is good – except in those instances where that absence of pleasure boils down to depriving a living entity of something that is generally experienced as beneficial and that it already has. If there is nobody or nothing that is or can be deprived of any kind of pleasure, the net result is a situation where the absence of both pain and pleasure must be preferred to a situation in which pain is absent and pleasure is present. Such a situation, of course, could only be attained in a world in which no sentient being exists.

On the basis of this asymmetry Benatar then argues that, while we do have a duty to try and prevent the reproduction of suffering beings, we do not have a duty to produce happy (in the sense of only-pleasure-experiencing) beings. We always regret the suffering of beings that exist, but we do not regret the absent pleasure of those who have never and will never exist. According to Benatar, his position differs from that of traditional utilitarianism, although it need not necessarily differ (2006:36–37). He distances himself from a position which he calls “positive utilitarianism”, i.e. a position according to which we have a moral duty to not only avoid suffering but also to promote and increase pleasure. In the case of this latter form of utilitarianism, the absence of beings that are subjects of pleasure must be regretted. However, Benatar claims to promote another kind of utilitarianism – that version in which our duty does not coincide with the obligation to produce *more happiness*, but rather to produce *more happy people*; note the invocation of the acclaimed distinction between “making
people happy” and “making happy people”. The ideal of “making happy people” is the position in which happiness is considered to be the primary moral good, whereas people are at most seen as the vehicles of happiness, i.e. the ones that actually experience the happiness produced in the world. The idea of “making happy people”, however, introduces somewhat of a Kantian element into this brand of utilitarianism. This view acknowledges that people, as Kant would have it, are ends in themselves,\(^6\) and that the striving after happiness is secondary to the recognition of personal autonomy. This latter version is, for Benatar, the more desirable kind of utilitarianism. The implication, however, remains that, if/when people do exist, they in fact never experience exclusive happiness (i.e. a happiness characterized by the total absence of pain or suffering). The only way to guarantee that there is no pain or suffering in the world – that which, morally speaking, ought to be our foremost ideal – is when no sentient life form ever comes into existence.

I will not here deal with all the provocative implications that Benatar draws from this highly controversial position. I only refer to the implications for the abortion debate (2006:163–200). He argues that, whereas this debate normally deals with the question when, if ever, abortion becomes permissible, the question must rather be turned on its head. The moral question is not: “when is abortion permissible?”, but rather: “when is abortion avoidable?” The position thus defended is not a “pro-life” stance, but a “pro death” stance. The moral duty of people who indulge in sex is to avoid procreation as far as possible. All women who do fall pregnant, ought to have abortions unless they have excellent reasons not to – reasons that Benatar seriously doubts could ever obtain. A further implication of his position is the conviction that both the human and other sentient species should become extinct (Benatar 2006:194–200). He remains unclear as to exactly how that is to be brought about.

\(^6\) Bear in mind the formulation of Kant’s categorical imperative (one of a number of versions) that is relevant in this respect: “Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end” (Kant 1964:96).
Critical assessment

In discussions of his position on the Internet and elsewhere, many have taken recourse to ridicule and or *ad hominem* arguments in response to Benatar. Comments range from asking: “how does this professor end his semester course? By machine gunning all the students?” to comparing Benatar’s claim that “life is not worth living” with the judgments of the Nazi’s on the worth of the lives of Jews. Questions are asked about the consistency with which a position such as his can be argued, without the author in person resorting to suicide.

I strongly disagree with such approaches. Benatar’s position is certainly highly controversial and prone to be very unpopular, but that in itself is no justification for the recourse to ridicule and personal insults. To read his book is to come deeply under the impression of the seriousness of his convictions and the admirable logical rigour with which he argues his case. It could possibly be argued that his book is an illustration of the absurdity to which a consistent application of the logic of his brand of utilitarianism leads. However, I am of the opinion that Benatar’s position deserves serious, albeit it critical, attention. The fact that a position leads to thoroughly counter-intuitive conclusions have on condition that the arguments are well developed, never been an impediment to serious philosophical discussion.

Terminating the dynamo of the biological world?

A first line of argument could be to point out that Benatar, in advocating the desirability of ending all sentient life, sets himself up against the most primordial of all drives or instincts that characterize the phenomenon of life. One need only be reminded of the claims of someone like Richard Dawkins’s genetic brand of evolutionism in this respect. According to Dawkins, life persists because of the single-minded, non-teleological obsession of genes, as the primary agents of natural selection, to continue replicating themselves indefinitely. To quote Dawkins:

> Genes have no foresight. They do not plan ahead. Genes just *are*, some genes more so than others, and that is all there is to it” (Dawkins 1989:24). [And later]: What is a single selfish gene trying to do? It is trying to get more numerous in the gene pool” (Dawkins
This represents the most forceful and irresistible drive known in nature. To argue the way Benatar argues is to assume that a rational case can be made for terminating the very dynamo of the biological world. However, I would be hesitant to pursue this line of argumentation further since this argument is essentially of a scientific and not a philosophical nature.

I would like to limit myself to two other lines of argument against Benatar. The first has to do with the “all or nothing” nature of his argument: if there is any pain in the world, continued life becomes unjustifiable. The second has to do with a more fundamental matter, viz. the consistency with which a rational argument for the extinction of all sentient, and thus also rational, life can be made.

**If there is pain, sentient life ought to disappear**

The first serious philosophical problem that I have with Benatar’s argument is the unpalatable consequences that he derives from his brand of utilitarianism. Simply because it can, theoretically, be argued that there is an asymmetry between the presence and absence of pain and pleasure, and that the epithets of “good” and “not bad” are yielded by the evaluation of situations characterized by the absence of both pain and pleasure, continued life, as the subject of possible and actual pain, should become extinct. The implication of this line of argumentation is that the possibility of pain – any pain, however small or insignificant – cancels the justifiability of continued life. Put the other way round: No amount of good in the world can, in any way, justify the presence of any amount of pain, however small or insignificant.

---

7 Cf. the following, more elaborate explanation of Dawkins’s claim: “A gene travels intact from grandparent to grandchild passing straight through the intermediate generation without being merged with other genes … a gene … does not grow senile; it is no more likely to die when it is a million years old than when it is only a hundred … The genes are the immortals, or rather, they are defined as genetic entities that come close to deserving the title. We, the individual survival machines in the world, can expect to live a few more decades. But the genes in the world have an expectation of life that must be measured not in decades but in thousands and millions of years” (Dawkins 1989:34).
If this is the consequence of Benatar's brand of radical utilitarianism, it proves to my mind the thorough untenability of such a position. This kind of argument feeds on such an abstract idea of what the term “life” implies that it severs any relation to what the concept normally entails. The idea of life without pain is not only unrealistic; it is, I would argue, inconceivable. This has early in the Western tradition been persuasively argued by Irenaeus in his well-known eschatological theodicy. Grappling with the question as to whether God, if He exists, could be held responsible for the presence of evil in the world, Irenaeus concludes that God is responsible for the possibility of evil – not because God himself is evil or unloving, but because evil is a necessary condition for moral growth: that which is often called “soul-making”. In support, Irenaeus develops what has been called the method of the counterfactual hypothesis: the challenge to conceive of a world where all evil as well as the possibility of suffering has vanished (Irenaeus 1952:47–56).

John Hick, in one of his discussions of the Irenaeun theodicy, argues that such a world would, for example, be a world in which nobody could ever hurt or disadvantage anybody else. Hick writes:

\[
\text{[T]he murderer's knife would turn to paper or the bullets to thin air; the bank safe, robbed of a million dollars, would miraculously become filled with another million dollars; fraud, deceit, conspiracy, and treason would somehow leave the fabric of society undamaged. No one would ever be injured by accident: the mountain climber, steeplejack or playing child falling from a height would float unharmed to the ground; the reckless driver would never meet with disaster. There would be no need to work, since no harm could result from avoiding work; there would be no call to be concerned for others in time of need or danger, for in such a world there would be no real needs or dangers (Hick 1963:46).}
\]

A world which renders all forms of pain and suffering impossible must, by definition, be a world without natural laws, without moral sensitivity and eventually, also, without freedom. It would be a world without any needs or dangers. It would also be a world without science, because due to the adaptability of nature in order to prevent all forms of harm, there would
be no durable structure of/in the world with consistent characteristics and patterns of behaviour that could be studied. As Hick writes:

In eliminating the problems and hardships of an objective environment with its own laws, life would become like a dream in which, delightfully but aimlessly, we would float and drift at ease (Ibid).

This may, *prima facie*, seem like a quite pleasant possibility. Irenaeus’s most significant objection, though, is that *moral concepts* would lose all meaning in such a world. In a hedonistic paradise, there could not be wrong deeds, because nobody could ever disadvantage another. This also means, however, that there could be no right or good deeds. The virtues of courage and perseverance have no meaning in an environment in which no danger or resistance occurs. Without life in a stable, objective environment, it is impossible to give any content to concepts like love, generosity, temperateness, self-control, etc. A “perfect” (for Benatar “painless”) world would thus hardly be able to contribute to the development of the moral characteristics of the human personality. In fact, for Irenaeus a world without these characteristics is not the best but is indeed the worst of all possible worlds.

A related perspective is raised by Len Doyal in his review of *Better never to have been* in the *Journal of Medical Ethics* (Doyal 2007). Responding to Benatar’s “dismiss[al] of declarations that the benefits of life still outweigh its harms” (2007:573), Doyal foresees the possibility, if not probability, of a significantly different outcome of the hypothetical Rawlsian original position where rational negotiators deliberate about the merits of coming into existence or not – a thought experiment that Benatar indeed pursues in the book (Benatar 2006:178–181). For Benatar, the outcome of such deliberations can only be that all lives are wrongful and that no lives are worth starting. Doyal rightfully suggests that it is unacceptable that such a negotiation could only yield the conclusion that the only possible world to come out of it, would be a world where there is no harm at all – “not even a pinprick”! (2007: 574). It is more than feasible to expect that other types of world might emerge from the original position – “worlds about which it could rationally be argued that, compared with Benatar’s paradise of oblivion, the potential harms of existence might be accepted as being worth
the potential benefits, including the harm of experiencing the absence of such benefits” (Ibid). Doyal continues:

Original negotiators might design a world in which they would be willing to be born that might have the following characteristics: a dramatic excess of benefit over harm; a range of goods and services which, when experienced, would make most harms seem secondary; a wide spectrum of effectively enforced positive and negative freedoms that are always exercised in the interest of the least well-off … (Ibid). [Therefore], the issue that faces Benatar is whether or not his analysis is flexible enough for original negotiators rationally to decide in any circumstances approximating actual human existence that his view is false and that they rather would prefer to exist rather than never experience anything (Doyal 2007:575).

Benatar’s argument is an “all or nothing” argument: either there is only pleasure, in which case life is justifiable, although not necessarily desirable, or there is pain – even a “pinprick” (Doyle) – in which case all life becomes unjustifiable. This, it seems to me, is not only a non sequitur; it is in fact a conclusion that forfeits any recognizable correspondence to life as we know it or as it can be known.

The consistency of arguing in favour of non-existence

The last issue I wish to raise in critical response to Benatar’s position and argument, has to do with the rational consistency with which the desirability of non-being is, or can be argued by a rational agent. My point is not that one can never be rationally persuaded to end one’s own life; cases of severe suffering where a terminal patient desires the end of his or her life above all else, are well known to us. Rationally justified suicides can and do occur; with that fact, I have no quarrel. My question rather has to do with the conditions for the attainment of the morally based insight that all sentient life is harmful and therefore undesirable.

Benatar is not arguing that there ought to be nothing rather than something. What such a claim in case could mean, is hard to fathom. What he argues instead, is that sentient life is necessarily harmful and therefore undesirable. In addition, he is not arguing that sentient life will, in fact,
disappear – as we all know it will, if current day physics theories about
the emergence and necessary demise of not only our planet and sun, but
in fact the entire universe are valid. He argues that sentient life, in view of
the moral unjustifiability of the phenomenon of suffering that seemingly
necessarily accompanies is, ought to become extinct. His argument, as
mentioned at the start, is throughout presented as a moral argument – an
argument that proposes how things ought to be, in contradistinction to
how they in fact are.

My problem with the consistency and sustainability of such an argument is
the following. Benatar develops his position, unpopular and controversial
as it may well be, on the basis of a carefully construed argument – an
argument that, of necessity, is a rational achievement that could not have
been reached without the intellectual effort or labour that was required
to attain the insight which the rational process eventually yielded. Put
differently: that no sentient life rather than any sentient life is the desirable
state of affairs, cannot be an insight that emanates from a mental process
if the conditions for such a process do not obtain. It might be argued that
humans are not the only possible subjects for rational or mental processes;
these could arguably also be achieved by gods, spirits, angels, or demons.
This possibility, however, would catapult the argument to a level of
speculation that I (and hopefully also Benatar) deem imprudent to pursue.
What we know of mental states and processes, we only know on the basis
of human experience and their motivations and desirability.

Supposing there is a universe and a world such as we know, without any
sentient life. For the first 10 billion years after the big bang, this was indeed
the case. Such a “state”, according to Benatar, is morally preferable to a
situation where sentient life with the potential of experiencing pain or
suffering obtains. Note again: the claim is not factual; factually we know
the universe once was like this and will one day again be like this. The
claim, rather, is moral: because such a state represents the only guarantee
against the possible and actual occurrence of harm, and since harm is
morally reprehensible, such a state is “better” or “preferable” – morally
speaking! Does such a scenario make any sense? Why would such a situation
be eligible for moral evaluation? Moral deliberation and evaluation of
necessity requires the presence, experience, and interaction of human (not
only sentient) beings that are alive. For the kind of activity and intellectual
achievement that Benatar’s book and argument presuppose, human life as we know it seems to be a necessary condition. It therefore does not seem to make sense to promote a moral argument for the desirability of the non-existence of human life, since such an argument presupposes, for its very validity, due recognition as a moral argument and claim in an environment of morally responsible human interaction.

Conclusion

We have seen that, according to David Benatar’s anti-natalism, the extinction of all sentient life, including that of humans, is morally desirable. That claim is contested in this article. The gist of Benatar’s claim is not that sentient life will in fact disappear, although he does not dispute the cosmological claim that sentient life, together with all matter currently in the universe, will, in the long run, be destroyed or radically transformed. Benatar is also not primarily concurring with the view of some environmentalists that humans are destroying the earth and, if the price of preventing this eventuality is the destruction of sentient life, then so be it.

Benatar’s claim, as has been demonstrated and argued in this article, is a moral claim. Sentient life must become extinct because that eventuality will ensure that pain comes to an end – pain that, in all its guises, is always wrong and ought always to be prevented.

In my discussion of Benatar’s position, particular attention was paid to his argument about the “asymmetry of pleasure and pain”: the presence of pain is (always) bad, whereas the presence of pleasure is good. However, whereas the absence of pain is good, it cannot necessarily be argued that the absence of pleasure is bad; what can, at most, be claimed, is that the absence of pleasure is not bad.

In this article, I developed three arguments against Benatar’s position. Firstly, drawing on the work of Richard Dawkins, I pointed out that Benatar’s position contests the most primordial of all instincts that characterize the phenomenon of life, i.e. reproduction. Dawkins’s claim is, of course, not primarily philosophical, but inherently scientific. Although I concur with that claim, I do not belabour it in the article, since the article consists of essentially philosophical, and not scientific, arguments.
The first full-fledged philosophical argument that is developed against Benatar’s anti-natalism draws on an aspect of the theodicy originally construed by Irenaeus and extensively developed by John Hick in the 20th century, particularly in his books *Evil and the God of Love* (1985) and *Philosophy of Religion* (1990).

Irenaeus and Hick argue that the idea of a painless sentient existence is untenable. Pain cannot only be evaluated negatively. Pain plays a constructive role in moulding sentient beings into morally mature, well-rounded individuals. To put it differently, the constructive effect that pleasure has on sentient beings cannot be fully realized without at least the resistance that must be exercised against the actual and possible experience of pain.

My second philosophical argument, as suggested by the article’s title, is the claim that it does not make sense to promote a moral argument for the desirability of the non-existence of human life, since such an argument presupposes, for its very validity, due recognition as a moral argument or claim in an environment of morally responsible human interactions – that which Benatar’s argument proposes should disappear completely in the universe that he proposes.

Put differently: the onus rests on Benatar to explain how we must make sense of a moral argument (i.e. that all sentient life should disappear in order to ensure that pain ceases to exist) in a world that will have ceased to contain moral agents or the objects of the actions of moral agents. Where there are no moral agents left, moral argumentation ceases to matter. Moral argumentation that claims validity in a world without sentient life is the equivalent of moral argumentation in a world such as one of the moons of a planet in a different galaxy; it makes no sense, and it is quite unclear why it could possibly matter.

**Bibliography**


