CrossMark

Does the repugnant conclusion have any probative force?
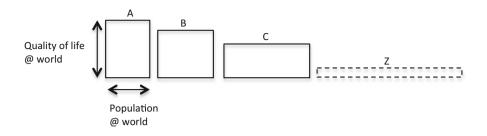
Christopher Cowie[1] ⓘ

Abstract In engaging with the repugnant conclusion many contemporary philosophers, economists and social scientists make claims about what a minimally good life is like. For example, some claim that such a life is quite good by contemporary standards, and use this to defend classical utilitarianism, whereas others claim that it is not, and use this to uphold the challenge that the repugnant conclusion poses to classical utilitarianism. I argue that many of these claims—by both sides—are not well-founded. We have no sufficiently clear sense of what a minimally good life is like. It is a result of this that the repugnant conclusion doesn't license us in drawing any interesting conclusions.

Keywords Repugnant conclusion · Wellbeing · Population ethics

Imagine a world, A, at which each inhabitant enjoys a very good quality-of-life. Now imagine a second world, B, comprised of an entirely different population, each of whom experiences a marginally lower quality-of-life than the inhabitants of A. Provided that the population of B is large enough in comparison with A classical utilitarianism entails that B is better than A (because the total welfare at B is greater than at A). Now imagine a third world, C. C stands to B as B stands to A: per-person quality-of-life at C is lower than at B, but the population of C is large enough that classical utilitarianism nonetheless entails that C is better than B. The series of worlds continues in this fashion; D, E, F and so on. At the limit imagine a world, Z, in which each inhabitant of an enormously large population leads a life that makes only a minimal positive contribution to the total amount of welfare at the world (such a life is 'minimally good'). We can represent it as follows:

✉ Christopher Cowie
   cdc33@cam.ac.uk

[1] Fitzwilliam College, University of Cambridge, Storey's Way, Cambridge CB30DG, UK

 Springer

Classical utilitarianism entails that Z is better than A. This highly counter-intuitive result is Parfit's famous 'repugnant conclusion'.[1] It is generally regarded as a major challenge to classical utilitarianism and as one of the most important tools for thinking about the correct approach to population ethics and axiology more generally.[2] In this article I argue that the repugnant conclusion lacks the probative force that it is often thought to have. It lacks probative force because we don't have a sufficiently clear sense of what life at Z is like. The core argument—which is elaborated on below—is as follows.

(1) We can't draw any interesting conclusions from the repugnant conclusion unless we have a sufficiently clear sense of what life at Z is like.
(2) We lack a sufficiently clear sense of what life at Z is like.
(3) So we can't draw any interesting conclusions from the repugnant conclusion (1, 2).

If (3) is correct, then the repugnant conclusion is not an effective tool for thinking about the correct approach to population ethics and axiology more generally. I defend this claim. I *don't* present this result as part of a defence of classical utilitarianism. I take my conclusions to be consistent with both the truth of classical utilitarianism and its falsity (though it is more likely to be welcomed by those who would defend classical utilitarianism).

I should acknowledge that—in broad outline—this strategy resembles some responses to the repugnant conclusion in the literature. Specifically, it resembles some of those responses that attempt to undermine the robustness of the intuition that Z is better than A. I discuss this in greater detail in Sect. 5. I also note at that juncture the important differences in method, emphasis and detail between my view and existing proposals. But before doing this I must set out my own view.

I begin in Sects. 1–3 by critically engaging with the popular view that a minimally good life ('life at Z') is quite good by contemporary standards; good enough that the repugnant conclusion is not repugnant after all. I focus on a defence of this view—made by both Partha Dasgupta and Clark Wolf—that I refer to as *the*

---

[1] The repugnant conclusion—i.e. that Z is better than A—can be presented as a theorem of classical utilitarianism, or as the result of reasoning resembling a paradox in its structure as it is in Parfit's 'mere addition paradox' (1984). An alternative way of getting there—that doesn't presuppose classical utilitarianism—is in Ng (1989) and Huemer (2008).

[2] An important source is Tannsjo and Ryberg (2004).

creation test. I claim that it fails. In Sect. 4 I state the positive view. In Sect. 5 I situate it in the literature.

1 Quality-of-life and the creation test

The repugnant conclusion can be—and often is—presented as a challenge to classical utilitarianism. There are two basic responses available. The first strategy is to modify or abandon classical utilitarianism.[3] The second strategy is to accept classical utilitarianism and with it the repugnant conclusion. Amongst those who use the second strategy, one popular approach is to claim that the appearance of repugnance can be 'explained away' by attributing to us a mistaken view of what a minimally good life ('life at Z') is like. As I'll bring out in the course of this paper, there are actually several different ways of doing this. In this section I'll focus on—and argue against—one rather strong way of doing it (I outline a weaker, more plausible version in Sect. 4). It is that we imagine life at Z to be much worse than it is fact is. Life at Z is actually surprisingly good by contemporary standards. Once we see this, the appearance of repugnance goes away. We can represent this as follows.

> *The QoL Strategy*: Life at Z is better than we initially think; it's actually quite good by contemporary standards. Once we see this it no longer appears repugnant that Z is better than A.

Arguing for this requires taking a stand on what life at Z is like. There is no consensus on how to do this (some even question whether it is possible). But I'm going to begin by focusing on an—as of yet critically unexamined—approach that owes to Clark Wolf and Partha Dasgupta.[4] They have converged on a common approach that I refer to as *the creation test* (though their views differ importantly in other respects). The creation test is a way of thinking about what life at Z is like. The basic idea is as follows. We imagine creating someone, when thinking only about the welfare that he or she would enjoy across his or her life. We now ask ourselves whether it would be ok to create that person. If it would be, this is an indicator that that person's life isn't bad. Wolf puts it explicitly as follows:

> Life is marginally above the neutral level just in case if you knew that any child someone else conceived would enjoy at least that level of well-being, you would not regard that person as having a reason (deriving from consideration of the child's welfare) not to conceive a child. (77).

---

[3] Options include person-affecting utilitarianism (Narveson 1967), non-maximising utilitarianism (Blackorby et al. 1995) perfectionism (Parfit 1984) and views that are either consequentialist but non-utilitarian (Mulgan 2006) or non-consequentialist.

[4] All references to Wolf (2004) and Dasgupta (ms) unless stated. I am grateful to Professor Dasgupta for allowing me to see the manuscript and for permission to cite. Given that Dasgupta's work is a manuscript, it is subject to possible changes. The final version may not be susceptible to all of the criticisms offered here.

We find what looks like a very similar idea being appealed to by Dasgupta. He writes:

> [T]he acid test for the conclusion that life, all in all, is invariably a good is to ask ourselves whether we shouldn't pause before creating a person so as to imagine the kind of life that is likely to be in store for the potential child. (16).[5]

Both Dasgupta's and Wolf's thought is that this is preferable to a perhaps more obvious way of thinking about life at Z. According to that more obvious way of thinking, we approach life at Z by thinking about the lives—and in particular the preferences—*of existing people*. Specifically, we think about an existing person who would only marginally prefer to live rather than to commit suicide or to die. This is a natural start-point. And it delivers a very low value for life at Z: it delivers the result that lives at Z are of the same quality as the lives of those who only marginally prefer continued life to suicide or to non-existence. If this is the value for life at Z, then, Dasgupta and Wolf agree, it would be repugnant that Z is better than A. But, they claim, this is the wrong way to think about life at Z.

Why do they think that it is the wrong way to think? Why should we prefer the creation test to thinking about the preferences of existing people for continued life? One reason offered by both Dasgupta and Wolf (as well as a number of other authors) is that there are many reasons that people would prefer not to commit suicide—or to cease to exist—*despite* their level of welfare. As Dasgupta puts it:

> Religious prohibition, fear of the process of dying… the thought that one would be betraying family and friends, and the deep resistance to the idea of taking one's own life that has been built into us through selection pressure would cause someone even in deep misery to balk. (15).[6]

Their claim, then, is that thinking about life at Z in terms of the preferences of existing people for continued life is likely to be a misleading guide to the welfare of those people. A better approach to life at Z, they claim, is to think about the kind of life that we would *create*. And thinking this way, they claim, drives up the value for life at Z. We can think of this in two steps.

Firstly, it allows for—and supports—the claim that a life in which one only marginally prefers continuation to suicide is a *bad* life (and not a marginally good life). It supports this claim because we would not, if thinking only about the welfare of someone who we would create, deem it ok to bring into existence someone who lives a life in which they only marginally preferred suicide to that life. As Wolf puts it:

> Suppose you knew that any child you conceived would live a life in which she was always just well enough off that she forbore suicide. Most people would regard this as a strong reason not to conceive a child. (75)

---

[5] We find something very much like this in Dasgupta's (1994: 115, 2005: 429).

[6] See also Wolf (75).

Secondly, it drives up the value for life at Z to a level that is actually quite good by contemporary standards: good enough that the repugnant conclusion no longer looks repugnant. Dasgupta puts what he acknowledges is a very rough figure on it when he writes that:

> That reasoning also steers us away from the thought that [zero well-being] is a poverty line in the sense that the notion has been articulated in figures such as the World Bank's "dollar-a-day", or even the "two-dollars-a-day" benchmark. [It] should be searched for in a higher range of figures, possibly a lot higher… (15)

Wolf makes a very similar point, making explicit that this can be used to explain away the repugnance of the repugnant conclusion. He writes that:

> [W]hen I consider the world of Parfit's repugnant conclusion to be inhabited by people who are as well off as I would have them… the repugnance has worn thin. In fact I find it a stretch to regard this conclusion as repugnant at all. (77)

In the following section I will scrutinise the basic proposal outlined above.

2 An epistemic challenge for the creation test

What should we make of the creation test as presented above? There are really three different claims being made:

(A) We shouldn't start thinking about life at Z in terms of the quality of life enjoyed by some living person who marginally prefers continued life to suicide.

(B) We should start thinking about life at Z in terms of the quality of life that someone would need in order for it to be ok to bring them into existence.

(C) If we employ (2), Z looks quite good by contemporary standards: good enough that the repugnant conclusion doesn't seem repugnant.

I shall begin by simply granting (A) and (B). My critical focus will be on (C). Certainly, the intended conclusion of (C)—that life at Z is quite good by contemporary standards—is contentious. It is explicitly denied by, for example, Nils Holtug:

> It seems to me that my life could be significantly worse than it actually is and be worth living. (2004: 154).

And Melinda Roberts:

> The deeper objection is… that it is just not plausible that a life has to be more or less like out own to be worth living. (2015: 22).

Importantly though neither of these (Holtug or Roberts) have engaged fully with the kind of argument—the creation test—that we find in Dasgupta and Wolf. So it is worth going into a bit of detail and assessing the arguments from scratch.

The idea of the creation test [at (C)] is—I think—to act as a kind of 'intuition pump'. The intuition pump works by putting yourself in the position of a 'creator': someone making decisions about bringing someone else into existence. What kind of level of welfare would someone have to have in order for *you* to think it ok to bring them into existence? Intuitively, wouldn't it be relatively high? Would *you* think it ok to bring into existence someone whose life is lived on the edge of suicide or only marginally better, who suffers from illness or disadvantage of various kinds? And so on. The idea of the creation test is that by facing these questions your intuitions crystallise somewhat. Specifically, they crystallise somewhat on something like the view that it wouldn't be ok to create someone—when considering only their welfare level—unless their welfare level were quite high by contemporary standards.

There is, I think, certainly something to be said for this way of thinking. But how robust is it? It will be useful to work with an example. Many people do *actually* choose to create lives with welfare levels that are too low to vindicate (C). Think about Dasgupta's remark (quoted above) that we should look much higher than the World Bank's $3-a-day figure for a value for quality of life at Z. Billions of people who live below this threshold choose to create lives: the lives of their children. Doesn't this put pressure on the claim that the creation test pumps intuitions that quality of life at Z must be surprisingly high by contemporary standards? After all, it seems that many people who practically apply the creation test seem to think otherwise.

Dasgupta and Wolf are, I think, likely to respond as follows. Those billions of people mentioned above aren't really engaging in the creation test as it is intended at all (which isn't, I should stress, a criticism of them). This is because they aren't choosing to create *only on the basis of the welfare of the resulting child*. As Dasgupta has discussed at length elsewhere, people choose to have children (insofar as procreation *is* a choice) for many reasons, not all of which are directly related to the welfare of the child. They may, for example, choose to procreate due to considerations of their own welfare or the welfare of their family (if, for example, having children is a means of achieving economic security). Wolf quite explicitly takes this into consideration in his formulation of the creation test by specifying that the created child is *someone else's*, *not one's own* (see above quotation).

This would be a sensible response for Dasgupta and Wolf to make. But it isn't sufficient to fully undermine the basic objection. This is because we can take a *third person perspective* on these kinds of cases. We can ask whether *we* think that it is ok for these billions of people who choose to have children in the world's poorest countries to do so. We are now applying the creation test as intended, without the bias of the perspective of parenthood or prospective parenthood. And, for my part at least, the intuition is very much that it is ok for them to do so. So having applied the creation test I *still* don't find the creation test pumping the intuition that welfare at Z is surprisingly high by contemporary standards.

Again, there is an obvious response on the part of Dasgupta and Wolf. It is that my judgment that it is ok for these billions of people to have children in these kinds of cases isn't really a pure application of the creation test either. Chances are, when I judge that it would be ok for extremely poor people to have children I am not just

thinking about the welfare of the prospective child (as I should be if I'm following the creation test). I am thinking about, or at least being influenced by, the procreative (or other political) rights of those people. What I *should* be doing—if I am correctly following the creation test—is asking whether it would be ok for these billions of people to have a child *when thinking only about the welfare of the child*.
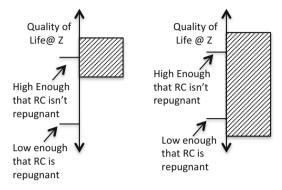
Let's try to meet this prescription (more on whether this is feasible below). Suppose we imagine someone who falls below the $3-a-day figure, who has a correspondingly small numbers of years spent in full-time education and low life-expectancy and who lacks various political liberties. Wouldn't it be ok to create this person when thinking only about his or her welfare? It is, I think, easy to imagine someone considering this question and quite reasonably responding that it *is* ok to create such a life. They might reason as follows: "This person will live a life that is like many billions of others, and that is better than many billions of people who came before her and who will come after her. She will experience a distribution of the highs-and-lows of human life that is, statistically, ordinary enough. Wouldn't it be ok to create her? Yes. Of course, it would be *better* if her level of welfare-level were higher, if she had greater economic, educational and socio-political opportunities. That could be said of many people who have lived, are currently living, and will come to live. But this is consistent with its being ok to create her when we think only about what life has in store for her."

This would seem to be at odds with Dasgupta and Wolf. They think, recall, that the creation test yields a value for life at Z that is really quite good by contemporary standards. And life on $2-a-day certainly isn't that. So what is going on here? Who is right and who is wrong? In a sense, neither. I don't think that Dasgupta and Wolf have misapplied the creation test, or applied it unreasonably in coming to the verdict that they do. There really is a way of applying the creation test that can make us think that life at Z must be quite good. Nor do I think that the—seemingly contradictory—imaginary use of the test above is unreasonable, or a misapplication. The truth is just is that the creation test leaves an awful lot of room for leeway. The result that one gets from applying it will depend on the kinds of properties of the created life that one is focusing on (e.g. income, life-satisfaction, hedonic states, etc.), how those properties are presented (e.g. the narrative structure, the stress, the comparative class, etc.), and of course, on any models for thinking about well-being that one has in mind (e.g. desire-fulfillment, eudaimonism, objective list, etc.).

The fact that the test leaves a lot of room for leeway won't, of course, be news to Dasgupta or Wolf. Neither claimed that it is a precise decision procedure that yields particularly determinate results. But it is important for what they want to it do—if I have understood correctly—that the leeway it leaves is constrained in a fashion that, if I am correct, it is not. What Dasgupta and Wolf want it to do, recall, is to help intuitions crystallise on the view that life at Z looks quite good by contemporary standards; good enough that the repugnant conclusion doesn't seem repugnant (I do consider an alternative, weaker interpretation of what they are trying to do in Sect. 5.1). That is claim (C) above. My worry is that it doesn't do this. In order to do this, the range of values for life at Z that the creation test generates must be quite high by contemporary standards (high enough that the repugnant conclusion isn't repugnant). But in fact the range of values that one could reasonably use it generate

is too large for this. The range includes values at the upper end that are probably high enough to do the work that Dasgupta and Wolf want done, but—crucially—it includes values at the lower end that are low enough to be consistent with the repugnance of the repugnant conclusion. The following diagrams can help to illustrate the basic point.

What Dasgupta and Wolf would need   vs.   What the Creation Test actually licenses

The vertical axes measure the (per-person) quality of life at Z: the higher up the scale, the better it is. The shaded rectangles represent the range of values for quality of life at Z as revealed by a reasonable application of the creation test. In order for the creation test to effectively help to pump the intuition that the repugnant conclusion isn't repugnant, it would need to yield the diagram on the left. But it actually yields the diagram on the right. It yields a view on which *both* competing hypotheses—i.e. (i) life at Z is good by contemporary standards and so the repugnant conclusion isn't repugnant and (ii) life at Z is bad by contemporary standards and so the repugnant conclusion is repugnant—are still in-play. Both, we might say, fall within the error-bars of the data given the experimental method (i.e. the creation test).

2.1 A further question

There is a further interesting question as regards how low the lowest boundary could be set given the creation test. Nothing much turns on the answer to this question as regards the central argument of this article. But it is interesting to speculate nonetheless.

I don't think it's, crazy to imagine that the lowest boundary could be set so low as to actually question Dasgupta and Wolf's initial claim that life at Z is better than the life of someone who only marginally prefers continued life to suicide. That is to say, someone could reasonably apply the creation test and get the result that, when considering only the welfare of the person that they would create, it would be ok to create someone who, across their life, marginally prefers continued life to suicide. There are two ways—one less extreme, one more extreme—in which I could see this going. Firstly, the less extreme case. In this case we imagine creating someone who marginally prefers continued life to suicide *but not because of* fear of death, religious prohibition and so on. That is to say, roughly, this person marginally prefers continued life to suicide just when thinking about their level of welfare. We

now apply the creation test. Would it be ok to create this person? I can certainly imagine someone thinking it ok to create this person (after all, they might reason, the person that we will create *does* prefer continued life to suicide and he or she isn't motivated by fear of death, religious prohibition and so on, so we should take this to be an accurate indicator of his or welfare). And again, this doesn't seem unreasonable.

Secondly, the more extreme case. In this case we imagine creating someone who marginally prefers continued life to suicide but is motivated in this by fear of death, religious prohibition and so on. We now apply the creation test. Would it be ok to create this person? Certainly, Dasgupta and Wolf think not. But there is at least a case to be made for thinking otherwise. Suppose that one reasons as follows: "Thinking about whether it is ok to create someone *when thinking only about their welfare* is a theoretical matter. Making a judgment about it is very different from making an ordinary judgment about whether it would be ok to create someone without the additional 'when thinking only about their welfare' rider. When answering this kind of theoretical question, it is best to approach it with some specific theory in mind. It is best to do so because we are outside of the realm in which our ordinary everyday intuitions are robust and reliable. The theory that we should use to approach questions about people's welfare is one based on the preferences of those people. Whether a life is going well or badly is determined by this. Given this, it is reasonable to think that if a person marginally prefers to stay alive rather than not, then their life is marginally good. Now it may be that this is because they fear death or are cognizant of religious prohibition (and so on). But can we really claim that these preferences aren't a legitimate input to the determination of their welfare? Do we, for example, really have any principled way of deciding which preferences fall within this category and which don't? Of course not! Or at least we don't unless we are willing to abandon the sensible theoretical claim that a person's wellbeing is determined by his or her preferences in the first place. So if someone marginally prefers to stay alive rather than not then their life is marginally good. And this is the case even if this marginal preference is borne out of a fear of death, of religious prohibition, and so on." Now I'm not claiming to explicitly endorse this. But I'm not sure that it seems like an unreasonable application of the creation test (indeed, I have spoken to some eminently reasonable people who have made claims very much like this). In any case, even if the lower boundary is higher than this extreme case would suggest, the repugnant conclusion will still look repugnant.[7]

---

[7] Wolf or Dasgupta might respond at this juncture by appealing to an alternative method to the creation test in order to show that life at Z isn't bad by contemporary standards. It is worth mentioning one in particular. Wolf claims at one point—actually, as a follow-up to his use of the creation test—that we can use *the repugnant conclusion itself* to calibrate life at Z: we should think of life at Z as being of whatever level would be necessary for us not to find it repugnant that Z is better than A (2004, 78). The idea is that this would end up generating a reasonably high value for life at Z by contemporary standards (or at least not a low value). I have some sympathy with this way of thinking. The problem with it, however, is that it isn't really a *response* to the challenge posed by the repugnant conclusion. It works by: (i) assuming that it isn't repugnant that Z is better than A, and (ii) fixing the value for life at Z so as to yield intuitions that are consistent with (i). This isn't a response to the challenge posed by the repugnant conclusion because it

3 A common problem

I don't think that Dasgupta and Wolf are the only authors to use the QoL Strategy. And similar problems apply to other attempts. Arguably (I return to why it is only 'arguable' in Sect. 5) Jesper Ryberg defends the QoL Strategy by arguing that, on balance, only relatively prosperous people will enjoy lives that rise to a level of minimal goodness. He writes:

> I believe a life worth living is not much different from a normal privileged life… [A] life contains a number of happy experiences and unhappy experiences, but it is not the case that the former vastly outweigh the latter. (Ryberg 2004, 242).

According to Ryberg then—at least as he is naturally read—life at Z is better than we might initially think. It is good enough that the repugnant conclusion isn't actually repugnant.

I don't think that this conclusion is warranted. It faces an epistemic problem that is very similar to the problem that the creation test faces. The problem is that we have an insufficiently precise idea of what a minimally good life would look like in order to warrant Ryberg's conclusion. To see this note that in order to work out what life at Z would look like on the broadly hedonic view that Ryberg is working with, we would need (a) a way of identifying and measuring the goodness or badness of the relevant experiential states (including a way of weighing length of state against intensity of state), (b) a wealth of empirical data to work out the distribution of these states in actual lives (presumably via lifelong longitudinal studies), and (c) a calculus for 'adding' the resulting values. But Ryberg does not provide anything like this information. Nor could he: unless I am much mistaken—and I stand to be corrected here—we don't even have anything close. In fairness, Ryberg is aware of this problems. Citing problems with measurement of wellbeing he writes that: "[i]t should be underlined that my judgment is highly impressionistic." (242). I agree with Ryberg about this. I am unsure why, given this, he takes himself to have a sufficient degree of confidence to make the claim that a minimally good life is probably "not much different from a normal privileged life" (I consider a possible explanation in Sect. 5).

It should be clear that this is closely related to the basic worry raised for the creation test in the previous section. The worry raised there was that a reasonable application of the creation test yields results that are insufficiently determinate to vindicate Dasgupta's and Wolf's claims. The worry with Ryberg's hedonic view is similar. There are serious problems—both conceptual and empirical—with

---

Footnote 7 continued

presupposes that the repugnant conclusion isn't repugnant. But the whole point of the challenge posed by the repugnant conclusion is that intuitively, many people *do* find it repugnant. Wolf's strategy here may nevertheless be correct. But showing that it is correct would surely require independent arguments for the truth of classical utilitarianism and its theorems (including the repugnant conclusion); arguments that is, that are independent of the QoL Strategy. For some of the most persuasive arguments to this end see Huemer (2008).

appealing to the hedonic view to show that life at Z is quite good by contemporary standards. These problems are such that, given our current state of knowledge, the boundaries for what a minimally good life looks like on the hedonic view will clearly be too broad to lend any confidence to the thesis that life at Z is quite good by contemporary standards.[8]

4 The probative force of the repugnant conclusion

What is the lesson of the arguments of Sects. 2 and 3 above? The lesson is *not*, I think, that quality-of-life at Z is lower than Dasgupta, Wolf, Ryberg and others claim it to be. That would be unwarranted. Rather, the lesson is that we don't really know—within reasonable limits—what life at Z is like. We don't know if it's really quite high by contemporary standards, or really quite low. We can put this as follows:

> *Insufficient Clarity View*: We don't have a sufficiently clear sense of what life at Z is like to know whether life at Z is surprisingly good, bad or neutral by contemporary standards.

This is, I think, a plausible thesis. It is a consequence of a fairly minimal epistemic humility. Suppose it is correct. What would it show? Obviously, it would show that Dasgupta and Wolf aren't entitled to claim that life at Z is surprisingly good by contemporary standards. So their appeal to the QoL Strategy won't work. But—and this point should be stressed—it would also show that their opponents who *deny* that life at Z is surprisingly good by contemporary standards (e.g. Roberts and Holtug as quoted above) are over-stepping the mark too. They claimed, recall, that life at Z *isn't* so good that our own lives are plausibly Z-like. If I'm right, however, then they aren't entitled to make this claim. This is because the grounds that I have provided for thinking that we are importantly ignorant of what a minimally good life is like are quite general: they should clearly make us sceptical of positive claims (i.e. it's surprisingly good) and negative claims (i.e. it's surprisingly bad) in equal measure. These are conclusions that all sides should accept. They are, as I say, a consequence of what seems to me to be a fairly, minimal epistemic humility about our knowledge of what life at Z looks like.

This is in itself a significant result. But it opens up the space for an even more interesting result. Specifically, it opens up the space for a *weakened* version of the QoL Strategy that should make us sceptical about the probative force of the repugnant conclusion. If we don't know whether life at Z is good, bad or indifferent by contemporary standards—and I don't think we do—then we can't confidently conclude that the repugnant conclusion is repugnant (or, for that matter, that it's not). We can put it like this.

> *The Weakened QoL Strategy*: In light of a lack of sufficient clarity as to whether life at Z is good, bad or neutral by contemporary standards, we can't tell whether Z is worse than A.

---

[8] Similar comments apply to Benatar's (2008) discussion.

This is weaker than the original QoL Strategy. Unlike the original QoL Strategy (that which Wolf and Dasgupta appear to be attempting to defend) it doesn't state that life at Z is surprisingly good by contemporary standards. Rather, it states that we don't know what life at Z is like by contemporary standards. But this is sufficient that we can't, with any degree of confidence, take the repugnant conclusion to undermine classical utilitarianism (or any other axiological view that entails it). More work would be needed. We can put this in terms of the argument that I presented in the introduction:

(1) We can't draw any interesting conclusions from the repugnant conclusion unless we have a sufficiently clear sense of what life at Z is like by contemporary standards.
(2) We lack a sufficiently clear sense of what life at Z is like by contemporary standards.
(3) (1, 2) So we can't draw any interesting conclusions from the repugnant conclusion.

If this argument is sound, this is a significant result. The repugnant conclusion is generally thought to be one of the more important thought-experiments in moral philosophy. It has certainly played a central role in axiology over the past 30 years. It has, for example, been one of the chief motivations behind rejections of classical utilitarianism in favour of non-maximising or non-person-affecting alternatives. But if the above argument is sound, then much of this is probably too quick.

I have already explained why I am sympathetic towards (2). I am sympathetic toward it because I take it to follow from a fairly minimal epistemic humility given the current state of our knowledge about what life at Z looks like. It is necessary to say something more about (1). Many of those who have something to say about what life at Z looks like (whether that it is good or bad by contemporary standards) will accept (1). But there will be resistance from those philosophers who believe the repugnant conclusion to be repugnant (and so to warrant rejection of classical utilitarianism) *whatever* life at Z turns out to be like. Their reasoning is as follows. What the repugnant conclusion *really* shows is that classical utilitarians are committed to a trade-off between quantity of lives and quality of lives. *That* is the real source of the repugnancy of the repugnant conclusion. And it remains as such even if life at Z is surprisingly good by contemporary standards. So the repugnant conclusion will still be repugnant—and will still warrant rejecting classical utilitarianism—even if life at Z is surprisingly good by contemporary standards. So (1) is false.[9]

I think that these philosophers are probably mistaken.[10] It is a mistake to think that the repugnancy of the repugnant conclusion is merely a consequence of the trade-off of quality of life for quantity of life and is unaffected by what life at Z is like. I can't defend this claim as fully as I would like here. But I will explain my

---

[9] We find something like this response in Holtug (2004) and Arrhenius (2004).

[10] I am not alone in this. This view is shared both by those who appeal to the QoL Strategy to defend classical utilitarianism and by some of those who don't. e.g. Roberts (2015).

reasoning in outline. Begin by thinking about how the repugnant conclusion is introduced in Parfit's initial 'mere addition paradox'. We are presented with a chain of reasoning that seems sound at each step but that yields a counter-intuitive conclusion (on the model of traditional sorites problems). At the first step we begin with a world, A, with high per-person-welfare for a low population. We compare this to a second world, B with lower per-person-welfare but higher overall welfare owing to a greater proportional increase in population. The conclusion—given the greater total welfare at B—is that B is better than A. We then compare B to C. The conclusion is that C is better than B. And so on. At the last step in the chain we compare Z to Y and conclude that Z is better than Y. We then take a step back and compare Z to A. We should—assuming the transitivity of the betterness of worlds[11]—conclude that Z is better than A. But, famously, we don't.
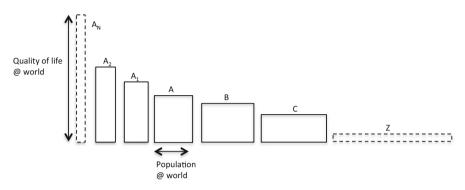
What is important to note about this reasoning is that we are *not*, initially, invited to have the intuition that A is better than B. We are initially invited to *accept*—or at least not to reject—the result that B is better than A (in virtue of its higher overall welfare). That's how the reasoning works. Each step seems ok. It seems ok that B is better than A, that C is better than B and so on. It is only when we get to Z and take a step back that something seems to have gone wrong. This is really important. It is important because this kind of reasoning simply wouldn't be possible if we started with the view that it is repugnant to trade-off quality of life for quantity of life. If we started with the view that it is repugnant to trade-off quality of life for quantity of life, we wouldn't accept the step from A to B or from B to C (because each of these steps are instances of trade-offs of quality for quantity). The reasoning would never get off the ground. So I don't think that the repugnancy of the repugnant conclusion is based on the rejection of views that trade-off quality for quality in a simple way.

What—if anything—is repugnant about the repugnant conclusion must be the *particular* trade-off that it describes. And at this point it becomes very natural to think that how we understand quality-of-life at Z must be relevant. Because it is surely not an accident that the trade-off that brings out the intuition of repugnance most strongly is one in which the inhabitants of A's comparator world (i.e. Z) have a *minimally good* quality of life (on this point, I think, Dasgupta and Wolf are quite right). So I think that—plausibly—we need to have an understanding of what life at Z is like in order to sensibly decide whether the superiority of Z to A is repugnant or not. But of course that is precisely what we *don't* have if the insufficient clarity view is correct.

Now I should acknowledge that this (i.e. life at Z) isn't the *only* place to look for an explanation of what is seemingly repugnant about the particular trade-off that the repugnant conclusion describes. The other obvious place to look is to the fact that there has been a *very large* trade-off between quality and quantity between A and Z. Perhaps *this* is what explains the repugnance: it is repugnant to make *very large* trade-offs of quality for quantity. If this is correct, then the repugnant conclusion can be of axiological significance—it can, for example, warrant us in rejecting classical utilitarianism—without us having the relevant knowledge about life at Z.

---

[11] Which is, of course, contestable (Temkin 1987; Rachels 2004).

But I'm not sure that this is correct. To see this, let's suppose that A is actually at the lower end of a further hierarchy of worlds: A to A_N. One place above A in the hierarchy is a world, A_1. A_1 stands to A as A stands to B. That is to say, A_1 has a lower population than A, but a higher per-person welfare. The hierarchy continues above A_1–A_2, A_3 and so on. We can represent it as follows:

Now we can imagine a world, A_N, which stands to A much as A stands to Z. The idea is that the trade-off between quantity and quantity that we find between A and Z is much the same as the trade-off that we find between A_N and Z. If you like, A is 'halfway' between A_N and Z. We can use this to test—and I think to undermine—the thesis that it is just the trade-off between quality and quantity that explains the repugnance of the repugnant conclusion. For if it were just this trade-off, then we would expect the superiority of A over A_N to be equally repugnant. But is it? Perhaps this is a point at which intuitions being to diverge. But certainly, the superiority of A to A_N doesn't seem to me to give rise to worries about 'repugnance' in anything lie the same way that the superiority of Z to A does. This suggests to me that Dasgupta and Wolf (and others) are right to think that judgments about quality of life at Z are playing a major role in shaping intuitions about repugnance: a sufficiently major role to make us question how much the repugnant conclusion can tell us without some kind of judgment about what life at Z is like. Hence (1). I acknowledge though, that this could certainly withstand further defence.

5 Situating the view

In this sub-section I discuss how the view that I have set out differs from, relates to, and builds on, neighbouring views in the literature. I focus on two different comparisons, one concerning how my view relates to other 'debunking' approaches to the repugnant conclusion (5.1), and the other concerning how it relates to John Broome's 'vagueness-based' approach (5.2).

5.1 Debunking

As I mentioned in the introduction, we can regard the basic strategy that I have used as an instance of a type of response to the repugnant conclusion that we sometimes see in the literature. The type of response is broadly *debunking* in character: it is based on querying the intuitive judgment that Z is worse than A. I'll now explain briefly how my own argument fits into this camp and how it differs from existing proposals of this kind.

The first point is that my proposal, unlike some instances of the debunking strategy, is based solely on considerations of quality of life at Z. Some existing instances of the debunking strategy appeal to altogether different sources of scepticism about the epistemic basis of the intuitive judgment that Z is worse than A. For example, both John Broome and Mike Huemer have argued that we should be sceptical of intuitions that Z is worse than A because we should be sceptical of intuitions in cases that involve large numbers.[12] This may be correct. But my argument is in any case concerned solely with considering quality of life at Z. It is also worth noting that, unlike both Broome and Huemer, I am not placing my argument within the context of a broader argument for the superiority of Z over A, let alone for classical utilitarianism. As I have made clear, I regard the results of this paper as compatible with either the superiority of Z over A or the superiority of A over Z. To adjudicate, we would also need independent arguments of the kind discussed at length by these authors. I discuss some further arguments from Broome below.

More interesting is the relation between my proposal and that defended by Torbjorn Tannsjo. Tannsjo has argued that we don't have a very good handle on what life at Z would be like, that it might well be rather good by contemporary standards, and so, that we shouldn't take the repugnant conclusion to count significantly against classical utilitarianism (Tannsjo 2015). Citing arguments that are in some respects similar to those that I presented in earlier sections, Tannsjo writes:

> Considerations such as these should either make the intuition that the A-world is (clearly) better than the Z world go away or, at least, they should give us pause if we want to use the intuition as evidence against utilitarianism… My claim is only that *there is no evidence to be found either way here*. (2015, 185. Italics mine.).

This sounds very much like the Weakened QoL Strategy.[13] There are however several differences between my account and Tannsjo's that it is worth noting. The first, most obvious difference is in the detail of how my view is established in comparison to Tannsjo's. I have taken as my departure point 'the creation test' as discussed by Wolf and Dasgupta. Tannsjo, by contrast, has primarily framed his argument—similarly to Ryberg—in terms of something closer to hedonistic

---

[12] Huemer (2008), Broome (2004).

[13] I thank an anonymous referee for *Philosophical Studies* for stressing this point.

utilitarianism. The second, much more interesting difference lies at the level of either emphasis or methodology. Tannsjo's account is, I think, somewhat ambiguous as regards whether it is an instance of the Weakened QoL Strategy or the QoL Strategy. In some places at least, Tannsjo makes claims that sound rather closer to the QoL Strategy (namely, that life at Z is actually quite good by contemporary standards). For example, he writes:

> In many situations we find ourselves momentarily below the line where our lives are worth living. Moreover many people probably live lives that, on the whole, are not worth living. When this is acknowledged, the repugnant conclusion does not seem repugnant any more. (2004, 223).

This sounds very much like the claim that the repugnant conclusion fails to undermine classical utilitarianism because life at Z is quite good by contemporary standards. That is to say, it sounds very much like the QoL Strategy (and not the Weakened QoL Strategy). As noted above however, in other places Tannsjo appears to defend a view that is much closer to the Weakened QoL Strategy. The view that I have defended is, I hope, less ambiguous in this regard. One can—as I do—question the probative force of the repugnant conclusion without committing oneself to the view that life at Z is probably quite good by contemporary standards.

It is worth briefly relating this back to the preceding discussion of Wolf, Dasgupta and Ryberg. As I presented it in Sects. 2 and 3 these authors made use of the QoL Strategy: certainly, the quotations that I gave seemed to support that. It may be however that this isn't the best way to read them. Perhaps their views are actually best thought of as an instance of the Weakened QoL Strategy. If so, then they are closer to mine than my presentation in Sects. 2 and 3 suggested. I don't think that the text allows us to decide on this. My view in any case is not, I hope, ambiguous in this way.

5.2 Broome on the vagueness of the neutral level

In *Weighing Lives* John Broome developed an important response to the repugnant conclusion. For anyone who is familiar with Broome's work, there will be obvious similarities between it and the view that I have presented above; similarities that make a comparison of the details—and the strengths and weaknesses—useful.

Broome is sensitive to the importance of what life at Z is like when thinking about the repugnant conclusion.[14] He argues that the neutral level—the level that falls between a life's being good and its being bad—is *vague*. By this he means that there are many lives for which it is in some sense indeterminate whether these lives are good, bad or neutral. He uses this as part of an argument in defence of classical utilitarianism in light of the repugnant conclusion. The argument is based on the application of a specific theory of vagueness: supervaluationism. According to this view, concepts with vague boundaries admit of 'sharpenings'. Sharpenings are precisifications of the concept that replace the vague boundaries of that concept with

---

[14] See Broome (2004).

a sharp boundary. For any sharpening, it is then a determinate matter whether the application of the concept in a given case is correct or incorrect. Each concept will admit of many different sharpenings. According to the supervaluationist—and this is their key move—one may make an assertion using that concept if and only if that concept applies *on all sharpenings*. It is a result of this, Broome argues, that we won't be able to assert that a life is better than neutral (i.e. that it is good) unless it is so on all sharpenings. But to be good on all sharpenings sets the bar for being good quite high: there might be some lives that are of quite a high quality but that are not better than neutral on some sharpening. If so, we may not assert that these lives are good. By setting the bar high for a minimally good life in this way, we drive up the value of life at Z. In this way, we avoid the repugnance of the repugnant conclusion.

I hope that it is clear why this view looks similar to the view that I have been defending. The key similarity is the use of a spectrum of different views about what life at Z is like. I claimed that the available tests for life at Z set wide boundaries for what life at Z is like: boundaries wide enough such that at the upper limit the repugnant conclusion isn't really repugnant after all, and at the lower limit it is. Broome claims that there is a similar width of boundary, only he explains this in terms of the vagueness of the relevant concept. One might wonder what the *real* differences between the views are (and hence, what my view adds to the existing literature). I'll briefly explain.

One obvious difference is in how the views are established. Broome's view that we should think about life at Z in terms of upper and lower boundaries is established by appeal to the vague nature of the relevant concepts. I arrive at the same view by different means. I arrive at it by reflection on the leeway left when we apply the best available tests to work out what a minimally good (or a neutral) life is like. These are very different approaches and they will lead to very different results. Our best tests are likely to leave a lot of leeway for thinking about what life at Z is like because of the vagueness of the relevant concepts (as Broome correctly claims) *as well as* shortcomings in the tests that we use, our best theory and a lack of empirical knowledge. My view is that all of these factors contribute to our lack of precise knowledge about what life at Z is like, and for all I have said, it could be that the vagueness of the concept contributes only very minimally to what we end up establishing about the relevant upper and lower boundaries from our current epistemic position. Broome would probably not disagree with this. But it is a very different approach nonetheless.

A second much more important difference between the two views is in their consequences. There is a sense in which both Broome and I are agreed: the repugnant conclusion fails to show that classical utilitarianism is false. Perhaps we are also agreed—although I am unsure on exactly whether Broome would consent to this point—that reflection on life at Z fails to show the repugnant conclusion to be either repugnant or not. But in any case there is an important difference in the *modal status* of our conclusions. Broome thinks that the source of the divergence in the upper and lower boundaries of a neutral life is due to the vagueness of the relevant concepts. As a result of this he is seemingly committed to the claim that the repugnant conclusion *couldn't* show classical utilitarianism to be false (unless it were recast using different concepts). I am not committed to this. My view is

consistent with the claim that the repugnant conclusion *could* show classical utilitarianism to be false (or at least seriously undermine it). My view is consistent with the claim that *if we managed to get sufficient clarity on the relevant tests and theoretical and empirical issues*, then we could establish something significant about the value for life at Z; we could potentially establish, for example that the value is high enough that the repugnant conclusion isn't repugnant or low enough that it is. When we put the point in this way, it naturally invites the question as to just why Broome *is* confident that the vagueness of the relevant concepts sets the boundaries for neutrality at the actual levels that he thinks it does.

6 Conclusion

We don't have a very good sense of what life at Z is like. This undermines the probative force of the repugnant conclusion. It follows that we can undermine the probative force of the repugnant conclusion without relying on the very strong—I think, overly strong—claim that life at Z is quite good by contemporary standards. We need only rely on a fairly minimal degree of epistemic humility about assessments of quality of life. This establishes that we don't know whether life at Z is good, bad, or neutral by contemporary standards. And this is sufficient to undermine the probative force of the repugnant conclusion.

References

Arrhenius, G. (2004). The paradox of future generations and normative theory. In T. Tannsjo & J. Ryberg (Eds.), *The repugnant conclusion: Essays on population ethics*. Dordrecht: Kluwer.

Benatar, D. (2008). *Better never to have been: On the harm of coming into existence*. New York: Oxford University Press.

Blackorby, C., Bossert, W., & Donaldson, D. (1995). Inter-temporal population ethics: Critical-level utilitarian principles. *Econometrica, 63,* 1303–1320.

Broome, J. (2004). *Weighing lives*. Oxford: Oxford University Press.

Dasgupta, P. (1994). Savings and fertility: Ethical issues. *Philosophy & Public Affairs, 23*(2), 99–127.

Dasgupta, P. (2005). Regarding optimum population. *The Journal of Political Philosophy, 13*(4), 414–442.

Dasgupta, P. (ms). Birth and death. *Time and the Generations* (**forthcoming**).

Holtug, N. (2004). Person-affecting moralities. In T. Tannsjo & J. Ryberg (Eds.), *The repugnant conclusion: Essays on population ethics*. Dordrecht: Kluwer.

Huemer, M. (2008). In defence of repugnance. *Mind, 117*(468), 899–933.

Mulgan, T. (2006). *Future people: A moderate consequentialist account of our obligations to future generations*. Oxford: Oxford University Press.

Narveson, J. (1967). Utilitarianism and new generations. *Mind, 76,* 62–72.

Ng, Y.-K. (1989). What should we do about future generations? Impossibility of Parfit's theory X. *Economics and Philosophy, 5,* 135–253.

Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.

Rachels, S. (2004). Repugnance or intransitivity: A repugnant but forced choice. In T. Tannsjo & J. Ryberg (Eds.), *The repugnant conclusion: Essays on population ethics*. Dordrecht: Kluwer.

Roberts, M. (2015). Population axiology. In I. Hirose & J. Olson (Eds.), *The Oxford handbook of value theory*. New York: Oxford University Press.

Ryberg, J. (2004). The repugnant conclusion and worthwhile living. In T. Tannsjo & J. Ryberg (Eds.), *The repugnant conclusion: Essays on population ethics*. Dordrecht: Kluwer.

Tannsjo, T. (2004). Why we ought to accept the repugnant conclusion. In T. Tannsjo & J. Ryberg (Eds.), *The repugnant conclusion: Essays on population ethics*. Dordrecht: Kluwer.

Tannsjo, T. (2015). *Taking life: Three theories of the ethics of killing*. New York: Oxford University Press.

Tannsjo, T., & Ryberg, J. (2004). *The repugnant conclusion: Essays on population ethics*. Dordrecht: Kluwer.

Temkin, L. (1987). Intransitivity and the mere addition paradox. *Philosophy & Public Affairs, 16*(2), 138–187.

Wolf, C. (2004). O repugnance, where is thy sting? In T. Tannsjo & J. Ryberg (Eds.), *The repugnant conclusion: Essays on population ethics*. Dordrecht: Kluwer.