[Forthcoming in *The Oxford Handbook of Population Ethics*.]

Our Intuitive Grasp of the Repugnant Conclusion

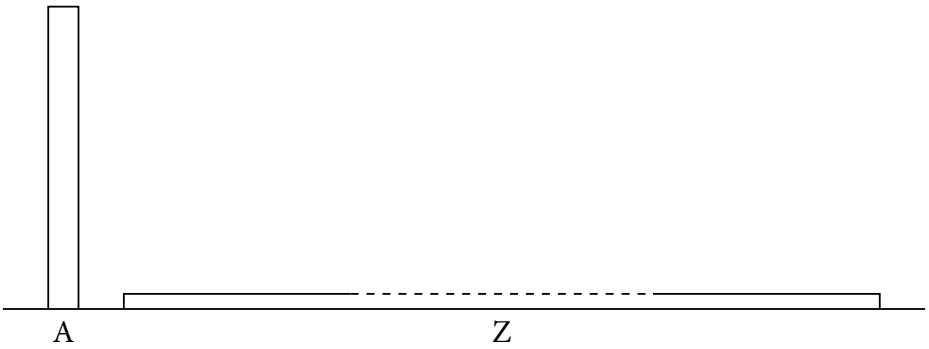
Johan E. Gustafsson*

The Repugnant Conclusion is a counter-intuitive implication of Total Utilitarianism. A compelling defence of the latter is that our intuitions are unreliable in this case because the Repugnant Conclusion involves very large numbers. This chapter surveys earlier proposals of this kind and proposes a variation based on the idea that our intuitions' unreliability is due to a slight insensitivity in our intuitive grasp of the relevant factors.

Consider

> *The Repugnant Conclusion*
> For any possible population of at least ten billion people, all with very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives barely worth living.[1]

The Repugnant Conclusion is often put in terms of a comparison between an *A* population—a large population where each person has very high well-being—and a *Z* population—a huge population where each person has positive but very low well-being. In the following figure, the populations are represented by boxes, where the width of a box represents the size of the population and the height represents their members' level of well-being.[2]



The Repugnant Conclusion says that for every population like A, there is a better population like Z.

As the name reflects, many people find the Repugnant Conclusion repugnant. This repugnance poses a significant challenge to views that entail the Repugnant Conclusion, such as, Total Utilitarianism:[3]

Total Utilitarianism
A first population is at least as good as a second population if and only if the sum total of well-being is at least as great in the first as in the second.[4]

We have that the Repugnant Conclusion is counter-intuitive and, if the Repugnant Conclusion is false, then so is Total Utilitarianism. This seems to be a compelling objection to Total Utilitarianism.

But should we trust our intuition that the Repugnant Conclusion is repugnant? If this intuition is used as evidence against Total Utilitarianism, we need to examine the reliability of this evidence. One might be able to defend Total Utilitarianism if one can explain why this intuition is unreliable and do so in a way that doesn't challenge the reliability of the main intuitions that supports Total Utilitarianism. Moreover, even if you happen accept Total Utilitarianism and also happen to find the Repugnant Conclusion plausible, this conclusion is still sufficiently unpopular to warrant an explanation.

In this chapter, I shall try to explain why our intuitions about the Repugnant Conclusion are unreliable as evidence against Total Utilitarianism (section 5). But, before I present my own proposal, I shall consider some alternative explanations (sections 2–4). These explanations, including my own, are variations of the idea that people's intuitions are misled by the large numbers involved in the Repugnant Conclusion. Finally, I shall defend explanations of this type from two objections in the literature (sections 6–7).

1. The Intuition of Neutrality

Before we go on, however, some words are needed on the intended scope of these explanations of why our intuitions about the Repugnant Conclusion are unreliable. Some people reject the Repugnant Conclusion because they accept *the Intuition of Neutrality*, which says that it is axiologically neutral whether a person with a life worth living is brought into existence. This intuition is memorably captured by Jan Narveson's slogan 'We are in favor of making people happy, but neutral about making happy people.'[5] Suppose that there are two populations P_1 and P_2, where Adam, Eve, and Cain exist in P_1 but only Adam and Eve exist in P_2, with well-being levels as follows:

	Adam	Eve	Cain
P_1	1	1	–
P_2	1	1	1

If one judges that P_2 is not better than P_1, one does not agree with one of

the basic tenets of Total Utilitarianism, namely, that people with positive well-being make the world better. And then one is unlikely to agree with the Repugnant Conclusion. Furthermore, since no large numbers are involved in the above example, worries about our ability to grasp large numbers are futile.

Yet I do not think that most people who find the Repugnant Conclusion repugnant do this because they accept the Intuition of Neutrality. If they did, there would be no need to involve large numbers in the Repugnant Conclusion to get a repugnant implication from Total Utilitarianism. Anyway, the attempts below, by me and others, to explain the unreliability of our intuitions about the Repugnant Conclusion are not intended to cover those who think it is repugnant because they accept the Intuition of Neutrality. These attempts might, however, do some work even for these people if they also think that there is something additionally repugnant about the Repugnant Conclusion, on top of their not being in favour of making happy people.

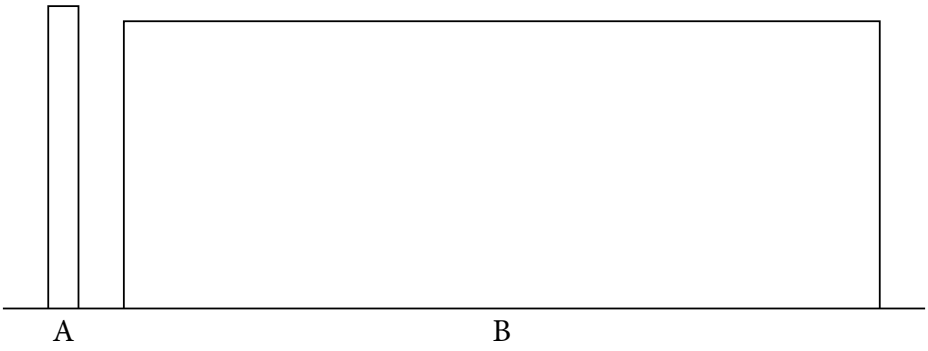
2. The Imaginative Alikeness of Large Numbers

Michael Huemer argues that our intuitions about the Repugnant Conclusion are unreliable because we are unable to imaginatively differentiate between large numbers. He claims that beyond a certain level, all large numbers are imagined roughly the same.[6] So, when we compare *A* and *Z*, which are both very large populations, we imagine their sizes the same way—that is, just as very large. Hence we compare them as follows:

A	*Z*
very large	very large
very high well-being	positive but very low well-being

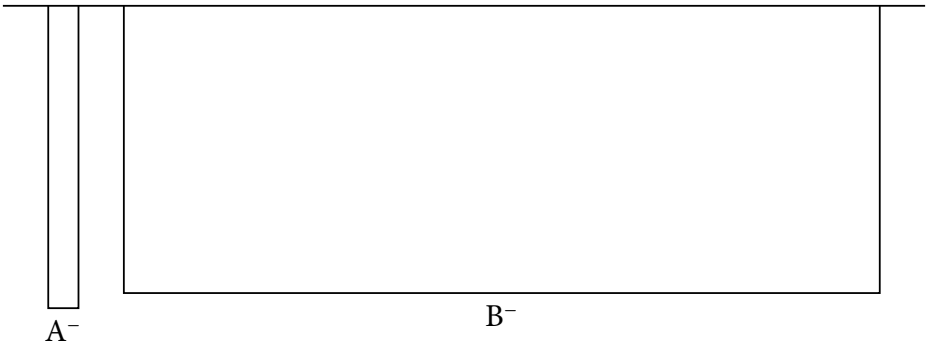
If this suggestion were correct, then the main feature that, according to Total Utilitarianism, makes *Z* better than *A* (that is, *Z*'s size) would be lost in the comparison. Hence it would not be surprising if our intuitions yielded the wrong answer.[7]

Yet, in some cases, it seems that we *can* take into account the relative sizes of very large populations.[8] Consider, for example, the comparison between a very large *A* population and a much larger *B* population in which people have only a slightly lower level of well-being:



In this case, I guess that most people intuitively judge that the size of *B* makes up for its lower level of well-being. And hence they judge that *B* is better than *A*.

Some people might have a clearer intuition about the corresponding negative populations, where people instead have negative levels of well-being. Consider the following negative variation of the last comparison of Parfit's case The Two Hells:[9]



Here, I guess that most people find B^- clearly worse than A^-. The fact that much less people suffer in A^- than in B^- makes up for the fact that the people in B^- suffer slightly less than the people in A^-. Hence it seems that for some comparisons of very large populations we can take into account their relative sizes.

Torbjörn Tännsjö offers a related account of how large numbers distort our intuitions, namely, that we have difficulties identifying with a large number of people. He claims that

> our actual moral sense seems to be based on identification. However, our capacity to identify with others is limited. Most of us care about our family, and those who are near and dear to us. We take less interest in our fellow countrymen but more interest in them than in people living far away from us. However, it is widely recognized that we *ought* to care about strangers. We ought to generalize our sympathy even to them. We have extra difficulties

> in doing so when it comes to very large numbers of people. Very large numbers *mean* very little to us. However, large numbers do matter. In the same manner that we generalize our sympathy to strangers we ought (mechanically, if necessary) to generalize our sympathy to large numbers of people, even to all the people living in Parfit's Z-world.[10]

If the objection here is that very large numbers mean very little to us and hence that they all mean roughly the same thing to us, then it is essentially the same as Huemer's imaginative-alikeness objection. And then it is unsuccessful for the same reasons.

On the other hand, the first part of the quote suggests instead that the objection is that it is hard to identify with the people in Z. The people in Z, we can assume, are neither fellow countrymen nor near and dear. They are anonymous members of a very large number of people, which makes them difficult to identify with. Yet the same holds for the people in A. Since the people in A and the people in Z are all anonymous members of very large populations, it seems that our difficulties having sympathy with large numbers of people would make it difficult for us to sympathize not only with the people in Z but also with those in A. And, if so, one would expect that we wouldn't intuitively judge one of A and Z, or one of A^- and B^-, as much better than the other. But those who find the Repugnant Conclusion repugnant seem to find A much better than Z and find B^- much worse than A^-.

3. Compounding small values

Huemer offers a further explanation of the unreliability of our intuitions about the Repugnant Conclusion, namely, that we are bad at compounding lots of small values. He claims that

> we find a tendency to underestimate the effect of compounding a small quantity. Of particular interest is our failure to appreciate how a very small value, when compounded many times, can become a great value. The thought that no amount of headache-relief would be worth a human life is an extreme instance of this mistake—as is the thought that no number of low-utility lives would be worth as much as a million high-utility lives.[11]

Huemer's proposal is based on the idea that we intuitively value the lives in a population one by one and then fail to properly compound their values to get the value of the population. But there are similarly repugnant cases where the compounding of the values of lives plays no role. Consider

> *The One-Person Repugnant Conclusion*
> For any possible life which is at all times of a very high quality, there is a better possible life which is at all times barely worth living.[12]

The One-Person Repugnant Conclusion can be put in terms of a comparison between an *A*-life—a long life that is at all times of very high quality and a *Z*-life—a very long life that is at all times worth living but barely so. The One-Person Repugnant Conclusion says that for every life like the *A*-life, there is a better life like the *Z*-life. Similarly to the regular Repugnant Conclusion, it is counter-intuitive that the *Z*-life would be better than the *A*-life. In the assessment of this one-person variant of the Repugnant Conclusion, the compounding of the values of lives plays no role, since we are just comparing individual lives. Hence, even if we were unable to reliably compound small values of a large number of lives, this would not explain why we have anti-utilitarian intuitions in one-person analogues of the Repugnant Conclusion.

Nevertheless, the compounding-small-values proposal might still be able to explain why our intuitions are unreliable about the One-Person Repugnant Conclusion. Any *Z*-life that is better on utilitarianism than some *A*-life must be extremely long, perhaps thousands of years long. Therefore, we might be unable to directly grasp the value of lives of such length, since we cannot intuitively imagine what living for that long would be like. So we have to make this intuitive comparison between a long *A*-life and a extremely long *Z*-life, not by intuitively valuing the whole of the *Z*-life, but instead by valuing shorter, more intuitively graspable intervals of the *Z*-life. And then we fail to properly compound the small values of the many short parts of the *Z*-life.

Yet consider a *Z*-life of the same length as the longest duration of life which we can grasp the value of without doing any compounding. Then consider a much shorter *A*-life with a just slightly less sum-total of well-being than the *Z*-life. Suppose, for the sake of the argument, that the *Z*-life would be a year long life that is at all times barely worth living and that the *A*-life would be a week long life that is at all times of a very high quality. Then it is still counter-intuitive that the *Z*-life would be better than the *A*-life. But in this case, since we can evaluate the whole *Z*-life without compounding the values of any shorter segments, the compounding-small-values approach does not apply.

4. Grasping large numbers

John Broome claims that

> we have no reason to trust anyone's intuitions about very large numbers, however excellent their philosophy. Even the best philoso-

> phers cannot get an intuitive grasp of, say, tens of billions of people. That is no criticism; these numbers are beyond intuition. But these philosophers ought not to think their intuition can tell them the truth about such large numbers of people.[13]

Broome's point, however, is not that we have unreliable intuitions about *every* principle that says something about very large populations. Some of the principles he wishes to rely on himself says something about populations of all sizes. For example, in his defence of a variant of Total Utilitarianism, he relies on

> *The Principle of Personal Good*
> Take two distributions A and B that certainly have the same population. If A is equally as good as B for each member of the population, then A is equally as good as B. Also, if A is at least as good as B for each member of the population, and if A is better than B for some member of the population, then A is better than B.[14]

Since the Principle of Personal Good quantifies over all populations, including very large ones, it says something about large populations. Still, Broome wishes to rely on its intuitive plausibility.

In response, Broome narrows his case against the reliability of our intuitions. Yet his example of an intuition that depends on large numbers is not the Repugnant Conclusion. It is the similarly structured claim that it is better to save someone from AIDS than to cure any number of people from headaches. Broome writes:

> I shall not try to formulate a general principle that distinguishes universally quantified intuitions we can rely on from those we cannot rely on. Instead, I shall identify a particular feature that makes some of these intuitions unreliable. The intuition about AIDS mentions a fixed event A and a variable event $B(n)$ that depends on the number of people. The fixed event is curing one person of AIDS; the variable event is curing n people of a short mild headache. The intuition has the form: for all numbers n, A is better than $B(n)$. An intuition of this form is exposed to doubt because the goodness of $B(n)$ may increase with increasing n. It does so in this case. The intuition is that, although $B(n)$ gets better and better with increasing n, it never gets better than A, however large n might be. This sort of intuition particularly depends on our intuitive grasp of large numbers. So it is unreliable.[15]

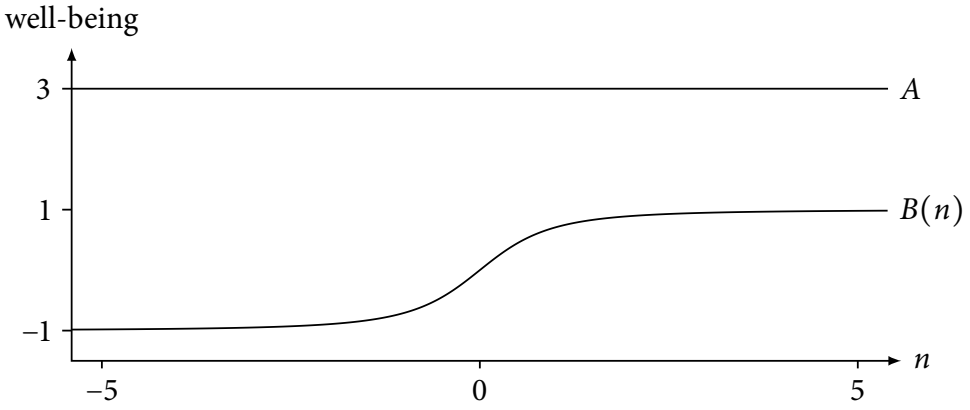
This narrowed proposal, however, still rules out too much. There are reliable intuitions of the form Broome wishes to discredit. Let, for example, A and $B(n)$ be distributions for the same population, such that

A has everyone at well-being level 3

and

$B(n)$ has everyone at well-being level $\frac{n}{\sqrt{1+n^2}}$

The following graph illustrates the relation between the well-being of the people in the population given distributions A and $B(n)$:



Plausibly, $B(n)$ gets better as n increases, since—other things being equal—it is better if everyone has a higher level well-being. Similarly, it seems intuitively that, for all numbers n, A is better than $B(n)$, since everyone has well-being level 3 in A whereas in $B(n)$ they have, for all numbers n, a well-being level below 1. This intuition seems neither unreliable nor particularly dependent on my intuitive grasp of large numbers. So it seems that, if an intuition is unreliable, it is not because it is of the above form. Moreover, the Principle of Personal Good entails that A is better than $B(n)$ for all numbers n. If we have reliable intuitions about the Principle of Personal Good, we should also have reliable intuitions about this logically weaker claim, which is of the form Broome exposes to doubt. Hence Broome's narrowed proposal is not entirely successful either.

One might object that this objection can be sidestepped if we just restrict the n parameter in Broome's proposal to population sizes. This move, however, cannot explain why our intuitions are unreliable in the One-Person Repugnant Conclusion, since in that case the population-size parameter is fixed. Moreover, this move seems a bit ad hoc unless we can explain why our intuition can reliably grasp very large numbers of some things but not reliably grasp very large numbers of people.

5. Extreme trade-offs and margins of error

While the previous proposals seek to explain the unreliability of our intuitions about the Repugnant Conclusion, I have a slightly less general aim. My aim is merely to argue that our intuitions about the Repugnant Conclusion are unreliable as evidence against Total Utilitarianism. So my proposal need not rule out that one's intuition that the Repugnant Conclusion is repugnant is reliable as evidence against some theories that yield this conclusion. I shall argue that it is likely that we would have the intuition that the Repugnant Conclusion is false even if Total Utilitarianism were true. Hence, if I am right, these intuitions are unreliable as evidence against Total Utilitarianism.[16]

My diagnosis is that it's not just the largeness of the numbers in the Repugnant Conclusion that causes our problems. Rather, the problem is making trade-offs where the relevant factors are extremely proportioned in opposite ways. The proposal's main assumption is that our intuitive understanding of morally relevant factors is inexact and comes with a slight margin of error.

To illustrate the basic idea, consider measuring two rectangular areas. The first is a football pitch. We measure the length of the pitch to be 100 m and its width to be 68 m. The second is one side of an extremely long and very narrow tape. We measure the length of tape to be 3000 km and its width to be just 2 mm. We get that the area of the pitch rectangle is 7140 m^2 and the area of the tape rectangle is 6000 m^2. Hence we get that the pitch rectangle has a much larger area than the tape rectangle. This conclusion, however, is only reliable if our measurement of the tape rectangle's width has a margin of error no greater than slightly below 1 mm. If our measurements are insensitive to a difference in width of just 1 mm, we cannot rule out that the width of the tape rectangle is 3 mm, which would make the area of the tape rectangle larger than 7140 m^2 (it would be 9000 m^2). And then we cannot rule out that the tape rectangle has a larger area than the pitch rectangle. So, even though it seems like the pitch rectangle's area is a lot larger than the tape rectangle's area and there is just a very minor margin of error for the width of the tape rectangle, this tiny margin of error is enough to make the comparison unreliable. This is because any error in the measurement of the width of the tape rectangle is multiplied by the rectangle's very great length, resulting in a huge margin of error for the calculated area. In this manner, I propose that a very slight dullness in the sensitivity of our intuitive understanding of the relevant factors in an intuitive comparison can have a crucial effect on the comparison's reliability.

I propose that

our intuitions about a comparative claim between x and y are unreliable as evidence against a theory T if

(i) there is, according to T, a trade-off between two relevant factors in the comparison and

(ii) there is a possible change in one of these two factors such that the change is small enough to fall within our intuitions' margin of error and

(iii) this change would make a difference to whether the comparative claim holds according to T.

The inexactness of our intuition might result in a great mismatch between a correct theory and our intuitive judgement if some change that falls within our intuition's margin of error can make a great difference to how the theory compares the options. This, I shall argue, is what is going on in case of Total Utilitarianism and the Repugnant Conclusion.

One class of comparisons that seem to fulfil (i)–(iii) given a theory T are comparisons between two options such that T evaluates the options by the product of two relevant factors F and G and one option is very little F and very much G and the other option is much F and little G. In comparisons of this kind, there is a straightforward explanation of why our intuitions are unreliable. The problem is that, when options are ranked by the product of F and G and one option has very little F and very much G, then very small variations in factor F will have a very large effect on the product of F and G. Hence a very small change in factor F might change which option has the greatest product of F and G. But, because our intuition is inexact, it might plausibly be insensitive to this small change. Thus, because of the inexactness in our intuitive understanding of factor F, our intuitions might deviate from T in these cases. Nevertheless, this inexactness doesn't rule out that our intuitions reliably track T in cases without this kind of extreme trade-offs.

It is crucial that the factors are proportioned in opposite ways for the two options, since if one option dominated the other in both of the relevant factors even by a tiny amount, one would reach the right judgement by mere dominance reasoning, without taking into account the products of these factors.

In the Repugnant Conclusion, there are two factors opposed in this way according to Total Utilitarianism. We have to compare the high well-being and few people in A—that is, few relative to the size of Z—with the very low welfare and huge number of people of Z. Hence we have two options with alternately very much and very little of two relevant factors such that Total Utilitarianism evaluates the options by the product of these factors. Note that, for every Z population that Total Utilitarianism

ranks as better than *A*, there is a population Z^* of the same size as *Z* but where people have an even lower but still positive well-being, such that Total Utilitarianism ranks Z^* as worse than *A*.[17] The people in *Z* and Z^* have very similar well-being, and they all have lives 'barely worth living'. And, when we intuitively compare *Z* and Z^* to *A*, our intuition is not sensitive enough to take into account the small difference between the lives in *Z* and those in Z^*. Since we cannot intuitively take into account the exact level of well-being in Z, which is crucial for Total Utilitarianism's ranking of *A* and *Z*, it seems that our intuitions about these cases are unreliable as evidence against Total Utilitarianism.[18,19]

With my proposal, unlike some of the earlier ones, there's no need to claim that our intuitive thinking departs from Total Utilitarianism—for example, by failing to compound small values. On my proposal, our intuition may very well follow the total-utilitarian principle. But there is a slight margin of error in our intuitive judgement of the parameters of the principle, which gets greatly amplified in cases where the principle multiplies two parameters such that one is a very small quantity and the other is a very large quantity.

It may be objected that my proposal would seem to predict that our intuitive evaluations of the *A* and the *Z* populations should be thoroughly equivocal, whereas there is widespread intuitive agreement that *Z* isn't better than *A*. That is, my proposal fails to explain the observed systematicity in the alleged failures of our intuitions about the Repugnant Conclusion.

But my above proposal is just an account of the *reliability* of our intuitive judgements; it is not an account of what intuitive judgements we do make. I think, however, that there is an explanation of the systematicity in our intuitive judgements in the these cases. Note that, in all of the discussed cases, our intuitive judgement seems to disregard the part of an alternative's good or bad features which would, according to my account, be within our intuition's margin of error. This suggests that, when we have trouble getting an exact intuitive grasp of a certain feature, we only take into account the amount of that feature which we clearly grasp is there. When we asses *A*, even though we do not have an exact grasp of the amount of well-being in an imagined life in *A*, we can still clearly grasp that it is at least above a fairly high amount. So, to a large extent, we take the high quality of the lives in *A* into account. Yet, when we asses *Z*, there is no positive amount of well-being such that we can clearly grasp that there is at least that amount of well-being in an imagined life in *Z*. This is because the positive but very small amount of well-being in those lives is within our intuition's margin of error. So we fail to take into account the positive quality of the lives in *Z*.[20]

It may next be objected that my proposal seems a little ad hoc if it

only explains why we have unreliable intuitions about the Repugnant Conclusion.[21]

Yet it also explains many other cases where otherwise intuitive theories yield counter-intuitive implications. A further counter-intuitive implication of Total Utilitarianism is

> *Hangnails for Torture*
> For any excruciatingly painful torture session lasting for at least two years to be experienced by one person, there is some large number of minute-long very mildly annoying hangnail pains, each to be experienced by a separate person, that is, other things equal, worse.[22]

Here, we have another case where the options score alternately very well and very badly in terms of two factors. We have a great loss in well-being for just one person—that is, the torture lasting at least two years—compared to a very small loss in well-being for a very large number of people—that is, the minute-long hangnail pains. And Total Utilitarianism evaluates the options by the product of these factors. Hence my proposal also applies to our intuitions about Hangnails for Torture.

My proposal also explains why our intuitions in the One-Person Repugnant Conclusion are unreliable as evidence against utilitarianism. We have one life that is very long with very high quality of life which is compared to a much longer life with a barely positive quality of life. And utilitarianism evaluates these lives by the product of their quality and their length.

Furthermore, my proposal does not rule out that we have reliable intuitions about the Principle of Personal Good, which caused Broome some concern. If, given the same population, a first distribution strongly dominates a second distribution in the sense that the first is at least as good for everyone and better for someone than the second, then there is no relevant better-making factor according to Total Utilitarianism in which the second beats the first. So my proposal does not entail that our intuitions about the Principle of Personal Good are unreliable as evidence in favour of Total Utilitarianism.

And, unlike Huemer's imaginative-alikeness proposal, my proposal does not rule out that we have reliable intuitions about the comparisons of A and B and of A^- and B^-. In these comparisons, there are trade-offs between population size and level of well-being, but these trade-offs are not extreme; no small change in either population size or level of well-being would change that B is better than A or that B^- is worse than A^- according to Total Utilitarianism.

6. Imagining Very Many Lives

Finally, we shall look at two general objections to this kind of defence of Total Utilitarianism. The first is due to Derek Parfit who responds to the worry that we might have trouble imagining the relevant populations involved in the Repugnant Conclusion. He claims that

> We can imagine what it would be for someone's life to be barely worth living. And we can imagine what it would be for there to be many people with such lives. In order to imagine *Z*, we merely have to imagine that there would be *very* many. This we can do.[23]

Yet, it's not merely the *Z* population that might be arbitrarily large, the *A* population might be so too. If *A* is sufficiently large, it would have to contain very many people. But, in that case, the lives in a relevant *Z* population would have to be imagined as being much more numerous than as being very many. Because otherwise our imagination wouldn't do justice to the much larger size of *Z* compared to *A*, which is *Z*'s main advantage according to Total Utilitarianism. Hence it is insufficient to just be able to imagine *Z* as very many lives barely worth living. We have to be able to take into account how much larger the *Z* population is even when the *A* population is huge.[24]

Moreover, even if we only consider cases where *A* consists of just ten billion people, there is another problem with Parfit's reply. The sum total of well-being in a population of ten billion people with a very high quality of life is very large. Hence, even on Total Utilitarianism, there are populations worse than *A* but consisting of very many lives barely worth living. So, if one could merely imagine *Z* as very many lives barely worth living, one couldn't be sure that one imagines a population that is better than *A* according to Total Utilitarianism.

7. The Extrapolation Argument

The second general objection is due to Theron Pummer. He argues that to defend counter-intuitive principles involving large numbers, like the Repugnant Conclusion, it is not enough to show that we have unreliable intuitions about large-number cases. He claims that, even if we have unreliable intuitions about large-number cases, we can extrapolate from small-number cases where we do have reliable intuitions. Instead of the Repugnant Conclusion, however, he focuses on the similarly structured Hangnails-for-Torture claim.

Pummer claims that we normally have reason to believe something if we have reason to believe that it would seem true under ideal circumstances. He argues that, if we had reliable intuitions about large-number

cases, we would find Hangnails for Torture counter-intuitive. That is, he argues in favour of

> (1) If we could relevantly imagine any number of mild hangnail pains, we would have the intuition that there is *no* number of such pains such that it is worse than 2 years of excruciating torture.

rather than

> (2) If we could relevantly imagine any number of mild hangnail pains, we would have the intuition that there is *some* number of such pains such that it is worse than 2 years of excruciating torture.[25]

For the Extrapolation Argument, Pummer introduces a variable version of Hangnails for Torture, where the number of years with hangnail pains is given by a variable:

> *The Variable Claim*
> Two years of excruciating torture is worse than X years of very mildly annoying hangnail pains (each a minute long), other things being equal.[26]

Pummer holds that we do not become less confident in the variable claim the larger we imagine X to be and that this counts in favour of (1) rather than (2). The Extrapolation Argument runs as follows:

> (3) If (2) were true, and thus if (1) were false, then we would become less confident in the Variable Claim, the larger we imagine X to be.
>
> (4) We do not become less confident in the Variable Claim, the larger we imagine X to be.
>
> So, (1) is true and (2) is false.[27]

While we might also question (4), I shall argue that (3) is false.[28] For small values of X, which are the ones Pummer tries to extrapolate from, I don't think that, if (2) were true, we would become less in the Variable Claim as X increases. To see why, consider first the utilitarian principle that the value of an option is proportional to the option's sum total of well-being. According to this principle, the torture is much worse than the hangnails for all relatively small values of X. So, even if utilitarianism and thus Hangnails for Torture were true, the value of the hangnails would never get close to the value of the torture for all small values of X. Now, let us assume, following Pummer, that we have reliable intuitions

about small-number cases. Then, I think one would plausibly remain fully confident in the Variable Claim for all relatively small values of X even if utilitarianism and Hangnails for Torture were true. This is because one wouldn't only evaluate the torture as worse than the hangnails, one would evaluate the torture as *much* worse than the hangnails for all relatively small values of X. Hence, if utilitarianism were true, then, even allowing for a wide margin of error, one would plausibly remain fully confident in the Variable Claim for all relatively small values of X. And the same would hold for any theory that evaluates torture and hangnails in roughly the same way as utilitarianism.

But, if (2) were true, it seems that some theory would be true that evaluates torture and hangnails in roughly the same way as utilitarianism. And then we have that if (2) were true, we wouldn't become less confident in the Variable Claim, the larger we imagine X to be. Thus (3) seems false. Hence the Extrapolation Argument is unconvincing.

Summing up, I have argued that the intuition that the Repugnant Conclusion is repugnant is not reliable evidence against Total Utilitarianism. My explanation for this is that our intuitive understanding of morally relevant factors is inexact and comes with a slight margin of error, which makes are intuitive judgement about cases with extreme trade-offs unreliable. This explanation has some advantages over earlier explanations of our intuitions' unreliability in large-number cases. Finally, I have defended explanations of this kind against some recent objections.

I wish to thank Gustaf Arrhenius, Campbell Brown, Krister Bykvist, Timothy Campbell, Marc Fleurbaey, Martin Peterson, Douglas W. Portmore, Daniel Ramöller, Jussi Suikkanen, Folke Tersman, Nicolas Olsson-Yaouzis, and the audiences at the Population Ethics Workshop, Eindhoven University of Technology, March 26, 2013, at the Higher Seminar in Practical Philosophy, Stockholm University, September 10, 2013, and at ISUS XIII 2014, Yokohama National University, August 21, 2014, for valuable comments.

Notes

[*]I would be grateful for any thoughts or comments on this paper, which can be sent to me at *johan.eric.gustafsson@gmail.com*.

[1]Parfit (1984, p. 388).

[2]Note that in this illustration the area of the box representing Z is indeed larger than the box representing A. In Parfit's (1984, p. 388) original illustration, the area of the A box is 1.6 times as large as the area of the Z box. Although the Z box's borders are partly dashed indicating an arbitrarily large size, this makes the canonical illustration of the Repugnant Conclusion visually misleading.

[3]Parfit (1984, p. 388). As far as I know, Sidgwick (1907, pp. 415–416) was the first to point out that Total Utilitarianism has this implication. Apart from his complaint that this type of conclusion is too exact for common-sense morality, Sidgwick (1907, p. 416) interestingly does not seem to find the Repugnant Conclusion repugnant. McTaggart (1927, pp. 452–453) claims that many moralists would find the Repugnant Conclusion repugnant, but he does not himself find it repugnant. He sees 'no reason for supposing that repugnance in this case would be right.' The repugnance rests, he thinks, on a mistaken conviction, which he does not share. For a list of some early sources of the Repugnant Conclusion, see Arrhenius (2000, p. 40n) who this general idea back to Whewell (1852, pp.237–238). But, as I shall show below, the distinction between average and total views were know before that. Of the classical utilitarians, I think that only Mill would have found the Repugnant Conclusion repugnant. Mill's (1965, p. 756) following views on population growth suggest average rather than total utilitarianism:

> There is room in the world, no doubt, and even in old countries, for a great increase of population, supposing the arts of life to go on improving, and capital to increase. But even if innocuous, I confess I see very little reason for desiring it. The density of population necessary to enable mankind to obtain, in the greatest degree, all the advantages both of cooperation and of social intercourse, has, in all the most populous countries, been attained. A population may be too crowded, though all be amply supplied with food and raiment. […] If the earth must lose that great portion of its pleasantness which it owes to things that the unlimited increase of wealth and population would extirpate from it, for the mere purpose of enabling it to support a larger, but not a better or a happier population, I sincerely hope, for the sake of posterity, that they will be content to be stationary, long before necessity compels them to it.

Bentham, on the other hand, seems to have been a total rather than average utilitarian; see Gustafsson (2018, p. 99n18). First of all he was aware of the distinction; Bentham (1952–1954, vol. 3, p. 318) writes

> Opulence, though so nearly of kin to wealth, or rather for that very reason, requires to be distinguished from it: opulence is *relative* wealth, relation being had to population: it is the ratio of wealth to population. Quantity of wealth being given, the degree of opulence is therefore not directly, but inversely, as the population, i.e. as the degree of populousness—as the number of those who are to share in it: the fewer the shares, the larger is each one's share.

Bentham (1839, p. 228) argues that it is better to distribute wealth so that 10,000 already existing rich people get richer than to use the money to give existence to some very poor people. Here, money is used as a convenient substitute for happiness—see Bentham (1998, p. 252) and Schofield (2006, p. 43). Bentham's argument for this is not that creating poor people would lower the average level of happiness but instead that

> a greater addition the to the aggregate quantity of happiness would be made by dividing among the first 10,000 the whole additional quantity of wealth, than by making any addition to the number of persons brought into existence.

This is because he thinks it is likely that the poor people who would live at the minimum of subsistence would soon suffer some accident and pass away.

[4]I am here following the usage of Arrhenius (2000, p. 39). In Broome (2004, p. 138), this axiological component of utilitarianism is called the total principle. Parfit (1984,

p. 387) calls a similar principle the impersonal total principle.

[5]Narveson (1973, p. 80).

[6]Huemer (2008, p. 908).

[7]Compare Greene (2001) who similarly objects that

> the fact that we are able to more or less fully appreciate the "quality of life" benefits with [A] and unable to fully appreciate the "quantity of life" benefits that come with [Z] may cause us to overestimate the repugnance of the repugnant conclusion.

[8]To see that we can take some factor in to account when we make comparisons, find a case where we reach a different verdicts in a comparison depending solely on a change in this factor.

[9]This is a variation of Parfit's (1984, p. 406) case The Two Hells, where the smaller population consists of just ten people, whereas in my variant both populations consist of at least ten billion people.

[10]Tännsjö (2002, p. 344).

[11]Huemer (2008, pp. 909–910).

[12]Parfit (1986, p. 169) presents a similar one-person variant of the Repugnant Conclusion. His variant, however, has the drawback of involving infinity, which is known to easily mislead intuition.

[13]Broome (2004, p. 57).

[14]Broome (2004, p. 120).

[15]Broome (2004, pp. 58–59).

[16]Let R be that we have the intuition that the Repugnant Conclusion is false, and let U be that Total Utilitarianism is true. We note that

$$P(U \mid R) = \frac{P(R \mid U) \cdot P(U)}{P(R \mid U) \cdot P(U) + P(R \mid \neg U) \cdot P(\neg U)}$$

given that $P(R) \neq 0$ and $P(U) \neq 0$. For example, suppose that $P(R \mid U) = 0.8$ and your prior credence in U is 0.8, then your posterior credence in U given R should be

$$P(U \mid R) = \frac{3.2}{3.2 \cdot P(R \mid \neg U)}$$

At worst, if $P(R \mid \neg U) = 1$, we have that $P(U \mid R)$ will be roughly 0.76. And, if there is a chance (as seems plausible) that we would have the intuition that the Repugnant Conclusion is false even if Total Utilitarianism is false, then $P(U \mid R)$ will be even higher. For example, suppose that also $P(R \mid \neg U) = 0.8$. Then R should have no effect on our credence in U, that is, $P(U \mid R) = P(U) = 0.8$. Hence, if $P(R \mid U)$ is high, R cannot be reliable evidence against U.

[17]I am assuming here that there is no lowest possible positive level of well-being. Hence I am assuming denseness for well-being levels rather than discreteness. See Arrhenius (2000, p. 163). If one gives up this this assumption, one might get around my objection by imagining a Z population of lives with the lowest positive level of well-being. But this would not help much unless one is clear about which well-being level is the lowest positive one, which seems implausible. If one does not know what lowest positive level of well-being is, one cannot be sure that one imagines a population of lives with that level of well-being.

[18]Note that I do not claim that one cannot take the difference in the level of well-being between Z and Z^* into account when one compares them with each other, since that is a simple case of dominance, without trade-offs.

[19]It may be objected that, if we allow that the *Z* population can be infinitely large, then it should be sufficient to judge that the well-being of the people in *Z* is positive in order to judge that it will be better than any finite population. But, if we allow infinite populations, then the *A* population may be infinitely large and hence infinitely good and then there will be no *Z* population such that it is better than *A*; so the Repugnant Conclusion would not follow from Total Utilitarianism. To sidestep this issue, however, we can consider a weakened version of the Repugnant Conclusion where the *A* population must be finite. A first reply is that our intuitions about infinity are known to be unreliable. So, if our intuition that the Repugnant Conclusion is repugnant depends on crucially on cases involving infinity, then it is unreliable. A second reply is that, if the is a margin of error in out intuition, then we could not reliably tell whether we imagine an infinite population with barely good lives or an infinite population with barely not good lives.

[20]One might be unconvinced. Suppose we stipulate that each person in *Z* has a life that is slightly better than a life in which the only conscious experience is the short-lived pleasure of a single lick of a lollipop (perhaps the people in *Z* get two licks rather than just one). Since we all know what it is like to lick a lollipop, this description of the case gives us a clear idea that there is at least that amount of well-being (that is, one lollipop lick's worth) in an imagined life in *Z*. Yet, one might doubt that this will have any significant impact on people's intuitions about which population is better in this case. (I thank Timothy Campbell for raising this objection.) Note, however, that, though you might know what a life consisting one lollipop lick would be like, you do know what the exact well-being (that is, personal value) would be for that life. When you asses the well-being of the lives in *Z*, the dullness of our intuition comes in (and brings about a margin of error for the well-being level), and it does so even if you could imagine exactly what their lives would be like in all non-evaluative respects.

[21]Similarly, Temkin (2012, p. 122) worries about a related attempt to explain away our intuitions about a version of the One-Person Repugnant Conclusion:

> To be sure, advocates of additive aggregation may continue to insist that our intuitions about such cases are not to be trusted, even for cases involving intuitively graspable numbers. So, for example, they might insist that the difference between oyster-like lives of 100, 1,000, and 10,000 years is on a trajectory that would eventually amount to a great difference, but that the slope of the trajectory is so slight that we don't intuitively notice it, or perceive its long-term implications. But though it is possible such a view is correct, it has the air of an untestable article of faith that advocates of additive aggregation are compelled to invoke to explain away our intuitions, and I doubt that many will find it sufficiently compelling to alter their judgments about such cases.

Since my proposal applies not just to the Repugnant Conclusion but to any similar structured case, it makes some predictions that are to some extent testable. It predicts that there will be a lot of otherwise plausible theories *T* that yields counter-intuitive implications in cases where (i)–(iii) hold. Moreover, the possibility that the proposal, even if correct, would not compel many to alter their judgements about the Repugnant Conclusion seems irrelevant if our aim is not popularity but truth.

[22]Pummer (2013, p. 37). The example is a variation of Temkin (1996, p. 179) and Rachels (1998, p. 73). This explanation works equally well, changing what needs to be changed, for the similar lollipops-for-life case in Temkin (2012, p. 34).

[23]Parfit (1984, p. 389).

[24]Greene (2001) similarly objects that

> Certainly we can imagine a very large number of people living such

> lives, but can we effectively imagine the *difference between* two very large numbers of people living such lives? This is what our task requires, and it's not at all clear that we are up to it.

[25]Pummer (2013, p. 41). In this and the following quotes, I have changed the numbering.

[26]Pummer (2013, p. 41).

[27]Pummer (2013, p. 42). Pummer's argument for (3) is, more or less, a restatement of (3). He writes:

> (3) seems plausible because if (2) were true and thus there were some value for X, call it *n*, such that if we imagined *n* years of mild hangnail pains we would have the intuition that together they are worse than 2 years of excruciating torture, it seems we would gradually lose confidence in the Variable Claim as our imagined value for X gets closer to *n*. This seems true even if *n* were very large; we would presumably lose at least *some* confidence as X gets larger, if (2) were true.

[28]One problem with (4) is that if you have some moral uncertainty with at least some positive credence in Total Utilitarianism, it seems that you should become slightly less confident in the Variable Claim the larger you imagine *X* to be. (I thank Timothy Campbell for this point.

References

Arrhenius, Gustaf (2000) *Future Generations: A Challenge for Moral Theory*, Ph.D. thesis, Uppsala University.

Bentham, Jeremy (1839) *The Works of Jeremy Bentham, Published under the Superintendence of His Executor, John Bowring*, vol. III, Edinburgh: William Tait.

— (1952–1954) *Jeremy Bentham's Economic Writings: Critical Edition Based on His Printed Works and Unpublished Manuscripts*, ed. W. Stark, London: George Allen & Unwin, 3 vols.

— (1998) *'Legislator of the World': Writings on Codification, Law, and Education*, eds. Philip Schofield and Jonathan Harris, The Collected Works of Jeremy Bentham, Oxford: Clarendon Press.

Broome, John (2004) *Weighing Lives*, Oxford: Oxford University Press.

Greene, Joshua D. (2001) 'A Psychological Perspective on Nozick's Experience Machine and Parfit's Repugnant Conclusion', Presentation at the Society for Philosophy and Psychology Annual Meeting, Cincinnati, OH.

Gustafsson, Johan E. (2018) 'Bentham's Binary Form of Maximizing Utilitarianism', *British Journal for the History of Philosophy* 26 (1): 87–109.

Huemer, Michael (2008) 'In Defence of Repugnance', *Mind* 117 (468): 899–933.

McTaggart, John McTaggart Ellis (1927) *The Nature of Existence Volume II*, Cambridge: Cambridge University Press.

Mill, John Stuart (1965) *Principles of Political Economy with Some of Their Applications to Social Philosophy: Books III-V and Appendices*, vol. III of *Collected Works*, Toronto: University of Toronto Press.

Narveson, Jan (1973) 'Moral Problems of Population', *Monist* 57 (1): 62–86.

Parfit, Derek (1984) *Reasons and Persons*, Oxford: Clarendon Press.

— (1986) 'Overpopulation and the Quality of Life', in Peter Singer, ed., *Applied Ethics*, pp. 145–164, Oxford: Oxford University Press.

Pummer, Theron (2013) 'Intuitions about Large Number Cases', *Analysis* 73 (1): 37–46.

Rachels, Stuart (1998) 'Counterexamples to the Transitivity of *Better than*', *Australasian Journal of Philosophy* 76 (1): 71–83.

Schofield, Philip (2006) *Utility and Democracy: The Political Thought of Jeremy Bentham*, Oxford: Oxford University Press.

Sidgwick, Henry (1907) *The Methods of Ethics*, London: Macmillan, seventh edn.

Tännsjö, Torbjörn (2002) 'Why We Ought to Accept the Repugnant Conclusion', *Utilitas* 14 (3): 339–359.

Temkin, Larry S. (1996) 'A Continuum Argument for Intrasitivity', *Philosophy and Public Affairs* 25 (3): 175–210.

— (2012) *Rethinking the Good*, Oxford: Oxford University Press.

Whewell, William (1852) *Lectures on the History of Moral Philosophy*, Cambridge: Cambridge University Press.