

*Our Intuitive Grasp of the Repugnant Conclusion**

Johan E. Gustafsson[†]

ABSTRACT. According to Total Utilitarianism, a first outcome is at least as good as a second outcome if and only if the sum total of well-being in the first is at least as great as the sum total of well-being in the second. Total Utilitarianism implies the Repugnant Conclusion, which says that, for every population consisting of ten billion people with very high well-being, there is a population consisting of a very large number of people whose existence would be better even though its members have lives that are barely worth living. The Repugnant Conclusion is counter-intuitive. But is this intuition reliable as evidence against Total Utilitarianism? A potential defence of Total Utilitarianism is that our intuition is unreliable in this case because the Repugnant Conclusion involves very large numbers. This chapter considers some earlier proposals of this kind and proposes a variation based on the idea that the unreliability of our intuition about the Repugnant Conclusion is due to a slight insensitivity in our intuitive grasp of the morally 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

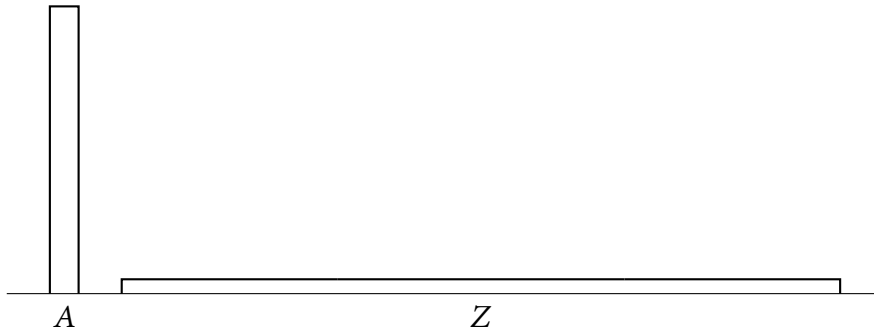
The Repugnant Conclusion is often put in terms of a comparison between an outcome *A*—which has a large population where each member has very high well-being—and an outcome *Z*—which has a huge population

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

[†] 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. Parfit's (1982, p. 142) earlier formulation contained, in addition, the normative claim that we ought to bring about the larger population.

where each member has positive but very low well-being. In the following figure, the outcomes are represented by boxes, where the width of a box represents the size of the outcome's population and the height represents the well-being of its members:



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

While its name is tendentious, many people do find the Repugnant Conclusion repugnant. This repugnance poses a significant challenge to views that entail the Repugnant Conclusion, such as, Total Utilitarianism:²

² Parfit 1984, p. 388. Sidgwick (1907, pp. 415–416) was the first to point out that Total Utilitarianism has this implication. Apart from his complaint that this conclusion is too exact for common-sense morality, Sidgwick (1907, p. 416) doesn't seem to find the Repugnant Conclusion repugnant. McTaggart (1927, pp. 452–453) claims that many *other* moralists would find the Repugnant Conclusion repugnant, but *he* sees 'no reason for supposing that repugnance in this case would be right.' The alleged repugnance rests, he thinks, on a mistaken conviction. For a list of some early sources of the Repugnant Conclusion, see Arrhenius 2000, p. 40n who traces this general idea back to Whewell 1852, pp.237–238. But the distinction between average and total views were known long before that. Bentham was clear about the distinction and endorsed a total view; see Gustafsson 2018, p. 99n18. Of the classical utilitarians, it seems that only Mill would have found the Repugnant Conclusion repugnant. Mill's (1965, p. 756) following remarks on population growth suggest an average, rather than total, view:

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 unlim-

Total Utilitarianism A first outcome is at least as good as a second outcome if and only if the sum total of well-being in the first is at least as great as the sum total of well-being in the second.³

Since the Repugnant Conclusion is counter-intuitive, we seem to have evidence that the Repugnant Conclusion is false. And, if the Repugnant Conclusion is false, then so is 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 should examine the reliability of this evidence. We may be able to defend Total Utilitarianism from the charge of repugnance if we can explain why this intuition is unreliable (and do so in a way that does not challenge the reliability of the main intuitions that support Total Utilitarianism).

In this chapter, I shall try to explain why our intuition about the Repugnant Conclusion is unreliable as evidence against Total Utilitarianism (section 5).⁴ 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 very large numbers involved in the Repugnant Conclusion. Finally, I shall also answer two general objections to explanations of this kind (sections 6–7).

ited 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.

Yet some of Mill’s (1972, p. 1414) later remarks suggest a total view:

when I said that the general happiness is a good to the aggregate of all persons I did not mean that every human being’s happiness is a good to every other human being; though I think, in a good state of society & education it would be so. I merely meant in this particular sentence to argue that since A’s happiness is a good, B’s a good, C’s a good, &c., the sum of all these goods must be a good.

³ Here, I am following the usage in 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.

⁴ For a discussion of the reliability of our intuition about the Repugnant Conclusion from the perspective of the psychological literature, see Mogensen forthcoming.

1. The Intuition of Neutrality

Before we go on, however, some words may be needed on the intended scope of these attempts to discredit our intuition. Some people reject the Repugnant Conclusion because they accept *the Intuition of Neutrality*, which says that it's axiologically neutral whether a person with a life worth living is brought into existence. This intuition is memorably captured by the slogan 'We are in favor of making people happy, but neutral about making happy people.'⁵

Consider, for instance, a first outcome, where only Adam and Eve exist, and they have the same high quality of life. And consider a second outcome, which is just like the first except that, in addition, Cain exists with the same high quality of life as Adam and Eve. If one judges that the second outcome is not better than the first, then one disagrees with a basic tenet 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, regardless of any repugnance. Furthermore, since this example only involves small numbers, worries about our ability to grasp large numbers do not apply here.

Yet I don't think that most people who find the Repugnant Conclusion repugnant do so 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 intuition about the Repugnant Conclusion aren't intended to cover those who reject the Repugnant Conclusion because they accept the Intuition of Neutrality. Still, these attempts may, to some extent, apply to these people too if they think that there's something additionally repugnant about the Repugnant Conclusion, on top of their not seeing any value in making happy people.

2. The imaginative likeness of large numbers

Michael Huemer argues that our intuition about the Repugnant Conclusion is 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*, we

⁵ Narveson 1973, p. 80.

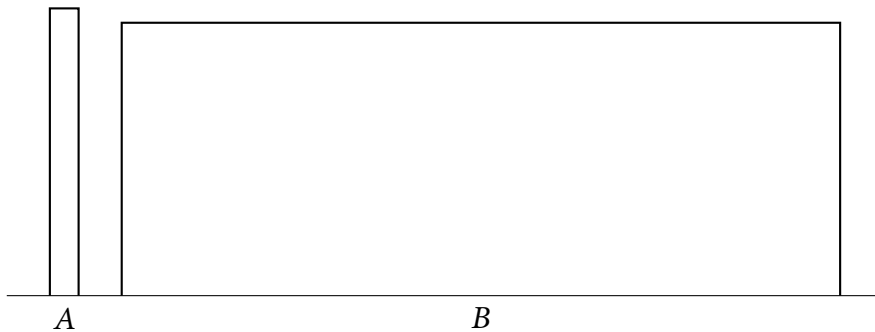
⁶ Huemer 2008, p. 908.

imagine their population sizes the same way—that is, just as very large. Hence we compare them as follows:

<i>A</i>	<i>Z</i>
very large population	very large population
very high well-being	positive but very low well-being

If this suggestion were correct, then the main feature of *Z* which makes it better than *A* according to Total Utilitarianism (that is, the size of its population) would be lost in the comparison. And then it wouldn't be surprising if our intuitions gave us the wrong answer.⁷

Yet, in some cases with very large populations, it seems that we can take their relative sizes into account. For example, consider the comparison between outcome *A*, which has a very large population, and an outcome *B*, which has a much larger population whose members have a slightly lower well-being:



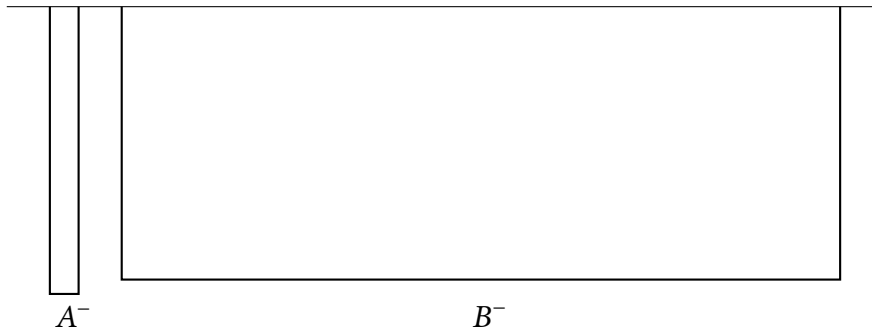
In this case, I guess that most people intuitively judge that the size of the *B* population makes up for its lower level of well-being. So they judge that *B* is better than *A*.

Some people may have a clearer intuition about the corresponding outcomes with negative well-being:⁸

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 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 this variant both populations consist of at least ten billion people.



Here, I guess that most people find B^- clearly worse than A^- . The fact that many fewer 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 their relative sizes into account.

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.⁹

⁹ Tännsjö 2002, p. 344. Tännsjö (2020, pp. 390–396) argues that the repugnance of the Repugnant Conclusion is debunked by the existence of instances where it is not repugnant that an outcome like Z is better than an outcome like A given a certain specification of what the lives in these outcomes are like. But this response does not work against the following strengthening of the Repugnant Conclusion (which also follows from Total Utilitarianism):

The Invariant 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 with lives that are barely worth living whose existence would be better, *regardless of what these lives are like in other respects*.

It is hard to identify with the very large number of people in Z .¹⁰ 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 . The people in A and the people in Z are all anonymous members of very large populations, so 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 intuition about the Repugnant Conclusion, namely, that we are bad at compounding lots of small values. Huemer 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

This 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. There are, however, 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

¹⁰ If, instead, the objection is that very large numbers mean very little to us and hence that they all mean roughly the same thing to us, then it's essentially the same as the imaginative-alikeness objection. And then it is unsuccessful for the same reasons.

¹¹ Huemer 2008, pp. 909–910.

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 wouldn't explain our anti-utilitarian intuition about this one-person analog of the Repugnant Conclusion.

Nevertheless, the compounding-small-values proposal might still be able to explain the unreliability of our intuition about the One-Person Repugnant Conclusion. Any *Z*-life that's better than some *A*-life, according to utilitarianism, must be extremely long—perhaps thousands of years long. Plausibly, we are unable to grasp directly the value of such lives, because we can't intuitively imagine what living for that long would be like. So we have to make the intuitive comparison between a long *A*-life and an extremely long *Z*-life, not by intuitively valuing the whole of the *Z*-life, but by valuing shorter, more intuitively graspable intervals. And then we fail to properly compound the small values of these shorter parts of the *Z*-life.

But consider a *Z*-life of the same length as the longest duration of life such that we can grasp the value of that duration of life without doing any compounding. And consider a much shorter *A*-life with a just slightly lower sum-total of temporal well-being than the *Z*-life. Suppose, for instance, 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's still counter-intuitive that the *Z*-life would be better than the *A*-life. But, in this case, we can evaluate the whole *Z*-life without compounding the values of any shorter segments. And then the compounding-small-values proposal doesn't work.

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 mislead intuition.

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 philosophers 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 all principles that say something about very large populations. Broome wishes to rely on some principles that do so. For example, in his defense of Total Utilitarianism, Broome 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 distributions (including those with very large populations), it says something about very large populations. Still, Broome wishes to rely on its intuitive plausibility.

To avoid this problem, Broome narrows his case against the reliability of our intuitions. His example of an intuition that depends on large numbers is not the Repugnant Conclusion. Instead, it's the similarly structured claim that it is better to save someone from a severe disease like AIDS than to cure any number of people from mild headaches. Broome writes:

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 $B(n)$

¹³ Broome 2004, p. 57.

¹⁴ Broome 2004, p. 120.

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

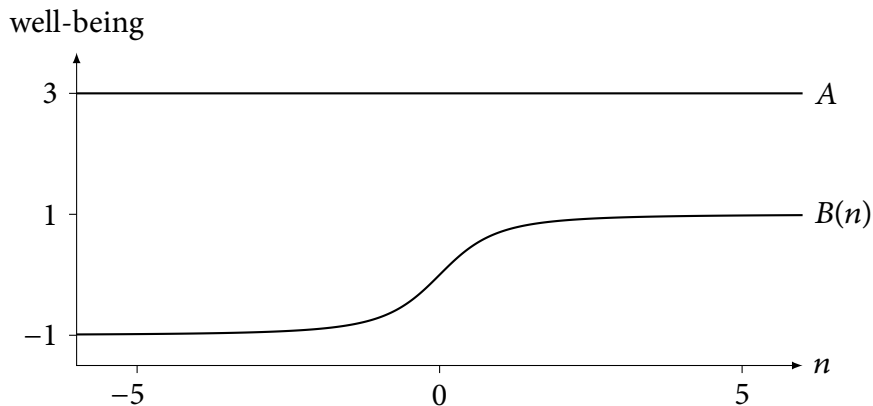
This narrowed proposal still rules out too much, however. There are reliable intuitions of the form Broome wishes to discredit. Let, for example, A and $B(n)$ be distributions for the same population, such that

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 people's well-being in A and their well-being in $B(n)$:



Plausibly, $B(n)$ gets better as n increases, because—other things being equal—it's better if everyone has a higher well-being. Similarly, it intuitively seems that, for all numbers n , A is better than $B(n)$, since everyone has a well-being of 3 in A whereas in $B(n)$ they have, for all numbers n , a well-being below 1. This intuition seems neither unreliable nor particularly dependent on our intuitive grasp of large numbers. So, if an intuition is reliable, it's not unreliable because it has the form Broome targets.

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 (as Broome claims), we should also have reliable intuitions about this logically weaker claim, which has the form Broome

¹⁵ Broome 2004, pp. 58–59.

exposes to doubt. Hence, by his own standards, Broome’s narrowed proposal also rules out too much.

One might object that this objection can be sidestepped if we just restrict the n parameter to population sizes in Broome’s narrowed proposal. This move, however, cannot explain why our intuition is unreliable in the One-Person Repugnant Conclusion, since in that case the population-size parameter is fixed. Besides, the move seems 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 intuition about the Repugnant Conclusion, I have a slightly less general aim. My aim is merely to argue that our intuition about the Repugnant Conclusion is unreliable as evidence against Total Utilitarianism.¹⁶ So my proposal need not rule out that this intuition could be reliable as evidence against some other theories that imply the Repugnant Conclusion. I shall argue that it’s likely that we would have the intuition that the Repugnant Conclusion is false even if Total Utilitarianism were true. Hence, if I’m right, this intuition is unreliable as evidence against Total Utilitarianism.¹⁷

My diagnosis is that the underlying problem isn’t merely the large-

¹⁶ In Gustafsson 2020, I argue that the total view is compatible with the Repugnant Conclusion, given that we accept an axiology where value bearers can differ not just in goodness (the value dimension where differences make value bearers better and worse) but also in undistinguishedness (the value dimension where differences make value bearers incomparable). That two-dimensional response to the Repugnant Conclusion is an alternative to the one suggested here. With the response presented here, we don’t need the undistinguishedness dimension.

¹⁷ 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

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

given that $P(R) \neq 0$ and $P(U) \neq 0$. For example, suppose that $P(R | U)$ —that is, your conditional probability of R given U —is 0.8. And suppose that $P(U)$ —your prior credence in U —is 0.8. Then your posterior credence in U given R should be

$$P(U | R) = \frac{3.2}{3.2 + P(R | \neg U)}$$

ness of the numbers in the Repugnant Conclusion. Rather, the underlying problem is making trade-offs where the relevant factors are extremely proportioned in opposite ways. My proposal's main assumption is that our intuitive understanding of the 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 105 m and its width to be 68 m. The second is a side of an extremely long and very narrow tape. We measure the length of the tape to be 3000 km and its width to be just 2 mm. We get that the area of the pitch is 7140 m² and that the area of the tape is 6000 m². So we get that the pitch has a much larger area than the tape. This conclusion, however, is only reliable if our measurement of the tape'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 can't rule out that the tape's width is 3 mm, which would make its area much larger than 7140 m² (it would be 9000 m²). And then we can't rule out that the tape has a larger area than the pitch. So, even though it seems like the area of the pitch is much larger than the area of the tape and that the margin of error for the width of the tape is tiny, this tiny margin of error still makes the comparison unreliable. This is because any error in the measurement of the tape's width is multiplied by its very great length, resulting in a huge margin of error for the calculated area.

In this manner, I propose that a 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 intuition about a comparative claim between x and y is 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

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 wouldn't 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 additionally that $P(R \mid \neg U) = 0.8$. 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 .

- (ii) there is a possible change in one of these two factors such that the change is small enough to fall within our intuition's 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 judgment 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's going on in the case of Total Utilitarianism and the Repugnant Conclusion.

A clarification of clause (i) may be needed. The point of this clause is merely to allow that our intuition can be reliably guided by dominance considerations even when the respects in which the options differ fall within our intuition's margin of error. When one option is superior in all relevant respects, we don't need to take the exact magnitude of this superiority into account in order to rely on dominance considerations. And, if so, our intuition's margin of error for this magnitude doesn't matter as long as we grasp the dominance.

One class of comparisons that seem to fulfil clauses (i)–(iii) given a theory T are comparisons between two options such that T evaluates the options by the product of two 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's a straightforward explanation of why our intuition is unreliable. The problem is that, when options are evaluated 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 may change which option has the greatest product of F and G . But, since our intuition is inexact, our intuition may plausibly be insensitive to this small change in factor F . Thus, due to the inexactness in our intuitive understanding of this factor, our intuition might deviate from T in these cases. This inexactness, however, does not rule out that our intuition reliably tracks T in cases without this kind of extreme trade-offs.

In the Repugnant Conclusion, there are, according to Total Utilitarianism, two factors opposed in this way. We must compare the high well-being and few people in A —that is, few relative to the size of the population in Z —with the very low positive well-being and huge number of people in Z . So 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 outcome Z with lives that are barely worth living such that Total Utilitarianism evaluates Z as better than A , there is an outcome Z' which is just like Z except that people in Z' have a slightly lower but still positive well-being and Total Utilitarianism evaluates Z' as worse than A .¹⁸ The people in Z and Z' have very similar well-being, and they all have lives that are barely worth living. And, when we intuitively compare Z and Z' to A , our intuition isn't 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 evaluation of A and Z), it seems that our intuition in this case is unreliable as evidence against Total Utilitarianism.¹⁹

(Note that I do not claim that we can't reliably take the difference in the level of well-being between Z and Z' into account when we compare them with each other, because that comparison is a simple case of dominance without trade-offs.)

With my proposal, 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 be guided by Total Utilitarianism. But there is a slight margin of error in our intuitive judgment of the parameters of this principle, which gets greatly amplified

¹⁸ I am assuming that there is no lowest possible positive level of well-being. I am assuming denseness for well-being levels rather than discreteness; see Arrhenius 2000, p. 163. If one gives up this assumption, one might get around my objection by imagining a Z outcome where each life has the lowest positive level of well-being. But this wouldn't help much unless we are clear about which well-being level is the lowest positive one, which seems implausible. If we don't know what the lowest positive level of well-being is like, we can't be sure that we're imagining lives with that level of well-being.

¹⁹ It may be objected that, if we allow the population in Z to be infinitely large, then it should suffice to judge that the people in Z have positive well-being in order to judge that Z with an infinite population is better than any outcome with a finite population. But, if we allow infinite populations, A could also have an infinitely large population and be infinitely good. And, if so, there wouldn't be any outcome like Z that is better than A ; so the Repugnant Conclusion wouldn't follow from Total Utilitarianism. To sidestep this issue, 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's likely to be unreliable. A second reply is that, if there is a margin of error in our intuition, then we couldn't reliably tell whether we imagine an infinite population with barely good lives or an infinite population with barely not good lives.

in cases where the principle evaluates an option by the product of 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 *A* and *Z* should be thoroughly equivocal, whereas there's widespread agreement that, intuitively, *Z* is not better than *A* (and very little agreement that, intuitively, *Z* is better than *A*). That is, my proposal fails to explain the observed systematicity in the alleged unreliability of our intuition about the Repugnant Conclusion.

There is, however, an erroneous assumptions behind this objection. The unreliability in our intuitions need not result in equivocal judgments. In cases where our intuitions are unreliable, our judgments may still have a bias in a certain direction. So far, I have only offered an account of some circumstances where our intuitions are unreliable, rather than an account of what intuitive judgments we make under those circumstances. I do think, however, that there is an explanation of the systematicity in our intuitive judgments in these cases.

Note that, in the cases we have discussed, our intuitive judgment always 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. Even though we have no exact grasp of the amount of well-being in an imagined life in *A*, we can still clearly grasp that it's at least above a fairly high amount. So, to a significant extent, we take the high quality of the lives in *A* into account. Yet, when we assess *Z*, there's 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

²⁰ One may be worried about the following kind of case: suppose we stipulate that each person in *Z* has a life that's slightly better than a life where 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's 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 our intuition about which outcome is better in this case. Note, however, that, though you might know what a life consisting of one

Note that this systematicity explanation still allows that we can make reliable judgments in some comparisons where the differences in the relevant factors fall within our margin of error. In cases where there are no trade-offs between the relevant factors, our intuition can still be reliably guided by dominance considerations. So we can reliably judge with the help of dominance considerations that, other things being equal, tiny improvements to people's lives make the world better and that the addition of barely good lives also make the world better. Likewise, we can reliably judge with the help of dominance considerations that, other things being equal, tiny detriments to people's lives make the world worse and that the addition of barely bad lives also make the world worse.

It may next be objected that my proposal seems a little ad hoc if it only explains the unreliability of our intuition about the Repugnant Conclusion.²¹ Fortunately, the proposal also explains many other cases where otherwise intuitive theories yield counter-intuitive implications. A further counter-intuitive implication of Total Utilitarianism is

lollipop lick would be like, you don't know what the exact well-being level (that is, the exact personal value) would be for that life. When you assess 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 testable to some extent. It predicts that there will be a lot of otherwise plausible theories *T* that yield counter-intuitive implications in cases where clauses (i)–(iii) hold. Moreover, the possibility that my proposal, even if correct, wouldn't compel many to alter their judgments about the Repugnant Conclusion seems irrelevant if our aim is not popularity but truth.

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 high and very low 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, since utilitarianism evaluates the badness of these options by the product of these factors, this is an instance of an extreme trade-off between morally relevant factors such that clauses (i)–(iii) hold. Hence my proposal also applies to our intuition about Hangnails for Torture.²³

My proposal also explains why our intuition about the One-Person Repugnant Conclusion is unreliable as evidence against utilitarianism. In this case, we compare a very long life with very high quality of life to a much longer life with a barely positive quality of life. And, since utilitarianism evaluates these lives by the product of their quality and their length, we again have an instance of the problematic form of extreme trade-offs between morally relevant factors. So my proposal also applies to our intuition about the One-Person Repugnant Conclusion.²⁴

Furthermore, my proposal does not rule out the reliability of our intuition about the Principle of Personal Good, which utilitarians typically rely on to motivate their view. If, given the same population, a first distri-

²² Pummer 2013, p. 37. The example is a variation of examples in Temkin 1996, p. 179 and Rachels 1998, p. 73.

²³ This explanation works equally well, changing what needs to be changed, for the similar lollipops-for-life case in Temkin 2012, p. 34.

²⁴ It is easy to come up with more examples. Lots of counter-examples to expected-utility theory have the following structure: we have a choice between having a significant amount of money for sure and a gamble which is almost certain to give us nothing but might with a very low probability result in a huge win; see, for example, the truncated (that is, finite) version of the St. Petersburg paradox in Hájek and Nover 2006, p. 706. A tiny difference in the probability of the huge win will have a large effect on the expected value of the gamble. So, if there is a margin of error in our intuitive grasp of this probability, our intuition that it's better to have the significant amount for sure than to have the gamble won't be reliable as evidence against expected-utility theory. The structural similarity between the Repugnant Conclusion and the St. Petersburg paradox is noted in Cowen 2004, pp. 83–84.

bution 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 distribution beats the first. So my proposal does not entail that our intuition about the Principle of Personal Good is unreliable as evidence in favour of Total Utilitarianism.

And, unlike the imaginative-alikeness proposal, my proposal doesn't 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 sizes and well-being levels, but these trade-offs aren't extreme; no small change in either population sizes or well-being levels 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 consider two general objections to this kind of defense of Total Utilitarianism. The first comes from Derek Parfit who responds to the worry that we might have trouble imagining the relevant populations 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.²⁵

Yet it's not merely the population in Z that may be arbitrarily large, the population in A may be so too. If the population in A is sufficiently large, it would have to contain very many people. But, in that case, 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 the Z population compared to the A population, which is the main advantage of Z according to Total Utilitarianism. Hence it's insufficient to just be able to imagine Z as an outcome with very many lives barely worth living. We have to be able to take into account how much larger the Z population is when the A population is huge.²⁶

²⁵ Parfit 1984, p. 389.

²⁶ 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 num-

Moreover, even if we only consider cases where the *A* population consists of just ten billion people, there's another problem with Parfit's reply. The sum total of well-being for a population of ten billion people with a very high quality of life is very large. Hence, even on Total Utilitarianism, there are outcomes that are worse than *A* but which still contain very many lives barely worth living. So, if we can only imagine *Z* as an outcome with very many lives barely worth living, we can't be sure that we imagine an outcome that is better than *A* according to Total Utilitarianism.

7. The Extrapolation Argument

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

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 favor 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 two 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 two years of excruciating torture.²⁷

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:

bers of people living such lives? This is what our task requires, and it's not at all clear that we are up to it.

²⁷ Pummer 2013, p. 41. Here and elsewhere, I have changed Pummer's numbering.

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.²⁸

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 favor 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, Pummer concludes, (1) is true and (2) is false.²⁹

While we might also question (4), I shall argue that (3) is false.³⁰ For small values of X (which are the ones Pummer extrapolates from), I don't think that, if (2) were true, we would become less confident in the Variable Claim as X increases. To see why, consider first the utilitarian principle that the value of an option is proportional to that 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 we would plausibly remain

²⁸ Pummer 2013, p. 41.

²⁹ 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.

³⁰ 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.

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 we wouldn't merely evaluate the torture as worse than the hangnails, we 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, we 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 unreliable as evidence against Total Utilitarianism. My explanation is that our intuitive understanding of the morally relevant factors is inexact and comes with a slight margin of error, which makes our intuitive judgments unreliable in cases with extreme trade-offs between these factors. Given Total Utilitarianism, the Repugnant Conclusion involves an extreme trade-off of this kind. Therefore, if Total Utilitarianism is true, our intuition about the Repugnant Conclusion is unreliable. So this intuition is unreliable as evidence against Total Utilitarianism.

I wish to thank Tim Campbell for especially helpful comments. I also wish to thank Gustav Alexandrie, Gustaf Arrhenius, Campbell Brown, Krister Bykvist, Roger Crisp, Marc Fleurbaey, Johan Jacobsson, Martin Peterson, Douglas W. Portmore, Theron Pummer, Wlodek Rabinowicz, 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.

References

- Arrhenius, Gustaf (2000) *Future Generations: A Challenge for Moral Theory*, Ph.D. thesis, Uppsala University.
- Broome, John (2004) *Weighing Lives*, Oxford: Oxford University Press.
- Cowen, Tyler (2004) 'Resolving the Repugnant Conclusion', in Jesper Ryberg and Torbjörn Tännsjö, eds., *The Repugnant Conclusion*, pp. 81–97, Dordrecht: Kluwer.
- 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.
- (2020) 'Population Axiology and the Possibility of a Fourth Category of Absolute Value', *Economics and Philosophy* 36 (1): 81–110.
- Hájek, Alan and Harris Nover (2006) 'Perplexing Expectations', *Mind* 115 (459): 704–720.
- 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*, ed. V. W. Robson, vol. III of *Collected Works*, Toronto: University of Toronto Press.
- (1972) *The Later Letters of John Stuart Mill 1849–1873: Part III*, eds. Francis E. Mineka and Dwight N. Lindley, vol. XVI of *Collected Works*, Toronto: University of Toronto Press.
- Mogensen, Andreas (forthcoming) 'Against Large Number Scepticism', in Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan, eds., *Essays in Honour of Derek Parfit, Volume 2: Population Ethics*, Oxford: Oxford University Press.
- Narveson, Jan (1973) 'Moral Problems of Population', *Monist* 57 (1): 62–86.
- Parfit, Derek (1982) 'Future Generations: Further Problems', *Philosophy & Public Affairs* 11 (2): 113–172.
- (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.
- 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.
- (2020) 'Why Derek Parfit Had Reasons to Accept the Repugnant Conclusion,' *Utilitas* 32 (4): 387–397.
- Temkin, Larry S. (1996) 'A Continuum Argument for Intransitivity,' *Philosophy & Public Affairs* 25 (3): 175–210.
- (2012) *Rethinking the Good*, New York: Oxford University Press.
- Whewell, William (1852) *Lectures on the History of Moral Philosophy*, Cambridge: Cambridge University Press.