

Resolving The Repugnant Conclusion

Tyler Cowen

Department of Economics

George Mason University

Fairfax, VA 22030

January 14, 2001

*The author wishes to thank Robin Hanson, Ron Heiner, and especially David Schmidtz and Erik Tallrot for useful comments.

Abstract

The Repugnant Conclusion is closer to infinity-based arguments, such as Pascal's Wager, than it at first appears. Both rely on an unbounded set of payoff comparisons. It is possible to restructure Pascal's Wager to resemble the Repugnant Conclusion more closely, as the use of infinity is not central to the former. I then consider settings in which the set of comparisons is bounded, so as to differentiate Parfit's problem from the more general issues involved with very large numbers. We then find the Repugnant Conclusion no longer necessarily arises as a matter of logic rather is an empirical contingency. I then present some plausible intuitions under which the Repugnant Conclusion never arises. The paradoxical nature of Parfit's Repugnant Conclusion is traced to the simultaneous application of two inconsistent outside observer constructs: one to judge the Repugnant Conclusion as repugnant, and another to define the utility scale for a marginally worthwhile life. Once the two constructs are made consistent, the Repugnant Conclusion can be defused.

I. Introduction

The Repugnant Conclusion has proven to be one of the most durable and powerful paradoxes in normative ethics. Through the comparison of different population and utility patterns, Derek Parfit has stressed the difficulty of producing a coherent consequentialist theory that avoids deeply counterintuitive conclusions.

To recap briefly, the Repugnant Conclusion postulates a society with a large amount of total utility. This large total, however, is obtained by having very many persons living at levels of utility just slightly above zero. The welfare of this society will dominate many highly attractive alternatives, simply by virtue of having a very large population. In essence, adding more marginally worthwhile lives can always produce an arbitrarily high utility total.

Parfit's (1984, p.387) statement of the Repugnant Conclusion reads as follows: "The Repugnant Conclusion. For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population, whose existence, if other things were equal, would be better, even though its members have lives that are barely worth living."

This comparison deliberately ignores feasibility considerations. It does not matter whether the highly populated society is possible. Rather, the question is whether we would prefer it, if it were possible. This hypothetical question is posed to determine how utilities should be compared to other values, and to ask what prevents total utility from dominating all other consequentialist values, when the sum of total utility becomes high enough.

Invoking non-utility values does not automatically sidestep the Repugnant Conclusion. In contrast, the Repugnant Conclusion illustrates the difficulty of preventing non-utility values from being overwhelmed in the moral calculus by total utility. It might be the case, for instance, that the less populated society has significantly greater amounts of

justice, aesthetic beauty, or dignity. If this is true, the Repugnant Conclusion alternative simply needs to make up for this deficiency by having more people to increase its utility total.

Philosophers and economists have tried to address the Repugnant Conclusion in several ways. Ng (1989a) argues that the Repugnant Conclusion is not repugnant and that we should welcome the more highly populated alternative, if indeed it has higher total utility. Temkin (1993, chapter 7) argues that the Repugnant Conclusion can be prevented by capping the number of "points" that can be awarded to total utility in a pluralistic social welfare function. Cowen (1996) argues that utility needs to be bounded to prevent the Repugnant Conclusion. Blackorby and Donaldson (1984) and Blackorby, Bossert, and Donaldson (1997) present "critical-level utilitarianism," which suggests that lives below a certain absolute level of utility do not count in the social welfare function, even if they are otherwise "worth living" from the individual point of view.

In this article, I add some new twists to the idea of boundedness. I consider versions of the Repugnant Conclusion based on bounded and unbounded population comparison sets. An unbounded population comparison set allows us to consider any possible population, no matter how large; a bounded population comparison set requires us to restrict our attention to some finite list, albeit a potentially large one. We will see that the unbounded versions of the Repugnant Conclusion do not differ significantly from such well-known problems as Pascal's Wager and the St. Petersburg Paradox.

The bounded versions allow for possible resolution. First, the Repugnant Conclusion does not always arise under the bounded versions of the problem. Second, we can find intuitions under which the Repugnant Conclusion never arises. More specifically, the Repugnant Conclusion postulates the existence of a very large number of people, all with lives "just slightly worth living." Once this notion of a "marginally worthwhile life" is spelt out in more detail, in a bounded setting, the Repugnant Conclusion can be avoided altogether. It cannot be demonstrated that these intuitions are the best moral theory, but they do show at least one way out of the Repugnant Conclusion. Boundedness of the

population list, and scrutiny of the zero point for utility, taken together, provide a possible candidate for Parfit's "Theory X."

I also show that this solution need not violate what Parfit calls the "mere addition principle." In the schema to be outlined, the addition of another happy life to the population can be considered desirable across all margins. Nor must it be postulated that the value of additional lives somehow "diminishes at the margin" or asymptotically approaches zero, or that moral judgments are intransitive. These assumptions provide "easy ways out" of the Repugnant Conclusion, and I do not regard them as final answers.

The proposed resolution of the Repugnant Conclusion is distinct from the extant literature. The idea of a bounded population comparison list is not currently explicit in the literature, and is distinct from the "capped utility" idea considered by Cowen (1996) and Temkin (1993), as will be explained in more detail below. Furthermore, the literature does not consider how boundedness of the population list and the definition of the zero point might interact to resolve the Repugnant Conclusion.

How to define the zero point, and thus a marginally worthwhile life, has been raised by a few other sources, but it has not received extensive analysis. Mackie (1985, p.246) questions the meaning of the zero point when he writes "A level that is really marginally better than non-existence must already constitute a high degree of flourishing, and beyond this little further improvement is possible," although he does not pursue the matter any further. Dasgupta (1988, p.117) portrays the Repugnant Conclusion as being full of the "wretched of the earth," and thus rejects it on those grounds, but he does not consider whether a non-wretched definition of a marginally worthwhile life might be available. In Blackorby and Donaldson (1984) and Blackorby, Bossert, and Donaldson (1997) the location of the zero point for the social welfare function differs from the location of the zero point for an individual. Their version of "critical-level utilitarianism" does not positively count lives below a certain absolute level of well-being.

The paper proceeds as follows. Section II argues that the Repugnant Conclusion is closer to some infinity-based arguments, such as Pascal's Wager and the St. Petersburg Paradox, than it might appear. Section III considers how a "bounded" version of the Repugnant Conclusion might operate, and what status it would hold. Section IV examines the intuitions under which the Repugnant Conclusion can be avoided altogether. I will argue that these intuitions are not implausible, although they cannot be shown to be the best available moral theory. The paper closes with some remarks on why the paradox of the Repugnant Conclusion arose in the first place.

II. Boundedness

Moral philosophers typically reject arguments based on infinities or infinite utility magnitudes. Pascal's Wager comes in several versions, but let us start with the version based on infinite payoffs. The Wager then asks why we should not believe in God, given that the expected payoff of belief is infinite. By assumption, reaching heaven brings an infinite payoff of unending, unconstrained bliss.

Now this argument has well-known problems. First, if we think the idea of an infinite payoff is meaningful, any set of beliefs involves some non-zero chance of getting the infinite payoff, and thus it remains hard to choose among beliefs. Each set of beliefs would bring an infinite payoff. We might decide that the maximum chance of reaching the infinite payoff dominates the lesser chances (though both have infinite expected value, perhaps they can be rank-ordered according to other decision criteria). But then we are left uncertain as to which beliefs maximize this probability.¹

¹ For a survey of the literature on Pascal's Wager, see the essays in Jordan (1994). McClennan (1994) and Sorensen (1994) take particular care to focus on the infinity issue. More generally, see also Rescher (1985) and Armour (1993).

These problems aside, most philosophers do not find this form of Pascal's Wager compelling. The infinite expected payoff appears to have "too much power" in our calculations.

To see the problem in starker form, consider the following two gambles:

.01 of infinite bliss

.99 ordinary life

vs.

.02 of infinite bliss

.98 of very painful eternal damnation, though not infinitely negative

Most individuals would prefer the former gamble over the latter, though both have infinite expected value. If we maximize the probability of getting an infinite payoff, as some versions of Pascal's Wager suggest, that would direct us to the latter. But again we tend to prefer the former because we believe that the infinite payoff receives "too much weight" in the expected value calculation.

Similarly, most individuals, acting as ideal outside observers, would decline a gamble that would end the world with probability 0.9999 but otherwise would create a world with infinite utility. We do not believe that the infinity should receive infinite weight in the final expected value calculation.

Nor do we pay an infinite or near-infinite amount to play the St. Petersburg game, even though properly specified payoffs can create an infinite expected return. The St. Petersburg game involves coin tosses. The greater the number of consecutive heads (tails), the greater the payoff. Obviously a very large number of consecutive heads in a row has only a small probability, but would bring a very large payoff. If the payoffs are specified properly, the game has infinite expected utility, as demonstrated by Samuelson

(1977). If we treat the marginal utility of wealth as constant, the game has the simple form of paying off one cent for one consecutive heads flip, two cents for two consecutive flips, four cents for three consecutive flips, eight cents for four consecutive flips, sixteen cents for five consecutive flips, and so on. To the extent the marginal utility of wealth declines with higher wealth, we simply need to raise the dollar payoffs to maintain the infinite expected utility of the game.

Arguably no one has solved the problem of exactly why expected utility calculations go wrong when they encounter infinities.² For the purposes of this paper I take it as axiomatic that the infinity-based calculation does not overwhelm our other moral intuitions. That is, I side against the infinities.

In siding against the infinities, I do not mean to dismiss their importance. Often we make scientific progress by pushing on the hard cases and demanding answers. When all is said and done, when we have examined the problem more carefully, we might end up siding with the infinities, so to speak. But if this is what the Repugnant Conclusion proves to come down to, we have made progress on that problem. We would have shown that the Repugnant Conclusion was the problem of infinities in another guise. We would have turned the problem into a more general one, namely the problem of our intuitions against the infinities.

Now unlike the problems presented above, the Repugnant Conclusion does not rely on infinity. It does not postulate an infinite number of people, all living at a very small level of utility, and then compare that infinity to other possible population sizes. Nonetheless by considering an unbounded list of possible populations, the Repugnant Conclusion "deals in infinity" through other means. The Repugnant Conclusion finds out, ex post, how large a number is needed to "act like infinity" in the comparison, and then chooses that number to be the population size. The very large number acts "as if" it were an

² To consider one example, Sorensen (1994) attempts to translate the problem of infinities into the Sorites problem.

infinity, given that we can choose the size of that number ex post and make it sufficiently large to establish the Repugnant Conclusion.

Let us reexamine the infinity-based problems above, and see how they can be redone without infinity. Most of the moral arguments that appear to rely on infinity, such as Pascal's Wager, in fact require unboundedness only. The formal mathematical definition of "bounded" is as follows: a set is bounded if and only if there exists some finite distance 'e', such that every point in the set S is no farther from the origin than 'e' (see Jehle 1991, p.21). Unboundedness thus says we can produce a very large number ex post, by choosing from the list of all numbers, but without requiring recourse to infinity. The alternative assumption of boundedness would require us to specify the list of possible numbers in advance, implying we cannot choose any number we want, be it for the utility of heaven (Pascal's Wager) or the number of people postulated in a given world (the Repugnant Conclusion). Boundedness, however, remains compatible with the selection of very, very large real numbers in the relevant context.

Now let us redo Pascal's wager, to show that the paradox relies on unboundedness, rather than on infinity per se. Consider Pascal's Wager, and assume that the life of an unbeliever reaps 'k' utils. Assume further that the probability that God exists is given by 'p,' which need not be large. There now exists some utility of heaven, "UH," very large but less than infinity, so that it is "rational" to believe in God (at least given Pascal's logic) and give up the life of the unbeliever. No matter how large the k and how small the p, we can find such a real number UH to make belief rational, but without resorting to infinity.³

We can perform a similar transformation for the other infinity-based arguments mentioned above, including the St. Petersburg Paradox and the .01 vs. .02 probability of

³ McClennan (1994, p.115) notes that Pascal's own presentation of the wager did not typically require infinite utility.

infinite bliss. Each problem can be reformulated using unboundedness, based on knowledge of how large a real number is needed to "act like infinity" in each case.

In this regard an argument that relies on unboundedness, such as the Repugnant Conclusion, is closer to the logical structure of infinity arguments than it may at first appear. No matter how low the utility level is set for the muzak and potatoes life, Parfit wishes to reach into the unbounded list and pull out a high enough number for total population so that the Repugnant Conclusion arises. Parfit requires a sequential, two-step process where first the utility description of the muzak and potatoes life is determined, and then total population is chosen from an unbounded list. The Repugnant Conclusion picks a large enough number for population so as to "act like infinity."

I therefore conclude the following: if we are suspicious of infinity-based arguments, we should be suspicious of unboundedness as well. Unboundedness of the comparison set allows finite numbers, albeit very large ones, to play the role typically attributed to infinities. So if we reject arguments from infinities, we should be willing to consider boundedness for the set of population comparisons, when it comes to the Repugnant Conclusion. Boundedness, in this context, simply means that we must specify a maximum greatest possible population. This number can still be very, very large, and this does not constitute an arbitrary attempt to rule out the Repugnant Conclusion by fiat that all large populations are impossible.

Note that boundedness for population comparisons differs from the "capped utility" view as discussed by Temkin (1993) and Cowen (1996). Capped utility doctrines are set in a context of pluralistic values, where a society can "win" only so many "points" for the contribution of utility towards total value. If the contribution of utility to the overall worth of a society is capped in this fashion, it is easy to avoid the Repugnant Conclusion.

Adding more people will increase the utility total, but utility can add only so much value to overall worth.⁴

Capped utility is not convincing, however, because it denies what Parfit calls the mere addition principle. At some point the addition of a normal, happy life does not make a society better in the overall rankings. Either the additional life must be worth nothing, or the value of that life must asymptotically approach nothing. These conclusions seem implausible. Capped utility also implies the implausible conclusion that all our lives would be worth much less, possibly no more than epsilon, if we discovered the existence of many additional human beings on other planets. If those other planets were sufficiently populous, the value of the human beings on Earth would have to asymptotically approach zero.

Bounding population comparisons does not violate the mere addition principle or require a diminishing marginal value to extra lives. For any set of population comparisons we consider, the mere addition principle can hold within that list. Unlike capped utility, population boundedness places no particular maximum on how much total utility can contribute to the goodness of a situation. It implies only some limit, albeit a potentially very very high one, to how large a population we are willing to consider in our moral comparisons.

In addition to the Pascal's Wager analogy, two other reasons provide support for the assumption of boundedness. First, an unbounded list does not allow us to identify any

⁴ Cowen (1996) used the term "boundedness," but to refer to a cap on total utility, rather than a limit on the population options we are willing to consider. The earlier paper considered what I am now calling capped utility. I view this second paper as an extension of the first one. The earlier paper suggested that solving the Repugnant Conclusion required some new way to think about boundedness, without specifying what that might be. I attempt to provide such a reconceptualization of boundedness in this paper.

well-defined maximum for the social welfare function, be it in favor of the Repugnant Conclusion or against. Any postulated alternative can always be beat by another alternative, and there is no "best" outcome, any more than there is a largest real number. With or without the Repugnant Conclusion, we cannot generate a consistent set of "outside observer" rankings with an unbounded comparison list. The Repugnant Conclusion would be simply another concrete illustration of this point

Empirical plausibility is the second reason for assuming boundedness of the comparison list. The number of actual population/utility alternatives available in the world is finite, even if it is very, very large. More metaphysically, the universe appears to contain only a finite amount of energy and matter, which will again bound the list of possible population comparisons.

Invoking empirics is a tricky defense of boundedness, and needs to be used carefully. It has already been argued (see the Introduction) that the Repugnant Conclusion is not about feasibility constraints (i.e., whether such a large population is possible), but rather about how to compare utilities in principle. So the very large population postulated in the Repugnant Conclusion need not be empirically possible. Nonetheless the comparisons postulated by the Repugnant Conclusion should not deny fundamental principles of reality, such as the finite nature of the possibilities for human society. In this sense, imposing boundedness on the comparison list has stronger force than trying to rule out the Repugnant Conclusion on the grounds of its empirical unlikelihood. In this context, boundedness is not saying that any particular population is impossible, boundedness implies only that not all populations are possible or necessary to consider.

III. The Repugnant Conclusion with Boundedness?

The above arguments do not provide a knockdown case for assuming boundedness more generally, or for the more particular case of the population comparison list. But at the very least they suggest that boundedness for the population comparison list is a reasonable approach. That is all we will need to find a way out of the Repugnant

Conclusion. The Repugnant Conclusion takes its force as an impossibility result; i.e., that there is no plausible way to avoid it without violating other compelling moral axioms. It would constitute progress to find any way out of the impossibility result at all, even if it could not be demonstrated that the specified escape was necessarily the correct or best available moral theory.

Without having proved the necessity of boundedness, let us take boundedness as an experimental variable, to see how and if moral theory would change. In the context of the Repugnant Conclusion, boundedness implies a list of possible population comparisons with their utility profiles. The list would compare “fifty billion people with average utility of 800 vs. forty trillion people with average utility of 0.3,” “fifty-one billion people with average utility of 900 vs. forty-one trillion people with average utility of 0.4,” and so on, giving a very long (but bounded) list of possibilities. The assumption of boundedness forces us to restrict our attention to some part of the list with a well-defined maximum, rather than the entire list without end.

We see that in this set-up, the Repugnant Conclusion might arise, but need not arise. The bounded population comparison list may or may not contain very highly populated alternatives that welfare-dominate less populous alternatives with higher utility, in a manner that we find repugnant.

Return to Parfit’s original comparison, involving a society of ten billion individuals, each with very high utility. Under a bounded population list, there is no guarantee of finding a large enough population of epsilon-utility individuals to welfare-dominate. Parfit’s original statement of the Repugnant Conclusion stated that “there must be” some welfare-dominating but repugnant alternative. If we consider boundedness, we can whittle this claim down to “there might be” some welfare-dominating but repugnant alternative. Under boundedness, the Repugnant Conclusion becomes a maybe, an empirical contingency, rather than an iron clad impossibility theorem or an unavoidable counterexample.

What intuitions rule out the Repugnant Conclusion altogether?

Given a bounded population comparison list, the next question is whether there exist intuitions or axioms, under which the Repugnant Conclusion never arises. The answer here is yes. Again, I will not provide a knockdown moral argument for these intuitions, but instead will show there are at least plausible. I will present these intuitions in terms of three major claims, to be examined more closely below:

Intuition 1: Reliance of cardinal utility measurements on moral intuition

Intuition 2: Modal interdependence of value, or the value of one unit depends on how we would evaluate its potential multiplication

Intuition 3: Consistency across moral intuitions

Before proceeding with this endeavor, however, we need to step back and consider the definition of the Repugnant Conclusion more closely, specifically the embedded definition of a marginally worthwhile life.

Defining a marginally worthwhile life

Parfit does not offer a formal definition of a marginally worthwhile life, but at one point (1986) refers to a life of "muzak and potatoes." We could, for instance, imagine an individual with a short life span who is fed a small amount of potatoes and played a small amount of muzak, and then dies immediately. Such an individual's life, however short and incomplete, might be just slightly better than no life at all. This is a description of a marginally worthwhile life, however, rather than a formal definition.⁵

⁵ The muzak and potatoes example does not appear in Reasons and Persons, but rather is taken from a 1986 essay. Parfit may have intended this example as a clarification of the concept, which is not defined in the 1984 book.

We can formally define a marginally worthwhile life in at least two ways. The first definition is a purely physical one and relies on an objectively measurable notion of cardinal utility. We must have, in principle, some means of measuring the amount of utility (or other relevant values) in a life. A life that contained positive utility would be above the zero point and a life that contained pain or negative utility would be below the zero point. A marginally worthwhile life would be just above the zero point, by a very small amount. I will refer to objectively measurable cardinalism to describe this view.⁶

A second definition of a marginally worthwhile life relies on intuitive moral judgments. Under this standard we look at a life and judge that life to be close to the zero point, or far away from the zero point, using some moral theory of what makes a life go well. I will refer to outside observer theories, without necessarily endorsing all aspects of outside observer theories in the narrow sense, as the term is often used. Note that the outside observer theory is likely to be cardinal as well, but its aggregation procedures are based on intuitive judgments rather than objective measurement of some real physical quantity. The outside observer construct is not the only alternative to objectively measurable cardinalism, but I will use it as a proxy for a variety of other approaches to defining a marginally worthwhile life in terms of philosophic intuition.⁷

⁶ Note that a pure ordinal standard cannot be used to define a marginally worthwhile life. Ordinalism can place a life above or below the zero point (would an individual prefer that life to no life at all?), or rank lives in terms of their closeness, but ordinalism offers no metric of distance to tell us “how close.” In ordinal terms a life is never “near” the zero point.

⁷ Cowen (1989) surveys some ideal observer theories, most notably that of C.I. Lewis, and discusses their relevance to population theory. Note that the Parfitian framework cannot reject such constructs altogether, since it relies upon them for the judgment that the Repugnant Conclusion is indeed repugnant.

Note that Parfit's "muzak and potatoes" example could fit either the objectively measurable or the outside observer conception of a marginally worthwhile life. An objective measurement, if possible, could presumably yield an amount of utility very close to zero. Alternatively, an outside observer might conclude that such a life is just barely worth living, based on some notion of what makes a life go well.

As a general approach to valuing a life, I view the outside observer construct as more plausible than objectively measurable cardinalism. Measurable cardinalism requires that various pleasures and pains can be made commensurable in terms of a single metric, just as a thermometer might record temperature. In many cases, however, an account of commensurability requires judgments that are not strictly the result of objective measurement.⁸

Pleasures and pains take many forms. Consider the pleasure of passionate sex, the joy of parenting a young child, the experience of hearing a first-rate concert, and the simple enjoyment of a friendly encounter with a stranger. To the list we could add the pain of physical torture, the pain when a loved one dies, and the frustration of having to wait at an excessively long traffic light. It is not obvious that all of these pleasures and pains can be reduced to a single, commensurable unit, measurable in physical terms. Even if they are so reducible ultimately, they are not so reducible under current scientific knowledge, or anything close to current knowledge. Instead, our notion of well-being, and the value of a life, relies on a judgmental component about the relative importance, intensity, and relational aspects of different pleasures and pains. Plato, in his dialogue Philebus, stresses that different pleasures can be reconciled, or made commensurable, only under the guidance of philosophic wisdom.⁹

⁸ Ng (1975) is one example of a modern author who still talks in terms of an unambiguous physiological measure of cardinal utility.

⁹ On Plato's Philebus, see Benardete (1993).

These same intuitive judgments serve as input into what a marginally worthwhile life consists of. More than one kind of life will stand just above the zero point. In contrast to the muzak and potatoes life, consider another life, ninety years long and rich with experience, joy, and love. That life is clearly well above the zero point. We can, however, now modify this life and add increasing degrees of pain, or perhaps experiences of torture. Given that pain and pleasure have no obvious homogeneous metric, we must rely on some kind of intuitive or philosophic judgment to decide how much pain would be needed to make the life just marginally worthwhile.¹⁰

Similarly, we can extend the comparison along dimensions of probability. We could balance the rewarding forty-year life against a 98 percent chance of great enduring pains at the end of that life, a rewarding forty-one-year life against a 98.5 percent chance of great pains, and so on. The richer the life, the greater the chance of extreme pains needed to bring that life close to the zero point. But again we rely on intuitions and philosophic judgments to judge whether the life is worth living and by how much it is worth living.

The outside observer view also accounts for the differing repugnance of cardinal utility profiles that are similar in the aggregate. Rather than the muzak and potatoes lives, consider a comparable number of individuals, where each life is full of both immense joys and great pains. Assume further that in objectively measurable terms the joys and pains are balanced such that each of these lives is just above the zero point. To many individuals this scenario seems less repugnant than the muzak and potatoes lives, even though both scenarios contain lives that are just above the zero point. The second society, whatever its great pains and tortures, contains many peaks of very high value. That society is more than just muzak and potatoes. Our different intuitions about these two cases suggest that the worth of a life is determined by more than objectively

¹⁰ It might be argued that forty years of wonderful life could not be outweighed by any amount of pain and suffering. If this is the case, make the initial life sufficiently short, say ten years long, so that some amount of pains can outweigh the pleasures.

measurable utility. In our intuitions, not all locations “just above the zero point for objectively measurable utility” are created equally.¹¹

Finally, cardinal measurements, even when they are available, do not coincide with our intuitive judgments about life choices. Several well-known moral conundrums illustrate the tension between objectively measurable utility and moral intuitions. Assume, for instance, that a man's wife fools around on him but does not inform him of her actions and he never knows or senses that his marriage is any worse. An objectively measurable standard will suggest that the man is not worse off, but this is not obviously the right answer. Or assume that I suffer under great stress from my mother's terminal illness. It is not clear that I should take a pill that would make me indifferent to her pains. We want to be the kind of individual that suffers under our mother's suffering, even at the expense of objectively measurable utility. Robert Nozick's “experience machine” (1974, pp.42-45) can simulate any experience we want, but the happenings are not real. Once we are hooked up, however, we no longer know that it is a machine. Nozick points out that not all individuals would choose to hook themselves up to the experience machine, even though doing so would maximize objectively measurable utility. In each case, we allow our intuitive judgments to modify or override the recommendation of the objectively measurable standard for cardinal utility.¹²

Modal interdependence of value

¹¹ On the possibility of constructing “multiple” versions of the Repugnant Conclusion, using such alternative utility profiles, see Ryberg (1996).

¹² Griffin (1986) surveys some of these conundrums. Most generally, the objectively measurable standard treats people as “receptacles of utility,” whereas the outside observer approach asks what kinds of utilities people would like to have. Even if we side with the objectively measurable answer in a given case, the fact remains that we must consider additional philosophic intuitions to make the appropriate judgment.

I now introduce an intuition that I call **modal interdependence of value**. Under this intuition, an outside observer cannot judge “what a single life is worth” without thinking about “what fifty billion comparable lives would be worth.” The value of the single life does not depend on whether the other fifty billion lives actually exist; rather, the very thought of what fifty billion lives would be like helps us determine the value of one life. In other words, the aggregative intuitions of the outside observer reflect a non-separability of value.

Some forms of holism imply modal interdependence. Under holism, the value of a single unit depends on how that unit stands in relation to other parts. These other parts can be hypothetical or modal, rather than real.

Consider, for instance, a painting and its frame. A good frame can greatly enhance the value of a painting. When we look at an unframed painting and ask how great it is, we are asking, in part, how good it would look with a proper frame. Similarly, we will be considering how good the painting would look if hung in the proper light, or on a properly painted wall. We value the painting by asking these questions, even if the painting is currently unframed or badly hung. It would be plausible to claim that Las Meninas is Velazquez's greatest painting, even if the current frame were disastrous. Holism, and the interdependence of parts, can thus refer to hypotheticals to some extent.

Similarly, when judging the moral stature of an individual, we can ask how that individual would respond to opportunities for personal charity. Our answer to this modal query will influence our judgment of how benevolent the person is, even if the person will never confront those charitable opportunities. Once again, modality influences our evaluation of a single unit or a smaller piece of a larger picture. Judgments based on properties of objects will typically refer to modalities in this manner.

The intuition of modal interdependence of value considers one very particular kind of modal fact, namely modal multiplication. To what extent does the value of a single unit depend on how we would regard its existence in greater numbers?

To refer to another context, we often place more value on artworks that, in principle, would retain their vitality if multiplied. We value some artistic images more highly if they can withstand widespread dissemination; this intuition is independent of what actually happens with the artwork. We sometimes downgrade a painting or musical composition if we decide: “I would think less of it if I heard or saw it all the time.” This thought induces us to conclude that the single manifestation is worth less than we otherwise would have thought.

Similar intuitions about interdependency influence our valuations of children (regular children, not of Parfit's muzak and potatoes variety). We believe that if we sired and raised many children, each child would continue to hold great value for us. Each child would remain special. Such a thought makes us realize just how valuable children are, and it causes us to upgrade how we value children, even if we only have one child or perhaps have no children at all.

We must be very careful to specify the modal nature of this intuition. If we have more children, we might in fact treat each child less well, due to constraints of time, affection, and money. Perhaps we cannot afford to send every child to Princeton. The intuition does not deny this. The intuition rather indicates that whatever high value we assign to children, it follows (partially) from contemplating what it would be like to have many children. Children are worth a great deal, as reflected by the fact that they remain special, even when multiplied. Again, we should ask questions about “the many,” before issuing a final judgment on how to value “the one.”

Presumably we do not hold the same intuitions when we consider the multiplication of mediocre hot dogs. If we think of a world where mediocre hot dogs are very plentiful, we do not then conclude that mediocre hot dogs are something special. If anything, the thought of such a world induces us to downgrade the value of a single mediocre hot dog. We realize more fully that mediocre hot dogs, unlike children, do not have some kind of

persisting and memorable uniqueness. Rather they are interchangeable parts of very little individual value.

The value of uniqueness is one way to approach the modal interdependence intuition. Typically we value the unique more highly in a variety of contexts, as illustrated by the examples above of art and children. By contemplating modal multiplication, we receive a better sense of whether an item has a unique value that matters. In the case of the mediocre hot dogs, we fail to see such a unique value for each hot dog when they are plentiful.

Now consider how modal multiplication influences our valuation of the "muzak and potatoes" lives of the Repugnant Conclusion. I wish to suggest that these lives might be more like the mediocre hot dogs than they are like children of the ordinary variety. When we consider what many such lives look like, in tandem, we decide that each life is worth less than we might otherwise have thought. We do not realize that each life remains truly unique and of great individual value. Contemplating modal multiplication induces us to downgrade the value of a single muzak and potatoes life. I have no knockdown argument for this intuition, but it is consistent with the very repugnance of the Repugnant Conclusion, which is based on the non-impressive nature of the multiplied muzak and potatoes life.

We can now see one possible way out of the Repugnant Conclusion. Let us add the consistency of moral intuitions to the mix. Intuitional consistency forces modal interdependence to interact with our judgment that the Repugnant Conclusion is truly repugnant. Given our judgment of repugnance, we must downgrade how much we value each individual life to the point where the Repugnant Conclusion no longer welfare-dominates.

More systematically, the Repugnant Conclusion is avoided by the following means. To evaluate a (bounded) list of possible population comparisons, we ask how much a single life, and by extension many lives, is worth. We have a non-separability of value; we

cannot judge “what a single life is worth” without thinking about “what fifty billion lives would be worth.” In other words, that we find the Repugnant Conclusion repugnant helps determine the value of a single life, especially of the muzak and potatoes variety. The greater the repugnance of the Repugnant Conclusion, the less the muzak and potatoes life should be worth. The value of the single life does not depend on whether the other fifty billion lives actually exist; rather, the very thought of what fifty billion muzak and potatoes lives would be like helps us determine the value of one life (modal interdependence of value).¹³

We then find that there exist sufficiently low valuations for the muzak and potatoes life to avoid the Repugnant Conclusion. Our intuitions can assign very low utility levels to the marginally worthwhile lives, so low that the highly populated outcome does not welfare-dominate the society with higher average utility.

The central intuition for defusing the Repugnant Conclusion is thus simple. The muzak and potatoes lives simply are not worth very much. The trick, so to speak, is to see that under boundedness this can suffice to stop the problem from arising.

To see the resolution in starkest form, consider what I call the Non-Repugnant Conclusion:

The Non-Repugnant Conclusion

“For any possible population of at least fifty billion people, all with a very high quality of life, the utility total will be very high. In comparison, consider some much larger

¹³ We can now see more clearly how the arguments of this paper relate to Blackorby and Donaldson. Blackorby and Donaldson implicitly accept an objective notion of cardinal utility and thus introduce the “minimum utility level,” below which lives do not count for the social welfare function. I am suggesting that modal interdependence reshapes the entire utility scale.

imaginable population, full of lives just slightly above the zero point. For any maximum size of population that is chosen, there exists a numerical value for each of those lives, sufficiently close to zero, so that the fifty billion people have a higher utility total.”

Given the assumptions enumerated above, there always exists some intuitional evaluation of the muzak and potatoes life so that the Repugnant Conclusion does not arise. While it cannot be proven that this intuitional judgment is the “correct” one, that intuition is consistent with our judgment that the Repugnant Conclusion is indeed repugnant. Nor does that intuitional judgment violate the mere addition principle or fall into any of the other “traps” that bedevil potential solutions to the Repugnant Conclusion.

The proof behind the result is simple. Take the largest possible population on the comparison list and give each individual a "muzak and potatoes" existence. Then compare this to another outcome with a high average utility and a relatively high population. Define p_1 as the population of the more populated outcome and p_2 as the population for the less populated outcome. Also define (M&P) as the utility level for a muzak and potatoes life, and AU as the average utility in the “normal” scenario with wonderful lives. The more populated outcome will have a higher total utility only if $(p_1 \times \text{M\&P}) > (p_2 \times \text{AU})$. In this setting, p_1 , p_2 , and AU are all fixed constants, once the comparison of population profiles is specified. We are really comparing whether there exists a magnitude for M&P that is smaller than $(p_2 \times \text{AU})/p_1$. The answer will always be yes, since there is no smallest finite real number greater than zero. In other words, there exists a positive but non-zero level for M&P such that a Repugnant Conclusion never arises in a population comparison list of bounded size.

Most simply, for any finite real number ‘n,’ there exists another finite real number ‘k’ (the utility description of the muzak and potatoes life) such that $n \times k$ is arbitrarily small. Given that $n \times k$ can be arbitrarily small, the welfare-dominance of the Repugnant Conclusion is stopped.

The Non-Repugnant Conclusion has much of the logical structure of the Repugnant Conclusion. Each comparison starts with a fixed size and a fixed utility level for the

population of fifty billion. Parfit's Repugnant Conclusion proceeds by fixing the utility difference between the marginal lives and the zero point, and then increases the number of lives to produce a welfare-dominating, highly populated alternative. In contrast, the Non-Repugnant Conclusion first fixes a highly populated alternative (or alternatives), and then finds that there exists a sufficiently small utility level for those marginal lives, such that the smaller population with higher average utility is welfare-dominating and the Repugnant Conclusion is not preferred.

The question, of course, remains why we consider the comparison of the Non-Repugnant Conclusion, rather than the comparison of the Repugnant Conclusion. It is here that modal interdependence and boundedness of the population comparison set reenter. Boundedness of the population comparison set implies that we cannot choose any population we want, ex post, but rather that a maximum possible population must be specified ex ante. Modal interdependence implies that the value of a single life is not determined until we have examined the additive consequences of any valuation decision. Thus we have a maximum population specified first and a value of life determined by the application of moral intuitions to the set of feasible comparisons. In other words, we start by placing some limit on the problem and then allow for simultaneous determination of various values through the application of moral intuitions. Parfit's approach is more sequential and assumes that the value of a life is defined in physical or objective terms, independently of a broader web of philosophic intuitions. Parfit first reports this value of life and then subsequently chooses a population size from an unbounded list.¹⁴

¹⁴ It could be argued that once a specified utility level is assigned to the muzak and potatoes life, we can then choose a sufficiently large population so that the larger population of muzak and potato lives has higher total utility than the fifty billion wonderful lives. This criticism, however, ignores the bounded nature of the population comparison list. The bounded part of the list we are examining is defined to include any subsequent population/utility comparisons we might consider. Any "extra" comparisons we might add later simply mean that the original specification of the list was not large enough and thus was a misspecification.

We now can see more clearly how modal interdependence -- one form of non-separability -- differs from the non-separability of value postulated by the capped utility approach. Capped utility postulates that the value of a single life depends on how many other lives exist; this leads to well known problems with the mere addition principle. We can call this “number-contingent non-separability.” Modal interdependence postulates that the value of a single life depends on how we regard the value of many lives, whether or not those other lives exist. This philosophic move leaves aside number-contingency and its familiar problems. Instead it recasts non-separability in terms of how our intuitions about the value of a part relate to our intuitions about valuations of the whole, regardless of how many people are out there in the world. Most responses to the Repugnant Conclusion recognize that some form of non-separability is necessary to defuse the paradox, but to date they have focused excessively on number-contingent forms of non-separability, and not enough on modal interdependence of value.

Is moral theory transitive?

The resulting moral theory does not violate the traditional assumption that moral relations are transitive (Temkin XXXX has suggested dropping the transitivity assumption as one way out of the Repugnant Conclusion). Some expositions of the Repugnant Conclusion present more than a comparison between the two world-states compared above, and in the Parfit quotation reproduced in section I of this paper. Instead, these expositions of the Repugnant Conclusion start with the less populated alternative with higher average utility, and successively transform that alternative, through small changes, into the Repugnant Conclusion.

To see this process in more detail, consider the comparison from Parfit (1986). We start with many individuals at relatively high levels of utility, which Parfit calls "A" and looks like the following:

Figure 1 goes here.

We then compare this outcome to an alternative where the previously existing people have the exact same utility level, but some new people are added, with lower utility levels. Parfit calls this outcome A+, as illustrated in Figure 2. Clearly A+ appears to be "not worse than" A and arguably better than A.

Figure 2 goes here.

Parfit then compares A+ with B, which has a slightly higher utility total than A+ and a much more equal distribution of utility, illustrated as follows:

Figure 3 goes here.

B appears to be not worse than A+, and perhaps even better than A. Through successive application of this reasoning, however, we eventually end up with an alternative with a very high utility total but a very low average. In other words, we end up with the Repugnant Conclusion, by consistent application of step-by-step comparisons of "not worse than." It is for this reason that Temkin has challenged transitivity to avoid the Repugnant Conclusion.

The approach of this paper takes issue with this presentation of the Repugnant Conclusion on several points, drawing upon the arguments above. First, the argument requires an unbounded population comparison list, as discussed in section II above. In that regard this version of the Repugnant Conclusion resembles the more general problems with infinity and unboundedness, as has already been discussed. Second, the height and meaning of the utility boxes is not fixed in advance, but rather depends on philosophic intuitions and reflective equilibrium, taking into account all their implications, including the Repugnant Conclusion. More specifically, how much we value one life will depend on modal interdependence.

We can now see how the Repugnant Conclusion may appear to violate transitivity. As we examine the step-by-step comparisons presented directly above, Parfit considers each small change in population without introducing modal interdependence along the way. Once all the small changes are complete, we have a very large change, and modal interdependence tells us that it is an unpleasant one; i.e., that the Repugnant Conclusion is repugnant. Moral relations will appear intransitive. But had we introduced modal interdependence along the way, the repugnance of the final result would feed back into our initial intuitions and stop the process from reaching unpalatable conclusions. In essence, modal interdependence tells us that we cannot rely on step-by-step comparisons, as does this presentation of the Repugnant Conclusion.

What are the roots of the Repugnant Conclusion?

This resolution also provides a means of thinking about why the Repugnant Conclusion arose in the first place (many other "solutions" to the Repugnant Conclusion do not do this, but rather appear ad hoc). Parfit's paradox arises because his exposition causes the two differing notions of outside observer to conflict. The model of outside observer used to define the zero point should have consistent intuitions with the model of outside observer used to judge the repugnance of the Repugnant Conclusion. But as Parfit has constructed the comparison, the two models of the outside observer do not necessarily agree (admittedly neither is fully defined in Parfit's writings).

The paradox can be defused by forcing the two notions of outside observer to be consistent, within a framework of bounded population comparisons. If the Repugnant Conclusion truly is repugnant, the muzak and potatoes lives should not have a utility description so far above zero as to make the highly populated outcome welfare-dominate.¹⁵

¹⁵ Mackie (1985, p.247) hints at this interpretation when he refers to: "the two terms 'welfare,' which tends to have connotations that suggest the quantitative view, and 'quality of life,' which suggests the non-quantitative view." Note also that an unbounded

An alternative (but less persuasive) means of making all intuitions consistent is to force our judgments about the repugnance of the Repugnant Conclusion into line with an objectivist notion of distances from the zero point. If cardinal utility truly is an objective, measurable, and directly commensurable entity, then the Repugnant Conclusion no longer is repugnant, as Ng (1989b) has suggested. Blobs of utility are simply blobs of utility, regardless of what shape and size of package they come in. Whether we get those blobs in the form of many lives of muzak and potatoes, or in the form of fewer but more noble lives, would not matter once the relevant premises are accepted. This resolution also imposes consistency on the problem, although for reasons discussed above, I do not find the objectively measurable vision of cardinal utility to be a compelling one.

IV. Concluding remarks

To sum up, I have presented some new ways of thinking about the Repugnant Conclusion. We first saw that the logical structure of the Repugnant Conclusion was closer to Pascal's Wager, and other infinity-related arguments, than was immediately apparent. We then pursued the "experiment" of considering a bounded population set, so as to differentiate the Repugnant Conclusion from these other, more general problems involving very large numbers. In this context we saw that the Repugnant Conclusion does not necessarily arise. Furthermore, there exists a set of plausible moral intuitions, under which the Repugnant Conclusion never occurs.

comparison list covers up the clash between these outside observer judgments. An unbounded comparison list, whether in this context or in others, means that no well-defined rank ordering of alternatives is possible. Any postulated alternative can always be trumped by some other alternative. There is no best alternative and no worst alternative. Unboundedness, by chipping away at the notion of a coherent ranking more generally, disguises the contradictory conceptions of cardinal utility, and of a "marginally worthwhile life," given by the postulation of the Repugnant Conclusion.

Again, this paper does not have a forcing argument in favor of modal interdependence of value. It cannot be proven that the very small utility description for the muzak and potato life – small enough to avoid the Repugnant Conclusion -- is the "correct" one. Instead a more modest proposition has been demonstrated. If an ideal observer or "objective" account of a marginal worthwhile life takes all intuitions into account, when setting the numerical value of that life, it is possible to value that life in such a way so that the Repugnant Conclusion never arises. There exists a set of reasonable and coherent intuitions that avoid the Repugnant Conclusion altogether, without violating the mere addition principle or other plausible moral axioms. The Repugnant Conclusion is no longer an impossibility result.

I view these results as favorable towards consequentialist ethics, while shaping the direction that such an ethic will take. We can be consequentialists, without fearing the Repugnant Conclusion. This bodes well for Parfit's search for what he calls "Theory X." At the same time, however, the results militate against strict cardinal utilitarianism. To avoid the Repugnant Conclusion we must at some level heed the lessons of Plato's Philebus, and use philosophic wisdom to help set the scale for cardinal utility and the value of life. Whereas Sidgwick sought to make intuitional values commensurate through the medium of utility, we have ended up making differing kinds of utility commensurate through the medium of philosophic intuition.

References

Armour, Leslie. "Infini Rien" Pascal's Wager and the Human Paradox. Carbondale and Edwardsville: Southern Illinois University Press, 1993.

Benardete, Seth. The Tragedy and Comedy of Life: Plato's Philebus. Chicago: University of Chicago Press, 1973.

Blackorby, Charles and David Donaldson. "Social Criteria for Evaluating Population Change." *Journal of Public Economics*, 1984, 25, 13-33.

Blackorby, Charles, Walter Bossert, and David Donaldson, "Critical-Level Utilitarianism and the Population-Ethics Dilemma." *Economics and Philosophy*, 1997, 13, 197-230.

Carlson, Erik. "Mere Addition and Two Trilemmas of Population Ethics." *Economics and Philosophy*, 1998, 14, 283-306.

Cowen, Tyler. "Normative Population Theory." *Social Choice and Welfare*, 1989, 6, 33-43.

Cowen, Tyler. "What Do We Learn From the Repugnant Conclusion?" *Ethics*, July 1996, 106, 754-775.

Griffin, James. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press, 1986.

Jehle, Geoffrey A. *Advanced Microeconomic Theory*. Englewood Cliffs, New Jersey: Prentice Hall, 1991.

Jordan, Jeff. *Gambling on God: Essays on Pascal's Wager*. Lanham, Maryland: Rowman & Littlefield Publishers, 1994.

McClennan, Edward F. "Pascal's Wager and Finite Decision Theory." In *Gambling on God: Essays on Pascal's Wager*. Lanham, Maryland: Rowman & Littlefield, 1994, 115-138.

Mackie, J.L. *Persons and Values, Selected Papers, Volume II*. Oxford: Clarendon Press, 1985.

*Nelson, Mark. "Utilitarian Eschatology." *American Philosophical Quarterly*, 1991.

Ng, Yew-Kwang. "Bentham or Bergson? Finite Sensibility, Utility Functions, and Social Welfare Functions." *Review of Economic Studies*, 1975, 545-569.

Ng, Yew-Kwang. "What Should We Do About Future Generations?" *Economics and Philosophy*, 1989a, 5, 235-251.

Ng, Yew-Kwang. "Hurka's Gamble and Methuselah's Paradox: A Response to Cowen on Normative Population Theory." *Social Choice and Welfare*, 1989b, 6, 45-49.

Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.

Parfit, Derek. "Overpopulation and the Quality of Life." In *Applied Ethics*, edited by Peter Singer. New York: Oxford University Press, 1986, 145-64.

Rescher, Nicholas. *Pascal's Wager: A Study of Practical Reasoning in Philosophical Theology*. Notre Dame, Indiana: University of Notre Dame Press, 1985.

Ryberg, Jesper. 1996. "Parfit's Repugnant Conclusion." *The Philosophical Quarterly*, April 1996, 202-213.

Samuelson, Paul A. "St. Petersburg Paradoxes: Defanged, Dissected, and Historically Described." *Journal of Economic Literature*, 15, 1, March 1977, 24-55.

Sorensen, Roy A. In *Gambling on God: Essays on Pascal's Wager*. Lanham, Maryland: Rowman & Littlefield, 1994, 139-159.

Temkin, Larry S. *Inequality*. New York: Oxford University Press, 1993.