What Do We Learn from the Repugnant Conclusion?

Tyler Cowen

Ethics, Vol. 106, No. 4 (Jul., 1996), 754-775.

Stable URL:
http://links.jstor.org/sici?sici=0014-1704%28199607%29106%3A4%3C754%3AWDWLFT%3E2.0.CO%3B2-3

Ethics is currently published by The University of Chicago Press.

What Do We Learn from the Repugnant Conclusion?*

Tyler Cowen

I. INTRODUCTION

In a series of articles on population theory, culminating in his 1984 book *Reasons and Persons,* Derek Parfit presented dilemmas for utilitarian and consequentialist moral theories.[1] Parfit's work has led to renewed interest in the theory of optimal population. More generally, Parfit is searching for a general theory of beneficence—"Theory X"—that also will cover population comparisons. Theory X corresponds to Kenneth Arrow's notion of a social welfare function—both attempt to provide a generic formula or algorithm for ranking social outcomes on the basis of their characteristics.

So far, normative population theory remains at an impasse. The proposed population standards imply implausible or counterintuitive moral conclusions at one point or another. Neither average utilitarianism, total utilitarianism, combinations of average and total principles, nor the consideration of nonutility values offers a clear path through the thicket of possible paradoxes.

Parfit's Repugnant Conclusion is the most serious obstacle which normative population theories must face. The Repugnant Conclusion (explained in more detail below) postulates a society with a large amount of total utility obtained by having very many persons living at near-zero levels of utility. Most (although not all) consequentialist

* I am grateful to Thomas Hurka, Gregory Kavka, Thomas Schelling, Alex Tabarrok, an anonymous editor, and especially to Richard Arneson, Derek Parfit, Daniel Sutter, and Larry Temkin for useful comments and discussions.

[1] See Derek Parfit, *Reasons and Persons* (New York: Oxford University Press, 1984). Parfit offers a slightly revised discussion in his "Overpopulation and the Quality of Life," in *Applied Ethics,* ed. Peter Singer (New York: Oxford University Press, 1986), pp. 145–64. A number of important essays are collected in *Obligations to Future Generations,* ed. Richard I. Sikora and Brain M. Barry (Philadelphia: Temple University Press, 1978). See also Jonathan Glover, *Causing Death and Saving Lives* (Harmondsworth: Penguin, 1977). More recent references are provided throughout this article.

Ethics 106 (July 1996): 754–775

moral theories must rank this highly populated outcome higher than other, presumably more desirable, societies. While few philosophers accept the desirability of the Repugnant Conclusion world, its endorsement has proven surprisingly difficult to avoid. Social welfare functions which avoid the Repugnant Conclusion usually run afoul of other moral intuitions.

I interpret the difficulties of normative population theory in terms of a more general problem with consequentialist ethics. Specifically, I examine an impossibility theorem for ranking social outcomes when we treat different ethical values as commensurable. I show that we cannot find a social welfare function that satisfies four chosen axioms. Apparently reasonable intuitions about the boundedness of particular ethical values—that is, the maximum weight that one value can receive when compared with all others—turn out to be inconsistent with other plausible moral intuitions or do not cover the entire range of potential paradoxical results.[2]

If we consider the four axioms behind the impossibility theorem acceptable, we should abandon the search for Theory X and should instead study the implications of impossibility theorems for how we reason about ethics. Alternatively, we might consider at least one of the four axioms unacceptable. In this case the impossibility theorem shows which moral intuitions we must revise to construct Theory X and gives us a clue about how Theory X might look. The following exercise should be interpreted as a diagnostic which clarifies the tradeoffs involved with accepting or rejecting particular moral axioms. The Repugnant Conclusion is implicitly an account of conflict between different values and how we aggregate those values.

The impossibility result is analogous to the violations of Arrow's impossibility theorem studied in the social choice literature. In the literature on social choice, certain specific paradoxes of voting, such as cycling (with majority rule it can be true that *A* beats *B*, *B* beats *C*, but *C* beats *A*), led to broader results—specifically, no method of aggregating ordinal preferences can avoid violating certain plausible axioms, such as transitivity. Just as cycling is a special case that illustrates a more general problem with voting rules, the Repugnant Conclusion (and the other paradoxes presented below) is a special case that illustrates a more general problem with aggregating ethical values to rank outcomes. Both Arrow's theorem and the Repugnant Conclusion serve a diagnostic function in helping us revise our moral intuitions.

[2] The notion of boundedness I consider involves trading off different ethical values at extreme ranges, not the evaluation of infinities or infinite expected values. On the relevance of the latter issue for population comparisons, see Tyler Cowen and Jack C. High, "Time, Bounded Utility, and the St. Petersburg Paradox," *Theory and Decision* 25 (1988): 219–23.

II. THE REPUGNANT CONCLUSION

We can imagine a society which welfare dominates many highly attractive alternatives, simply by having a very large population. Given that each life in the highly populous society is worth living, if only barely, we can multiply the number of marginally worthwhile lives to obtain a welfare-dominating result.

Parfit's 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."[3]

The more populous society can always welfare dominate if it contains enough lives. Feasibility considerations are ignored deliberately in this comparison. The point is not whether the highly populated society is possible but rather whether we would prefer it if it were possible. We ask this hypothetical question to clarify our thoughts on how utilities should be compared to other values. Specifically, it raises the question of how total utility is to be bounded, if at all.[4]

The Repugnant Conclusion does not suggest that the highly populated society is the best society imaginable. An equally populated society with higher amounts of prosperity and culture would be better. Instead, the argument implies that for any plausible social welfare function, we can always present a pairwise comparison where that social welfare function implies preferring an unappealing alternative.

The Repugnant Conclusion is truly repugnant. Parfit is not postulating a world where individuals eke out a mediocre living and drive home to a small home in the suburbs. Rather, each individual experiences only a very small amount of joy in his or her entire life. Human existence is a vast monotony which, although not painful, contains little of value. Parfit refers to a world of "muzak and potatoes."[5] In the limit, we can imagine the utility of this life as no more than the smallest nonzero epsilon.

[3] Parfit, *Reasons and Persons,* p. 387.

[4] Peter J. Hammond ("Consequentialist Demographic Norms and Parenting Rights," *Social Choice and Welfare* 5 [1988]: 127–45) attempts to escape the Repugnant Conclusion by invoking feasibility constraints. He focuses on the costs of large families. Of course the Repugnant Conclusion can be generated with a very large number of small families rather than by assuming large families.

[5] See Parfit, "Overpopulation and the Quality of Life," p. 148. Some, like Partha Dasgupta ("Lives and Well-Being," *Social Choice and Welfare* 5 [1988]: 103–26), portray Parfit's version of the Repugnant Conclusion as full of the "wretched of the earth" (p. 117) and thus full of lives with a negative standard of living or not worth living. But the muzak-and-potatoes version of the Repugnant Conclusion is more difficult to escape through this maneuver.

We cannot sidestep the Repugnant Conclusion simply by invoking nonutility values. The highly populated society may have less dignity, less culture, and less justice. But it still has a huge amount of utility. Even if we place nonutility values in the moral calculus, these values might be overwhelmed by sheer numbers of people once we attach a positive value to each life. The impossibility theorem in this article shows that various means of limiting the value of total utility violate at least one of the four basic axioms.

III. THE IMPOSSIBILITY THEOREM

I present the dilemma behind the Repugnant Conclusion in terms of a more general problem with comparing conflicting values. To generalize the relevant issues, I consider ethical theories which rank different social states by weighing and comparing the values that describe each state. No consequentialist moral theory can satisfy all of the following four axioms: (1) universal domain, (2) the value of total utility, (3) value pluralism, and (4) the nonvanishing value axiom.

Objections to these axioms are considered in more detail in Section IV below; I do not intend the initial presentation as a series of arguments for accepting the axioms as a moral theory. The axioms simply mark some starting points for the subsequent discussion of the relevant moral issues.[6] I will now consider each axiom in more detail.

Axiom (1).—We can rank all outcomes. Following Kenneth Arrow, I call this condition universal domain.[7] While few accept this postulate as literally true, we wish to avoid moral theories which beg agnosticism instead of facing up to unappealing conclusions. I use this axiom to force moral theories to rank the Repugnant Conclusion against alternatives. Admittedly, moral theory may not be able to rank a variety

[6] Several papers provide an axiomatic treatment of population issues. Bargaining axioms for dividing a fixed pie among different populations are studied by William Thomson in "Axiomatic Theory of Bargaining with a Variable Population: A Survey of Recent Results," in *Game-Theoretic Models of Bargaining,* ed. Alvin Roth (New York: Cambridge University Press, 1985), pp. 233–58. Charles Blackorby and David Donaldson ("Intertemporal Population Ethics: A Welfarist Approach," working paper [University of British Columbia, 1993]) axiomatize how the Repugnant Conclusion can be avoided when we drop the modified Pareto principle (defined below). Yew-Kwang Ng ("What Should We Do about Future Generations?" *Economics and Philosophy* 5 [1989]: 235–53) proves that the Repugnant Conclusion follows if we accept Parfit's mere addition principle (an additional happy life is to be preferred) and an axiom of nonantiegalitarianism. This axiom states that any distribution with both more utility and a more equal distribution of utility is to be preferred. Larry S. Temkin, in his *Inequality* (New York: Oxford University Press, 1993), does not formally list axioms but applies the general method of seeing what ethical views are necessary to avoid the Repugnant Conclusion; I consider Temkin's views in more detail below.

[7] See Kenneth J. Arrow, *Social Choice and Individual Values* (New York: Wiley, 1963).

of outcomes, but if we place the hard cases in this category, we give up the search for a normative population theory. Parfit, by searching for a universal or general moral theory, is at least hoping to satisfy this axiom with his Theory X.

Axiom (2).—Total utility is one value that matters in the social welfare function. I call this the value of total utility. A world-state with more utility than another is better in at least one respect, with regard to utility. This axiom does not require that the world-state with more utility is better, all things considered. Axiom (2) simply postulates total utility as a relevant moral value.

Both axioms (1) and (2) are necessary to define the problems under consideration, but they do not represent the centerpiece of the impossibility theorem. The primary clash between values is given by axioms (3) and (4). These two axioms, if understood properly, conflict to produce the impossibility result.

Axiom (3).—More than one value (e.g., total utility, freedom, justice, equality, etc.) should matter or supply relevant input for our evaluation of outcomes. I call this the value pluralism axiom.

What does it mean that more than one value should "matter"? I define this axiom more precisely as follows: First, the world contains plural values 1 through N. Second, there is no value N such that whatever the distribution of other values from 1 through $N-1$ across social states A and B, there exists some distribution of N across A and B yielding the outcome that the social state containing more N is socially preferred.

The value pluralism axiom rules out hierarchical or lexically ordered principles. The axiom also stipulates there are no interconnections between ideals such that large losses in some values can always be outweighed by gains in another value. More specifically, massive disparities between all values but one should not be outweighed by a skewed distribution of that remaining value. A society lacking in all facets but one should not be preferred on the basis of a single strength alone.

I use the value pluralism axiom to reflect the common moral judgment that the Repugnant Conclusion is indeed repugnant. If total utility were the only value that influenced social rankings, we would have no reason to object to the Repugnant Conclusion world. A pure total utilitarian should not find the Repugnant Conclusion repugnant. Our unwillingness to accept the Repugnant Conclusion therefore requires that we attach significance to some other value or values. We feel that one value, in that case total utility, ought not be able to trump all other values so easily.[8]

[8] A number of readers have suggested that the dilemma of the Repugnant Conclusion can arise with only a single value. According to this interpretation, the Repugnant

Axiom (3) does not imply that all cases of a single value trumping other values are repugnant. We might, for instance, imagine two societies with high levels of nonutility values but where one society is preferred over the other because of its higher level of total utility. There is nothing obviously repugnant in such a ranking. Axiom (3) does not rule out such a ranking (i.e., does not assert repugnance) but rather considers comparisons more generally—"Whatever the distribution of other values from 1 through $N-1$ across social states A and B." The view that total utility should dominate the final method of ranking, regardless of other values, is what we find repugnant and what the axiom rules out.

Axiom (4).—No value should become infinitely small in importance at the margin. A very large addition to that value, all other things being held equal, should never translate into an asymptotically insignificant contribution to the social welfare function. I call this the nonvanishing value axiom. In the discussion that follows, I apply this axiom to the particular value of total utility.

I define this axiom more precisely as follows: Consider a comparison of two social states, which differ with regard to several values, 1 through N. For any distribution of values 1 through $N - 1$, there should always exist a sufficiently large quantity of value N that is socially preferable to an increment of some other value or values. A precise mathematical interpretation of this claim requires that value N never be asymptotically diminishing at any margin.

The nonvanishing value axiom, like the value pluralism axiom, represents an intuition about boundedness. Value pluralism implied that no single value should be able to dwarf all others in importance, no matter how large that single value becomes. The nonvanishing value axiom tells us that no single value should be dwarfed in importance by any others, at the margin, no matter how large that single value becomes. Each axiom reflects a different aspect of boundedness—an ethical value should not be allowed to become infinitely large or infinitely small in importance.

These two intuitions about boundedness clash to provide the basic impossibility result. From axioms (1) and (2), it follows that utility must be commensurable with other values. If not, different alternatives cannot be compared with each other. We thus can postulate a

Conclusion can arise through a conflict between two aspects of a single value, such as total utility and average utility. This view is consistent with the substance of my presentation, although not with the semantics. I define average and total utility as two separate values. The relevant point is that the Repugnant Conclusion requires more than one input into a social welfare function; for the substance of the argument it does not matter whether we label these inputs as "different values" or "different aspects of the same value."

social welfare function, SW, which compares utility and nonutility features to produce a final evaluation. Without loss of generality, I call the nonutility values "culture" and "dignity." When comparing social state *A* and social state *B,* we compare

$$\text{SW (Total Utility}_{A}\text{, Culture}_{A}\text{, Dignity}_{A})$$

$$\text{SW (Total Utility}_{B}\text{, Culture}_{B}\text{, Dignity}_{B})$$

The Repugnant Conclusion postulates that there exists a Total Utility$_A$ sufficiently large that the social welfare of situation *A* exceeds the social welfare of situation *B,* regardless of the values assigned to culture and dignity in the two cases.

Here is where the central clash between axioms (3) and (4) enters the picture. If no single value is allowed to become asymptotically small in importance (axiom [4]), additions to that value must eventually trump all other values in importance. Consider some distribution of non-*N* values which inclines us to prefer social state *A* to social state *B*. Now increase the quantity of value *N* in state *B*. As the quantity of *N* increases, at some point one of two results must occur: (1) increases in *N* must dwarf the combined superiority of values 1 through $N - 1$ in state *A* or (2) further units of *N* must have virtually no effect on the social welfare function. Axiom (3) rules out the former alternative, and axiom (4) rules out the latter. The Appendix sketches a mathematical proof of this reasoning.

IV. WHICH AXIOMS SHOULD BE DISCARDED?

Since no social welfare function can satisfy all four axioms, the search for Theory X requires that we drop at least one axiom. Some individuals may see axiom (1)—universal domain—as the most vulnerable of the four axioms.[9] Dropping axiom (1) implies that a moral theory or social welfare function cannot sensibly rank all of the relevant alternatives. More specifically, social welfare functions may cease to provide relevant comparisons when the alternatives have a serious imbalance with regard to a particular value, such as utility.

Rejecting universal domain, however, constitutes a surrender, not a solution. We are discarding the axiom simply to forestall the Repugnant Conclusion and other paradoxes; we might as well admit that consequentialism has failed. The Repugnant Conclusion appears repugnant precisely because we initially believed that the two societies

[9] Julian Simon ("The Welfare Effect of an Additional Child Cannot Be Stated Simply and Unequivocally," *Demography* 12 [1975]: 89–105) denies that we can unambiguously measure the value or welfare effects of a life.

could be compared. If the two situations were truly incomparable or were somehow too heterogeneous to be usefully juxtaposed, no question of repugnance would have arisen in the first place.

Dropping axiom (2), the value of total utility, while it merits consideration, raises issues that lie outside the scope of this article. Purely deontological theories, such as those of Kant and Nozick, give utility no weight whatsoever and thus avoid the Repugnant Conclusion. Without rejecting such theories out of hand, I nonetheless wish to see how the Repugnant Conclusion fares in consequentialist frameworks which attach some value to total utility.[10]

Dropping axiom (3) of value pluralism, another option, leads us to accept the Repugnant Conclusion and some other morally counterintuitive conclusions presented below. Acceptance of such results would constitute dramatic ethical news. Yew-Kwang Ng, for one, accepts the unboundedness of total utility and denies that the Repugnant Conclusion is truly repugnant. Ng goes even further and denies that nonutility values matter at all, eliminating value pluralism altogether. Few other commentators, however, have accepted these conclusions, which returns us to the original dilemma.[11]

In principle, we could drop axiom (3) and try to replace it with some other axiom which avoids the Repugnant Conclusion. Yet, little would be gained by such a move. First, axiom (3) appears intuitively plausible; its main function in the impossibility theorem is to reflect the repugnance of the Repugnant Conclusion. Axiom (3) is not an obstacle to be avoided but rather reflects the dilemma that arises from aggregation. Second, axiom (3) appears to capture the relevant intuition behind our rejection of the Repugnant Conclusion. We feel that very many summed epsilon utilities ought not to count for very much when other values are lacking. Third, I have not found any alternative axiomatic means of rejecting the Repugnant Conclusion that avoids a clash with axiom (4). As long as we wish to reject the Repugnant Conclusion, we are left with axiom (4) as the most vulnerable target regardless of how plausible we find axiom (3) in a more global sense.

[10] Axiom (2) can be discarded through another maneuver. Charles Blackorby and David Donaldson ("Social Criteria for Evaluating Population Change," *Journal of Public Economics* 25 [1984]: 13–33) suggest postulating a zero value for individuals below a certain critical level of utility, either for prospective individuals or for all individuals more generally. Their suggestion violates axiom (2), the value of total utility. Furthermore, a difficult trade-off remains. If the critical level is set low, we can generate a Repugnant Conclusion with individuals with utility just above that level. If the critical level is set high, we are not counting valuable lives. For further criticism of Blackorby and Donaldson, see Yew-Dwang Ng, "Social Criteria for Evaluating Population Change: An Alternative to the Blackorby-Donaldson Criterion," *Journal of Public Economics* 29 (1986): 375–81.

[11] For Ng's defense of pure utilitarianism, see Yew-Kwang Ng, "Welfarism and Utilitarianism: A Rehabilitation," *Utilitas* 2 (1990): 171–93.

Axiom (4), the nonvanishing value axiom, is less forceful than value pluralism. Value pluralism is plausible because we believe that one value should not be able to make all other values irrelevant at the margin. Violating the nonvanishing value axiom, however, implies that only one value is made irrelevant at some margin. Even if we find it intolerable for all other complementary values to be made irrelevant, we might find it tolerable that one value, utility, sometimes be made irrelevant.

Capped Utility

Invoking capped utility allows us to sidestep the Repugnant Conclusion, albeit at the cost of violating axiom (4). Capped utility solutions place an upper limit on how much additions of utility can increase the value of the social welfare function. Under one set of solutions, increments of utility can only contribute a maximum number of points to social welfare, no matter how much utility that society contains. Alternatively, we might cap the quantity of utility that can be created through the addition of new lives or through the addition of new lives below a certain level of well-being. The most sophisticated forms of capped utility postulate a diminishing asymptote. An appropriately chosen function will imply that increasing the number of people always increases the number of utility points, but at a diminishing rate. With the proper mathematical specification, the number of total utility points can approach, but never reach, the target level.

The clash between capped utility and the nonvanishing value axiom is easy to see. Near the upper limit, or asymptotic bound, large increases in utility do not outweigh very small declines in other ethical value(s). No matter how slight the postulated decrease in other values, the net contribution of large sums of utility is even slighter, if we are close enough to the asymptote. In lieu of allowing utility to dominate all other values, we have created a margin where other values can dominate utility.

Capping the contribution of utility can prevent the Repugnant Conclusion. Even if we increase the number of people to a very large sum, the total utility of the resulting society cannot add more than a certain amount to the overall measure of goodness. We avoid having a single ethical value dominate all others in importance.

Several ethical intuitions can generate capped utility. We might believe that a good society, all things considered, is defined by certain objective goods, such as dignity or culture. These goods may hold a lexicographic priority over utility across certain margins, implying that no amount of utility can make up for very small quantities of these goods. A "perfectionist" concept of a well-ordered life or society might take precedence over the summation of constituent values, or we might believe that a "good society" is a holistic concept that cannot be broken down into different parts in additive fashion. A good society might require the strong participation of many different values, with no single value having a strong influence across all margins.[12]

The rejection of axiom (4) through capped utility appears even more plausible when we consider the formulations of some particular population dilemmas. These formulations talk of adding new individuals to a world already fairly well populated. We might be willing to attach asymptotically diminishing value to the creation of new individuals who do not currently exist, even if we do not wish to cap the importance of total utility for all other comparisons.

Despite these points, we should not hasten to discard axiom (4). The alternative ethical theories which violate axiom (4) are either vulnerable to objections or do not actually solve the aggregation problem in all contexts. The ethical theories which violate axiom (4), although they avoid narrow interpretations of the Repugnant Conclusion, do not provide generally satisfactory alternatives for normative comparisons. Section V attempts to generalize the dilemma behind the Repugnant Conclusion to analogous utility-theoretic puzzles in other contexts. By presenting these other dilemmas, I hope to show that the aggregation of conflicting values represents a deeper problem than can be solved by dropping axiom (4).

Asymmetric Treatment of Unborns

Some forms of capped utility give special status to persons already alive, relative to prospective or potential persons. The Repugnant Conclusion is avoided, for instance, if we postulate a zero or asymptotically diminishing value for additional individuals.[13]

This response does not damage the basic thrust of the impossibility theorem. The Repugnant Conclusion can be specified without reference to which individuals are born or not yet born. In its simplest form the Repugnant Conclusion (and other population dilemmas) involves a de novo comparison between two world-states with two

[12] Temkin (chap. 7) provides the seminal treatment of these issues; he defends capped utility and argues that no amount of utility can make up for very small quantities of certain other goods, if the quantity of total utility is already high. Parfit considers a perfectionist standard in his "Overpopulation and the Quality of Life," p. 163. Thomas Hurka ("Value and Population Size," *Ethics* 93 [1983]: 496–507, esp. pp. 505–6) also defends a perfectionist standard. Hurka considers different varieties of perfectionism, including lexicographic views, in his *Perfectionism* (New York: Cambridge University Press, 1993), esp. chap. 6. C. D. Broad (*Examination of McTaggart's Philosophy* [Cambridge: Cambridge University Press, 1938], pt. 2, chap. 56) defends capped utility on the basis of a holistic theory of value.

[13] Parfit, in *Reasons and Persons,* still provides the seminal discussion of this issue. See also Blackorby and Donaldson, "Intertemporal Population Ethics," and "Social Criteria for Evaluating Population Change."

different sets of individuals. The necessity of handling de novo comparisons springs from axiom (1), universal domain. The question of whether unborns should be granted special status need not arise; all potential individuals start with the same contingent status. Asymmetric treatment of unborns, even if plausible, therefore does not eliminate the basic dilemma. Similarly, dilemmas analogous to the Repugnant Conclusion can be constructed using only people who are initially alive (see Sec. V below).[14]

Asymmetric Treatment for Low-Utility Individuals

Another version of capped utility suggests that total utility becomes asymptotically unimportant when, and only when, that utility is created through the successive addition of low-utility individuals. Thomas Hurka and Yew-Kwang Ng provide one version of this view. They define a social welfare function which treats numbers in nonlinear fashion. Average utility is multiplied by the number of people in existence, with a dampening function being applied to the number of people as that number increases. As the number of people increases, N becomes successively less important in the social welfare calculations and average utility becomes successively more important. This proposal caps the amount of utility that can be created by adding low-utility lives, even though the value of improving already existing lives is not capped.[15]

The Hurka-Ng solution provides one of the more promising means of avoiding the Repugnant Conclusion. Nonetheless, this solution does not eliminate the more general problems of boundedness discussed below in Section V. Furthermore, the Hurka-Ng solution is operationally equivalent to postulating interaction effects, a view which I criticize directly below.

Capping the Net Contribution of Utility through Interaction Effects

Another set of asymptotic social welfare functions postulates negative effects on other values as the number of low-utility individuals rises. The increases in population and utility that produce the Repugnant Conclusion, for instance, might push down some other social value, such as dignity. As we add successive numbers of low-utility individu-

[14] Dasgupta refers to de novo comparisons as "Genesis Problems."

[15] Hurka ("Value and Population Size") defends what he calls a "variable value" view, where the value of a life depends upon how many others are alive. Hurka gives no mathematical specification, but to avoid the Repugnant Conclusion and to satisfy axiom (2) Hurka's approach must rely on asymptotically diminishing utility. Ng ("What Should We Do About Future Generations?" pp. 244–50) presents a version of capped utility, although he prefers to accept the Repugnant Conclusion. Ng had first suggested this solution in his "Social Criteria for Evaluating Population Change."

als, the total number of dignity points goes down as the total number of utility points rises. Once the number of individuals becomes high enough, dignity points become negative. With proper specification of the function, the net contribution of each individual to social welfare—a function of utility points and dignity points—diminishes asymptotically. Interaction effects between utility and other values can avoid the Repugnant Conclusion without rendering all utility increases of negligible value.

By dignity I mean the ability of an individual to achieve some pattern-based conception of the good life. Negative dignity implies that with regard to dignity, an individual detracts more from society than he or she adds.[16]

In this context I use dignity only as an illustrative example. The same arguments will hold if we can find some other value that declines as we multiply the number of low-utility individuals. The nonutility value, however, must decline through negative effects associated with the increase in utility itself. By assumption, the additional individuals do not harm others or in any way affect the rest of the world; we can imagine them being born on a distant planet. The new individuals bring no negative externalities. Interaction effects therefore arise only when the presence of new individuals or new utilities affects some pattern-based value used to evaluate the overall worth of a society.[17]

Under these social welfare functions, the creation of high-utility individuals does not necessarily lower dignity points. The net contribution of new, low-utility individuals to social welfare is capped and asymptotically diminishing, even though utility per se is not capped.

Interaction effects provide a relatively promising means out of the Repugnant Conclusion. Postulating a negative interaction between utility and dignity violates only axiom (4), the nonvanishing value axiom. As we have already seen, the nonvanishing value axiom is the least persuasive of the four axioms. Interaction effects do not rule out the importance of utility in more general situations. Utility has a vanishing importance at the margin only when we try to increase utility by adding large numbers of low-utility individuals.

The case for such interaction effects is nonetheless far from airtight. Interaction effects avoid the Repugnant Conclusion only by making the dignity value of a low-utility individual negative when the

[16] I do not wish to push the definition of dignity or use of the dignity concept too hard. As we will see below, I will reject this solution rather than argue for it, and thus I do not cover the other potential weaknesses that this argument may bring. For the remainder of this section I speak of dignity but also refer to whichever other nonutility values might turn negative with population growth.

[17] Shelly Kagan ("The Additive Fallacy," *Ethics* 99 [1988]: 5–31) stresses the importance of interaction effects in ethics.

total number of individuals in society is high. This link between large numbers and negative dignity values is vulnerable on a number of fronts.

Whenever the net contribution of a single life diminishes asymptotically with numbers, the social welfare function creates scope for a massive revaluation of human lives. The asymptotic decline of the importance of total utility, as numbers increase, implies that large numbers of human lives would suddenly be worth much less (or more) than we had thought if we discovered that mistakes in the census had underestimated (or overestimated) the earth's population or if intelligent, rights-bearing extraterrestrial beings were discovered.[18]

Some revaluation of existing lives, as new numbers are discovered, may well be morally defensible, perhaps owing to a holistic view of societal value. Nonetheless, interaction effects allow this revaluation to nearly eliminate the initially postulated value for a life or group of lives. A very large group of low-utility human beings could lose all of its initial value, minus epsilon, if a sufficiently large new population were discovered.

Interaction effects also do not capture our intuitions about the repugnance of the Repugnant Conclusion. We do not see the teeming masses of the Repugnant Conclusion as a huge benefit only to be outweighed in their magnificence by some even greater disgrace, such as lack of culture or dignity. Instead, we see each muzak-and-potatoes life as not counting for very much. We believe that the sheer addition of such lives should not add up to much; we do not necessarily feel that the addition of such lives entails some massive loss in terms of other values. In this regard, the cruder capped utility solutions, for all of their drawbacks, capture the relevant intuitions more closely than interaction effects do.

In addition, interaction effects do not capture our intuitions about the nature of dignity. For interaction effects to prevent the Repugnant Conclusion, large numbers of low-utility individuals must eventually acquire negative dignity, not merely zero dignity. The net asymptotic effect caused by population growth comes from increasingly negative values for dignity.

The use of interaction effects to derive an overall asymptotic value for total utility appears suspiciously motivated by the desire to avoid the Repugnant Conclusion rather than by any particular microfoundations from a theory of dignity. We have some intuitions about how numbers of people translate into social welfare; that is, we believe that some kind of cap should be present. Our inability to express this

[18] Blackorby and Donaldson ("Intertemporal Population Ethics") have even axiomatized this belief in terms of a separability axiom.

intuition in a convincing fashion leads us to give dignity a very particular role—supporting a net asymptotic effect for utility—in the social welfare function.

The postulated solution to the Repugnant Conclusion is not robust to small changes in our notion of dignity. We might, for instance, plausibly believe any of the following about dignity: (1) the lesser dignity of low-utility individuals, compared to high-utility individuals, springs from their lower level of well-being and not from their numbers, (2) negative dignity does not eat up most of the value of a life, even as the number of individuals becomes very large, and (3) an individual whose life is worth living can never have a negative dignity value.

Any of these three beliefs about dignity, if true, would invalidate the use of interaction effects, on the basis of dignity, to avoid the Repugnant Conclusion. The point is not that these three beliefs are necessarily compelling. Rather, it seems odd that our avoidance of the Repugnant Conclusion requires that we reject these three beliefs about dignity. Our dislike of the Repugnant Conclusion appears more fundamental than the positions we might take on the microfoundations of dignity or whatever other interacting value we choose.[19]

Most generally, the attribution of negative dignity to large numbers of low-utility individuals does not resolve the problem of boundedness per se. It only attempts to limit the net contribution of large numbers of people to the social welfare function. Capping utility across the dimension of numbers simply modifies the forms in which Repugnant Conclusion–like comparisons arise. Rather than multiplying the number of persons in society, we can create other utility-theoretic dilemmas by changing the distribution of utility among given individuals. Section V shows how additional conundrums can be created, even if we can cap the value of total utility obtained through population increases.

V. THE GENERALITY OF THE PROBLEM OF BOUNDEDNESS

The various means of bounding utility provide ad hoc attempts to avoid the Repugnant Conclusion and related paradoxes. The postulated bounds can be circumvented by constructing other comparisons or counterexamples with different logical structures.

[19] James L. Hudson ("Diminishing Marginal Value of Happy People," *Philosophical Studies* 51 [1987]: 123–37) offers further criticisms of theories of capped utility that relate the diminution of utilities for the number of people in existence. First, such theories assume a clear-cut standard for measuring personal identity, that is, how many people there are. Second, Hudson claims that such theories have special difficulty in dealing with the welfare of animals, which are presumably not part of the same numbering scheme.

The impossibility theorem presented above extends beyond the comparisons involved in population economics. The essence of the Repugnant Conclusion is to overwhelm nonutility values by adding together many small utilities. Similar comparisons of utilities arise in many situations, not just in comparisons between different populations. The theory of optimal population was simply an area where the relevant ethical dilemmas were discovered in some vivid and compelling forms.

The following discussion will present a number of aggregation dilemmas analogous to the Repugnant Conclusion. In each case we must compare an aggregation of many very small utilities to some other value or set of values. The point is not that these dilemmas resemble the Repugnant Conclusion in every regard or that they provide perfect analogies. Rather, the differences between the Repugnant Conclusion and these other aggregation dilemmas do not matter as long as the issue of boundedness remains unresolved. In the given comparisons, aggregate utility overwhelms all other nonutility values. The nonutility differences between the Repugnant Conclusion comparison and other aggregation comparisons become irrelevant for how the social welfare function ranks outcomes. I am not defending the moral propriety of this forced irrelevance but rather use it to show the far-reaching nature of the problem of boundedness.

The same point can be made by redescribing what is at stake in the literature on the Repugnant Conclusion. The Repugnant Conclusion appears to raise at least three different moral issues. Does adding more people make an outcome better? How do we trade off values of great intensity against values of lesser intensity? And how do we weigh the interests of the many against the few? Without meaning to downplay the importance of these questions, I am focusing on the even broader issue of boundedness. Rather than addressing the above questions directly, this article, through generalizing the Repugnant Conclusion, is asking a more primitive question: How can we have a moral theory where these questions matter at all? Without a defensible concept of boundedness underlying the social welfare function, any particular moral issue can be made to appear irrelevant to our final evaluations of social states.

The generalizations of the Repugnant Conclusion illustrate the difficulty of developing a satisfactory concept of boundedness. Specifically, these generalizations weaken the case for pinpointing axiom (4) as the source of the ethical dilemmas discussed above. The particular means of bounding utility, suggested in the preceding section above, all fall vulnerable to at least one counterexample. If we bound the utility generated through newborns, aggregation dilemmas can be reformulated with currently existing persons (see all four aggregation dilemmas which follow). If we bound the utility enjoyed by low-utility individuals, aggregation dilemmas can be reformulated using high-utility individuals (again, see the four aggregation dilemmas which follow). More generally, these other aggregation dilemmas suggest (but do not prove) that other potential means of bounding utility also will be vulnerable to counterexamples. Axiom (4) is admittedly not fully compelling, but these additional aggregation paradoxes suggest that dropping that axiom would not solve the deeper problem—we do not yet have satisfactory conceptual means for bounding ethical values.

Intrapersonal Analogies

The Repugnant Conclusion produces a welfare-dominating population solution by adding epsilon-valued lives. Analogously, we might produce a welfare-dominating solution for a single life by adding epsilon-valued years. Consider Methuselah's Paradox.[20]

Methuselah's Paradox.—For any possible ecstatically happy and profound life of, say, two hundred years, we can imagine another, much longer life which will welfare dominate it simply by adding many years of epsilon utility. We can have potatoes and muzak for aeons.

Methuselah's Paradox does not disappear if individuals discount utility positively. The social welfare function would still prefer a world full of Methuselah-like creatures with a zero discount rate for utility.[21]

Note that under Parfit's theory of identity, Methuselah's Paradox and the Repugnant Conclusion do become closely analogous. Parfit, in his *Reasons and Persons,* argues that personal identity is a matter of degree; there is no difference in kind between different persons and

[20] On Methuselah's Paradox, see Parfit, "Overpopulation and the Quality of Life"; Tyler Cowen, "Normative Population Theory," *Social Choice and Welfare* 6 (1988): 33–43. Yew-Kwang Ng ("Hurka's Gamble and Methuselah's Paradox: A Response to Cowen on Normative Population Theory," *Social Choice and Welfare* 6 [1989]: 45–49) argues that Methuselah's Paradox should be accepted; John McTaggart takes a similar position in his *The Nature of Existence* (Cambridge: Cambridge University Press, 1927), vol. 2, pp. 452–53. Broad (pt. 2, pp. 687–88) argues against accepting the long life. Broad claims that we cannot evaluate lives piecemeal but must consider the entire pattern of a life; in effect, he rejects the nonvanishing value axiom by denying that additional years always contribute significantly to the social welfare function.

[21] Charles Blackorby and David Donaldson ("Normative Population Theory: A Comment," *Social Choice and Welfare* 8 [1991]: 261–67) attempt to use utility discounting to defuse the paradox. For arguments that positively discounting utility is irrational, see Tyler Cowen and Derek Parfit, "Against the Social Discount Rate," in *Philosophy, Politics, and Society,* ed. Peter Laslett and James Fishkin, 6th ser. (New Haven, Conn.: Yale University Press, 1992), pp. 144–61. We should not prefer a smaller quantity of utility now over a greater quantity of utility later, once we have adjusted for uncertainty and applied the appropriate ceteris paribus conditions. Discounting utility cannot be justified by the arguments used to justify discounting dollar magnitudes or physical consumption streams. Utility is what is left over after we discount physical streams of consumption goods.

a "single" person at different points in time. Given this belief, it should not matter whether the epsilon utilities are distributed across different persons at one point in time or distributed across time in a "single" person.

Like many utility-theoretic paradoxes, Methuselah's Paradox can be reversed to produce a comparison of opposite extremes. Would we prefer a very happy life of one hundred years, or would we prefer a much shorter period of happiness, say, just one hour long, with a very intense burst of delight? Oliver Sacks cites a statement of the Russian author Dostoyevsky (who was an epileptic): "You all, healthy people, can't imagine the happiness which we epileptics feel during the second before our fit. . . . I don't know if this felicity lasts for seconds, hours, or months, but believe me, *I would not exchange it for all the joys that life may bring*" (Sacks's emphasis).[22]

Distributing Marginal Changes in Utility

The Conundrum of the Cure develops an analogue to the Repugnant Conclusion in a situation where we must distribute a life-saving technology.

The Conundrum of the Cure.—The earth is inhabited by a very large number of persons of equal age who are all faced with the prospect of immediate death through disease. Two different life-saving technologies are available, kidney dialysis and a complete cure for the threatening disease. Dialysis prolongs everyone's life for an additional thirty years but impoverishes society because the dialysis machines are so costly. Each remaining life would be worth living, but only by epsilon utility. The second alternative, the cure, would give two billion individuals an additional thirty years of healthy and happy life. But most persons would die immediately because not enough cures can be manufactured.[23]

Unlike in the Repugnant Conclusion, we are not choosing population size de novo; we are evaluating how future utilities should be distributed across individuals who are already living. Nonetheless, we are asking whether the remaining population should resemble the highly populated world of the Repugnant Conclusion or whether it should resemble a less populated society that is richer in other values. We again confront the boundedness of total utility—this time separated from the issue of how to value unborns and future generations.

[22] See Oliver Sacks, *The Man Who Mistook His Wife for a Hat* (New York: Summit, 1985), p. 137.

[23] For an earlier version of this example and its relation to the literature on inequality, see Tyler Cowen, "Distribution in Fixed and Variable Numbers Problems," *Social Choice and Welfare* 7 (1990): 47–56.

The issue of boundedness is not restricted to cases where individuals are close to the zero point for utility. The Paradox of the Chairs raises distributional dilemmas even when everyone is assured of life well above the zero point.

The Paradox of the Chairs.—A very large number of persons inhabit the world. These persons lead happy, fulfilled lives. A resource windfall now comes along. We can take twenty million of these people and give them fantastically happy lives. A greater total of utility can be created, however, by giving each person a very, very, small pleasure. We would increase the utility of each individual by epsilon by making all chairs just a little bit more comfortable. Which outcome should we prefer?[24]

Negative Utilities

Issues of boundedness arise with negative utilities as well as positive utilities.

Prometheus's Paradox.—Prometheus is chained to a rock and must endure dreadful tortures for the next two hundred million years. Providing that the number of individuals in society is large enough, we can imagine an outcome which is even worse. We might make all chairs slightly more uncomfortable for everyone. The sum of the negative utilities would be greater than the suffering experienced by Prometheus. If we can prevent only one bad, we should remedy the condition of the chairs.

VI. IMPLICATIONS

In each of the above dilemmas, the unbounded summation of utilities threatens to overwhelm the importance of other values. Like the Repugnant Conclusion, these examples raise the question of whether we are willing to accept total utility as a potentially unbounded magnitude. In addition, the multiplicity of such cases questions whether we have a morally feasible procedure for bounding ethical values. Even if we can think of particular contexts where axiom (4) might not be fully plausible, a utility aggregation dilemma could be reformulated in some other context.

We remain without a fully acceptable social welfare function for comparing situations with different, albeit commensurable, values. We

[24] Some writers attempt to avoid the repugnance of the Repugnant Conclusion by redefining the zero point for utility or by questioning whether we can think accurately about lives in the range of the zero point; see J. L. Mackie, *Persons and Values* (New York: Oxford University Press, 1985); Dasgupta. The Paradox of the Chairs shows that similar issues of boundedness arise even when we are not near the zero point for any particular life. We can imagine other versions of the chairs paradox which involve changes in the length of life. Rather than making all chairs more comfortable, we might give each person in society an extra minute of life containing an epsilon of utility.

have, however, made progress on several fronts. First, we have seen the difficulty of capturing apparently reasonable intuitions about the boundedness of any particular ethical value. Our intuitions about boundedness deserve closer examination, as do the technical and mathematical tools used to analyze boundedness.

Second, we should not treat philosophical thought experiments and counterintuitive examples as knockdown arguments. The impossibility theorem implies that every social welfare function (satisfying certain postulates) will be vulnerable to such examples. I am not suggesting that such examples be ignored or that they do not highlight defects in moral algorithms. But the importance of such examples may well be comparative. We need to consider which social welfare functions—and which apparently plausible axioms—imply the least unappetizing conclusions or deal with paradoxes most successfully.

Third, we should be especially skeptical of arguments that attempt to establish an ethical position by ruling out all feasible alternatives. We need to argue for moral views directly. If every possible social welfare function can be shown to look bad by some example or another, argument by elimination becomes a less convincing procedure.[25]

Finally, utility aggregation dilemmas imply that Arrow's impossibility theorem, in modified form, is more far-reaching than had been thought. Cardinal utilities do not offer a way out of Arrow's problem, as has often been suggested. Instead, cardinal utilities only shift the problem of dictatorship to another level.[26]

The impossibility theorem presented above attempted to weigh commensurable but potentially conflicting ethical values; the social welfare functions of Arrow's theorem consider only incommensurable ordinal pı eferences for outcomes. Starting with a fixed-number, fixed-individuals situation, Arrow examined algorithms for ranking social states of affairs. From such a base, Arrow showed that no social welfare function could satisfy simultaneously the following axioms.[27]

[25] As an example for the argument by elimination technique, I have in mind Shelley Kagan's *The Limits of Morality* (New York: Oxford University Press, 1989). Kagan attempts to establish an argument for strong duties to others by ruling out all the arguments that might weaken such claims.

[26] On the use of cardinal utilities to resolve Arrow's paradox, see Kenneth W. S. Roberts, "Interpersonal Comparability and Social Choice Theory," *Review of Economic Studies* 37 (1980): 421–39; Amartya Sen, *Choice, Measurement and Welfare* (Cambridge, Mass.: MIT Press, 1984); Charles Blackorby, David Donaldson, and John A. Weymark, "Social Choice with Interpersonal Utility Comparisons: A Diagrammatic Introduction," *Social Choice and Welfare* 25 (1984): 327–56.

[27] For an accessible presentation of Arrow's theorem, see Allan M. Feldman, "A Very Unsubtle Version of Arrow's Impossibility Theorem," *Economic Inquiry* 12 (1974): 534–46.

Axiom (1*a*).—The Pareto condition: If everyone prefers *A* to *B*, the social welfare function should rank *A* above *B*.

Axiom (2*a*).—Universal domain: The social welfare function must rank all possible outcomes.

Axiom (3*a*).—Transitivity: If *A* is preferred to *B* and *B* is preferred to *C*, *A* must be preferred to *C*.

Axiom (4*a*).—Nondictatorship: The final social ordering should not merely reflect the ordering of a single individual.

Axiom (5*a*).—Independence of irrelevant alternatives: How we rank *A* versus *B* should be independent of how we rank *C* versus *D*.

Arrow showed that when the social welfare function relies on ordinal rankings generated from conflicting preferences (or conflicting values), the social welfare function will be characterized by dictatorship. The logic behind this result is compelling, once we accept the axioms. If we consider a basic two-person situation, the absence of preference commensurability implies that we cannot resolve value clashes without giving one person his or her way. Arrow's proof showed that the more complicated *N*-person case always can be partitioned into a series of two-person comparisons. Any algorithm used to resolve two-person clashes of preference, when applied consistently, will make one person the dictator in any clash of preference.

Many economists and philosophers have pointed out that we can sidestep Arrow's theorem by the introduction of cardinal utility, that is, information about preference intensity. When the utilities of different individuals can be compared, social welfare functions can escape the dictatorship result. Returning to the simple two-person case, we might use a function that ranks outcomes according to their aggregate utility, thus abandoning axiom (5*a*).[28]

The utility aggregation dilemmas studied in the article show that introducing cardinal utility information and eliminating axiom (5*a*) does not ensure a reasonable social welfare function. The resulting social welfare function will satisfy Arrow's remaining axioms but will not simultaneously satisfy the four axioms outlined in this article, or avoid the dilemmas raised by Parfit.

Cardinal utility information, by allowing for commensurable preferences across persons, shifts but does not eliminate the problem of dictatorship. Arrow's dictatorship problem disappears because social

[28] Arrow's independence of irrelevant alternatives axiom, on examination, rules out cardinal utilities. By asserting the independence of the *A* vs. *B* comparison from the *C* vs. *D* comparison, this axiom appears innocuous. Limiting the social choice procedure to pairwise comparisons eliminates all information about intensity of preference. When placed in an axiomatic framework, the separate consideration of differing pairs allows the representation of no more than ordinal preferences.

welfare functions must no longer arbitrate between irreconcilable ordinal preferences. With cardinality, however, the potential for a different kind of dictatorship arises. Once values are treated as commensurable, one value may swamp all others in importance and trump their effects. In the case of the Repugnant Conclusion, total utility is the trumping value. The possibility of value dictatorship, when we must weigh conflicting ends, stands as a fundamental difficulty, regardless of how much information we allow into the social welfare function.

APPENDIX
SKETCH OF A MATHEMATICAL PROOF

We start with a basic mathematical property:

(1) A monotone bounded sequence is convergent in *R,* the set of real numbers. Calculus texts treat this as an axiom of mathematics, or as a basic property of the set of real numbers.[29]

Now compare two social states:

$$(\text{Total Utility}_{\text{A}},\ \text{Culture}_{\text{A}},\ \text{Dignity}_{\text{A}}) \text{ and} \tag{2}$$

$$(\text{Total Utility}_{\text{B}},\ \text{Culture}_{\text{B}},\ \text{Dignity}_{\text{B}}). \tag{3}$$

$\text{Culture}_{\text{A}}$ and $\text{Dignity}_{\text{A}}$, by assumption, are much smaller than $\text{Culture}_{\text{B}}$ and $\text{Dignity}_{\text{B}}$. How (2) ranks against (3), of course, may depend upon their respective levels of total utility. Axiom (1), universal domain, enters by requiring that (2) and (3) be ranked. Axiom (2) allows us to consider total utility as a relevant ethical value.

Now consider a sequence which results from mapping increasing levels of total utility, combined with a fixed $\text{Culture}_{\text{A}}$ and $\text{Dignity}_{\text{A}}$, into social welfare. The successive levels of social welfare compose the numbers of the sequence. Axiom (3), value pluralism, requires that this sequence be bounded. The technical definition of boundedness runs as follows:

> Definition 1: A sequence $(\text{SW}_n)\, n \in N$ is said to be bounded if there exists an $L \in R_+$ such that $|\, \text{SW}_n \,| \leq L$ for every $n \in N$.[30] Boundedness implies a maximum value for the social welfare that can result from very high levels of utility. All the terms of a bounded sequence must be contained within some finite interval. Without boundedness we must endorse the Repugnant Conclusion, given that multiplication of the population can produce a very high level of total utility.

Now, the sequence defined above will be monotone and increasing; that is, increases in total utility will increase social welfare, as implied by axiom (4), the nonvanishing value axiom.

[29] See Robert G. Bartle and C. Ionescu Tulcea, *Calculus* (Glenview, Ill.: Scott Foresman, 1968), p. 93.

[30] Ibid., p. 85.

DEFINITION 2: The sequence (SW_n) is monotone and increasing if $SW_1 \leqslant SW_2 \leqslant SW_3 \leqslant \ldots SW_n \leqslant \ldots$[31]

From this definition, and from Definition (1), the sequence (SW_n) is monotone and bounded. Axiom (1) presented above implies that a monotone bounded sequence is convergent in *R*, the set of real numbers. Now consider the definition of convergence.

DEFINITION 3: A sequence (SW_n) of real numbers converges to SW $\in$ R if for every $\tau > 0$ there exists some SW $\in N$ (depending on $\tau > 0$ and on the sequence) such that

$$| SW_n - SW | \leqslant \tau$$

whenever $n \leqslant$ SW.[32]

The definition of a converging sequence implies that for any positive τ, however small, we can find some comparison between successive elements in the sequence that is smaller than this positive sum. This directly violates axiom (4), the nonvanishing value axiom, thus validating the impossibility theorem.

[31] Ibid., p. 93. If we reverse the inequalities, the sequence still would be monotone but would also be decreasing, which does not fit the problem at hand, hence the addition of the qualifier 'increasing'.

[32] Ibid., p. 80.