CRITICAL-LEVEL POPULATION PRINCIPLES AND THE REPUGNANT CONCLUSION

by

Charles Blackorby, Walter Bossert and David Donaldson

OCTOBER 2002

Discussion Paper No.: 02-16

DEPARTMENT OF ECONOMICS
THE UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, CANADA V6T 1Z1

http://www.econ.ubc.ca

Critical-Level Population Principles and the Repugnant Conclusion*

Charles Blackorby, Walter Bossert and David Donaldson

October 2002

Charles Blackorby: Department of Economics, University of Warwick and GREQAM, c.blackorby@warwick.ac.uk

Walter Bossert: Département de Sciences Economiques and CIREQ, Université de Montréal, walter.bossert@umontreal.ca

David Donaldson: Department of Economics, University of British Columbia, dvdd@telus.net

* Financial support through a grant from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

October 2, 2002

Abstract

Critical-level generalized-utilitarian population principles with positive critical levels provide an ethically attractive way of avoiding the repugnant conclusion. We discuss the axiomatic foundations of critical-level generalized utilitarianism and investigate its relationship to the sadistic and strong sadistic conclusions. A positive critical level avoids the repugnant conclusion. We demonstrate that, although no critical-level generalized-utilitarian principle can avoid both the repugnant and strong sadistic conclusions, principles that avoid both have significant defects.

Keywords: Population Ethics, Critical-Level Generalized Utilitarianism, Repugnant Conclusion.

1. Introduction

Population principles extend fixed-population social goodness relations so that they can rank alternatives with different populations and population sizes. Most of the population principles commonly discussed are welfarist: the ranking of any pair of alternatives depends on the well-being of those alive in the two alternatives only. Thus, information about all those who ever live together with their levels of lifetime utility (interpreted as individual indicators of lifetime well-being) is sufficient to establish a welfarist social ranking. Furthermore, these population principles are typically anonymous: information about individual identities is not needed. Anonymity provides a solution to Parfit's [1984] 'non-identity problem' and ensures that individual interests receive equal treatment.

Because information about well-being plays such an important role in welfarist principles, it is important to couple them with a comprehensive account of well-being such as that of Griffin [1986] or of Sumner [1996]. In addition, the interpretation of individual utilities as indicators of lifetime well-being is essential to avoid counter-intuitive recommendations regarding the termination of lives.

In order to investigate the ethical properties of population principles, it is important to know the level of well-being that represents neutrality. We follow the standard convention and identify a neutral life with a lifetime-utility level of zero. See, for example, Blackorby, Bossert and Donaldson [1997, 2002] or Broome [1993] for discussions of neutrality and its normalization.

Within the class of welfarist population principles, variable-population extensions of fixed-population utilitarianism play a dominant role. Fixed-population utilitarianism ranks any two alternatives with the same individuals alive in both by comparing their total or average utilities. There are many ways of extending fixed-population utilitarianism to a variable-population framework, and we call a population principle whose fixed-population subprinciples are utilitarian a same-number utilitarian principle. Standard examples are classical utilitarianism and average utilitarianism. Classical utilitarianism ranks any two alternatives by comparing their total utilities, whereas average utilitarianism employs average utilities instead. Other examples include number-dampened utilitarianism which uses average utility multiplied by a positive-valued function of population size to rank alternatives. Classical utilitarianism is obtained if this function is proportional to population size, and average utilitarianism results if the function is constant. See Ng [1986] and, for variations and further discussions, Arrhenius [2000], Blackorby, Bossert and Donaldson [2001], Carlson [1998], Hurka [2000] and Sider [1991].

Critical-level utilitarianism (Blackorby, Bossert and Donaldson [1995, 1997, 2002] and Blackorby and Donaldson [1984]) is another class of principles which generalizes classical utilitarianism (but not average utilitarianism). It uses the sum of the differences between individual utility levels and a fixed critical level to make comparisons.[1] If the critical level is zero, classical utilitarianism results. For each value of the critical-level constant, a different principle is obtained.

Critical-level generalized utilitarianism makes same-number comparisons by using the sum of transformed utilities, where the transformation can be any continuous and increasing function. For convenience, we consider only transformations that preserve the level of utility representing neutrality; this involves no loss of generality. If the transformation is chosen to be (strictly) concave, (strict) inequality aversion obtains as a property of the principle. Broome [2002] argues that generalized-utilitarian orderings with strictly concave transformations provide the best fit with Parfit's [1997] 'prioritarianism' (see also Fleurbaey [2002]). A same-number generalized-utilitarian principle is any principle whose same-number subprinciples are generalized-utilitarian.

Parfit [1976, 1982, 1984] criticizes classical utilitarianism on the grounds that it implies the repugnant conclusion. A population principle implies the repugnant conclusion if and only if, for any population size, for any positive level of utility and for any level of utility between zero and the previous level, there exists a larger population size such that an alternative where everyone in the larger population has the lower level of utility is better than the alternative with the smaller population and the higher utility for everyone.[2] The higher utility level can be arbitrarily large and the lower utility level can be arbitrarily close to zero, the level that represents a neutral life. The generalized counterpart of classical utilitarianism suffers from the same problem.

There are many classes of principles that avoid the repugnant conclusion. The purpose of this paper is to defend the critical-level generalized-utilitarian principles with positive critical levels as the ones that avoid it in the most ethically attractive way.

Arrhenius [2000] introduces two versions of the sadistic conclusion and argues that it should be avoided as well. A principle implies the sadistic conclusion if and only the addition of individuals with negative utilities can lead to a better alternative than the addition of a possibly different number of individuals with positive utilities to a utility-unaffected initial population. The strong sadistic conclusion is implied if and only if, for any alternative in which everyone's utility is negative, there exists a worse alternative in which everyone's utility is positive. We argue that the requirement that the sadistic conclusion be avoided is too strong: virtually all same-number generalized-utilitarian principles imply it.

The critical-level generalized-utilitarian principles have an important property called existence independence: rankings of alternatives are independent of both the utilities and number of unaffected individuals. There are, however, no population principles that satisfy

[1] Fixed critical levels are proposed by Parfit [1976, 1982, 1984].

[2] Parfit's statement of the repugnant conclusion is somewhat weaker.

this axiom together with several basic axioms and avoid the strong sadistic and repugnant conclusions.[3] We therefore investigate several principles that avoid the repugnant and strong sadistic conclusions and satisfy the basic axioms. In particular, we discuss the restricted critical-level generalized-utilitarian principles. Like all other principles with the same properties, including restricted or unrestricted number-dampened generalized utilitarianism (Hurka [2000], Ng [1986]), these principles fail to satisfy existence independence. Using an example, we argue that principles that violate this axiom are inconsistent with widely held ethical intuitions.

In Section 2, we introduce critical-level generalized utilitarianism and discuss its axiomatic foundation. The repugnant conclusion, the sadistic conclusion and the strong sadistic conclusion are discussed in Section 3, along with a result that specifies the critical levels that allow critical-level generalized-utilitarian principles to avoid each of them. Section 4 presents and discusses restricted critical-level generalized utilitarianism and Section 5 concludes.

2. Critical-level generalized utilitarianism

A population principle ranks alternatives according to their social goodness. We assume that each social alternative is associated with a full description of all features that may be relevant to the ranking. In particular, all determinants of individual well-being are included. The goodness ranking is assumed to be an ordering, that is, a reflexive, transitive and complete at-least-as-good-as relation. Reflexivity requires every alternative to be ranked as at least as good as itself. Transitivity ensures that the ranking is consistent in the sense that, if one alternative is at least as good as a second which, in turn, is at least as good as a third, then the first is at least as good as the third. Finally, a relation is complete if and only if any two distinct alternatives are ranked. Two alternatives are equally good if and only if each is at least as good as the other. Alternative x is better than alternative y if and only x is at least as good as y and y is not as least as good as x.

We restrict attention to welfarist principles, each of which is equivalent to a single ordering defined on utility distributions. One alternative is at least as good as another if and only if the utility distribution corresponding to the first is at least as good as the distribution corresponding to the second.

A utility distribution consists of the utility levels of all the people who ever live in the corresponding alternative. Because we consider anonymous principles only, it is not necessary to keep track of individual identities. Consequently, the utility levels in an alternative can be numbered from one to the number of individuals alive. Thus, if there

[3] For further discussions, see Arrhenius [2000], Blackorby, Bossert, Donaldson and Fleurbaey [1998], Blackorby, Bossert and Donaldson [2001], Blackorby and Donaldson [1991] and Ng [1989].

are n people alive in an alternative, a utility distribution is an n-tuple $u = (u_1, \ldots, u_n)$ where each number in the list is the utility level of one of the members of society.

A welfarist population principle can be described by an at-least-as-good-as ordering of utility distributions. The corresponding equal-goodness and betterness relations are defined as above: utility distribution u is as good as utility distribution v if and only if u is at least as good as v and v is at least as good as u; and u is better than v if and only if u is at least as good as v and it is not the case that v is at least as good as u. In order to be a population principle, the ordering must be capable of different-number comparisons: any two distributions $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_m)$ are ranked, even if the population sizes n and m are different.

Because the principles we investigate are anonymous, if we relabel the utility levels in a utility distribution u, the resulting distribution is as good as u. Such a relabeling is called a permutation of a utility distribution. A permutation of $u = (u_1, \ldots, u_n)$ is a utility distribution $v = (v_1, \ldots, v_n)$ such that there exists a way of matching each index i in u to exactly one index j in v such that $u_i = v_j$. For example, (u_2, u_1, u_3) is a permutation of (u_1, u_2, u_3). Anonymity is defined as follows.

Anonymity: For all population sizes n, for all utility distributions $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$, if v is a permutation of u, then u and v are equally good.

We also assume that the ordering satisfies the strong Pareto principle. It requires unanimity to be respected.

Strong Pareto: For all population sizes n and for all utility distributions $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$, if $u_i \geq v_i$ for all $i = 1, \ldots, n$ with at least one strict inequality, then u is better than v.

Continuity is a condition that prevents the goodness relation from exhibiting 'large' changes in response to 'small' changes in the utility distribution.

Continuity: For all population sizes n, for all utility distributions $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ and for all sequences of utility distributions $u^1, u^2, \ldots$ where $u^j = (u_1^j, \ldots, u_n^j)$ for all j,

(a) if the sequence $u^1, u^2, \ldots$ approaches v and u^j is at least as good as u for all j, then v is at least as good as u;

(b) if the sequence $u^1, u^2, \ldots$ approaches v and u is at least as good as u^j for all j, then u is at least as good as v.

All of the above properties impose restrictions on same-number comparisons only and they are well established and accepted in the literature on same-number social evaluation.

To establish a link between utility distributions with different population sizes, we impose two further conditions. The first requires that well-being and population size can be traded off in at least a rudimentary way.

Weak existence of critical levels: There exist a utility distribution $u = (u_1, \ldots, u_n)$ and a utility level c such that $u = (u_1, \ldots, u_n)$ and $(u, c) = (u_1, \ldots, u_n, c)$ are equally good.

A critical level for a utility distribution u is a utility level such that, if an individual with the critical level is added to u, all other utilities unchanged, the augmented distribution and the original are equally good. Although the above axiom only requires the existence of a critical level for a single utility distribution, it does not require critical levels to exist for others. If critical levels exist, they may depend on both the number of people alive and their utilities. If strong Pareto is satisfied, each utility distribution can have at most one critical level: given transitivity, the assumption that there are two distinct critical levels immediately contradicts strong Pareto.

Finally, we introduce an independence condition. It requires the ranking of any two alternatives to be independent of the existence of individuals who ever live and have the same utility levels in both. It allows population principles to be applied to affected individuals only.

Existence independence: For all utility distributions $u = (u_1, \ldots, u_n)$, $v = (v_1, \ldots, v_m)$ and $w = (w_1, \ldots, w_r)$, the utility distribution (u, w) is at least as good as the utility distribution (v, w) if and only if u is at least as good as v.

To illustrate the condition, consider the following example (see Blackorby, Bossert and Donaldson [2001]). Suppose that, in the near future, a small group of humans leaves Earth on a space ship and, after travelling through space for several generations, establishes a colony on a planet that belongs to a distant star. The colonists lose all contact with Earth soon after their departure and, in all possible alternatives, the two groups have nothing to do with each other from then on. No decision made by the members of either group can affect the other in any way. Now suppose that the colonists are considering an important decision for their society and want to know which of the associated alternatives is best. If the population principle satisfies existence independence, the individuals that remained on Earth and their descendants can be disregarded: the ranking of the feasible alternatives is independent of their existence and, therefore, of both their number and their utility levels. In this case, the population principle can be applied to the colonists alone.

We find existence independence ethically attractive because of examples such as this. In the presence of anonymity, existence independence cannot be reserved for particular groups: if it applies to groups such as the long dead, it must apply to all groups. Existence independence is attractive for practical reasons as well. Information about the number and utility levels of the long dead or of future people whose existence and well-being are unaffected by decisions taken in the present is very difficult to obtain. The same observation applies to the spaceship example: it is impossible for the colonists to gather reliable information about the number of individuals on Earth and their utilities.

Two classes of principles that are of particular interest in this paper are the critical-level utilitarian principles and their generalized counterparts. According to critical-level utilitarianism, there exists a fixed critical level of utility α such that a utility distribution $u = (u_1, \ldots, u_n)$ is at least as good as a utility distribution $v = (v_1, \ldots, v_m)$ if and only if the sum of the differences between the utility levels in u and α is no less than the sum of differences between the utility levels in v and α. That is,

$$[u_1 - \alpha] + \ldots + [u_n - \alpha] \geq [v_1 - \alpha] + \ldots + [v_m - \alpha].$$

Critical-level generalized utilitarianism uses a continuous and increasing transformation g applied to the individual utilities instead of the utilities themselves to establish the social ranking. According to these principles, utility distribution $u = (u_1, \ldots, u_n)$ is at least as good as distribution $v = (v_1, \ldots, v_m)$ if and only if

$$[g(u_1) - g(\alpha)] + \ldots + [g(u_n) - g(\alpha)] \geq [g(v_1) - g(\alpha)] + \ldots + [g(v_m) - g(\alpha)] \qquad (1)$$

where, as before, α is the fixed critical level. Without loss of generality, we can assume that the transformation g preserves the utility level representing neutrality, that is, it satisfies $g(0) = 0$.

Same-number generalized-utilitarian principles give priority to worse-off individuals if and only if the transformation g is strictly concave. Suppose that a single person is to be chosen to receive a one-unit increase in his or her utility level. According to any of these principles, the best choice is the worst-off person, the second-best choice is the second-worst-off person, and so on (if two people have the same utility level, either can be chosen). Thus, strict priority is given to worse-off individuals.

The critical-level generalized-utilitarian principles are the only ones that satisfy the above axioms. This result, which is proved in Blackorby, Bossert and Donaldson [1998], provides a strong case in their favour.[4]

Theorem 1: *A welfarist population principle satisfies anonymity, strong Pareto, continuity, weak existence of critical levels and existence independence if and only if it is critical-level generalized-utilitarian.*

[4] See also Blackorby, Bossert and Donaldson [1995] for an intertemporal formulation. An alternative characterization can be found in Blackorby and Donaldson [1984].

3. The repugnant and sadistic conclusions

A population principle implies the repugnant conclusion (Parfit [1976, 1982, 1984]) if population size can always be substituted for quality of life, no matter how close to neutrality the well-being of a large population is. That is, there are situations where mass poverty is considered preferable to alternatives where fewer people lead very good lives. An informal definition of the repugnant conclusion is given in the introduction. The following formulation makes it more precise.

Repugnant conclusion: For any population size n, for any positive utility level ξ and for any utility level ε strictly between zero and ξ, there exists a population size $m > n$ such that a utility distribution in which each of m individuals has the utility level ε is better than a utility distribution in which each of n individuals has a utility of ξ.

The sadistic conclusion, introduced by Arrhenius [2000], refers to the comparison of two alternatives both of which are obtained by population expansions. The sadistic conclusion is implied if and only if it may be better to add people with negative utilities to a utility-unaffected population than adding a possibly different number of people with positive utility to the same utility-unaffected population.

Sadistic conclusion: There exist utility distributions $u = (u_1, \ldots, u_n)$, $v = (v_1, \ldots, v_m)$ and $w = (w_1, \ldots, w_r)$ such that all utilities in v are negative, all utilities in w are positive and the distribution (u, v) is better than the distribution (u, w).

The strong sadistic conclusion obtains if and only if, for every utility distribution of negative utilities, there exists a worse utility distribution of positive utilities.

Strong sadistic conclusion: For any utility distribution $u = (u_1, \ldots, u_n)$ containing negative utilities only, there exists a utility distribution $v = (v_1, \ldots, v_m)$ with positive utilities only such that u is better than v.

Clearly, the sadistic conclusion does not imply the strong sadistic conclusion; for example, critical-level utilitarianism with the critical level $\alpha = 0$, which is classical utilitarianism, leads to the sadistic conclusion but not to the strong sadistic conclusion. Conversely, the strong sadistic conclusion does not imply the sadistic conclusion. Define a value function V by letting $V(u) = -u_1$ if the utility distribution u has exactly one component u_1, and $V(u) = u_1 + \ldots + u_n$ if the utility distribution u has at least two components. The goodness relation is defined by declaring one utility distribution to be at least as good as another if and only if the value of V for the first is greater than or equal to the value of V for the second. This ordering leads to the strong sadistic conclusion because, for any utility distribution $u = (u_1, \ldots, u_n)$ containing negative utilities only, the distribution

$v = v_1 = (-1)(u_1 + \ldots + u_n - 1)$ which contains positive utilities only is worse. The sadistic conclusion is avoided because any two utility distributions (u, v) and (u, w) each have at least two components and, therefore, are compared according to total utility. But total utility is always greater for (u, w) than for (u, v) if the components of v are negative and those of w are positive.

The above example fails to satisfy the strong Pareto principle. If strong Pareto is added, the strong sadistic conclusion is stronger than the sadistic conclusion. We obtain

Theorem 2: *If a principle satisfies strong Pareto and the strong sadistic conclusion, then it satisfies the sadistic conclusion.*

Proof. Suppose strong Pareto and the strong sadistic conclusion are satisfied. Consider a utility distribution $u = (u_1, \ldots, u_n)$ with at least two individuals such that all utilities are negative. By the strong sadistic conclusion, there exists a utility distribution $v = (v_1, \ldots, v_m)$ with positive utilities only that is worse than u. We can distinguish two cases.

(i) $m > 1$. Because u_n is negative, strong Pareto implies that $(u_1, \ldots, u_{n-1}, 0)$ is better than $u = (u_1, \ldots, u_{n-1}, u_n)$. Analogously, because v_m is positive, strong Pareto implies that $v = (v_1, \ldots, v_{m-1}, v_m)$ is better than $(v_1, \ldots, v_{m-1}, 0)$. Because u is better than v, transitivity implies that $(u_1, \ldots, u_{n-1}, 0)$ is better than $(v_1, \ldots, v_{m-1}, 0)$. This means that adding the utility distribution $(u_1, \ldots u_{n-1})$ with negative utilities only to (0) is better than adding the utility distribution $(v_1, \ldots, v_{m-1})$, which has positive utilities only, to (0). Thus, the sadistic conclusion is implied.

(ii) $m = 1$. In this case, v has a single positive component v_1. By strong Pareto, $v = (v_1)$ is better than $(-v_1)$. The strong sadistic conclusion implies that there exists a utility distribution $w = (w_1, \ldots, w_r)$ with positive utilities only that is worse than $(-v_1)$. Strong Pareo implies that $r \neq 1$ and thus $r > 1$. Because u is better than v, $v = (v_1)$ is better than $(-v_1)$ and $(-v_1)$ is better than w, transitivity implies that u is better than w. Analogously to the argument used in case (i), strong Pareto implies that $(u_1, \ldots, u_{n-1}, 0)$ is better than $u = (u_1, \ldots, u_{n-1}, u_n)$ and $w = (w_1, \ldots, w_{r-1}, w_r)$ is better than $(w_1, \ldots, w_{r-1}, 0)$. Because u is better than w, transitivity implies that $(u_1, \ldots, u_{n-1}, 0)$ is better than $(w_1, \ldots, w_{r-1}, 0)$. This means that adding the utility distribution $(u_1, \ldots u_{n-1})$ with negative utilities only to (0) is better than adding the utility distribution $(w_1, \ldots, w_{r-1})$, which has positive utilities only, to (0). Thus, again, the sadistic conclusion is implied. ■

The following theorem identifies the values for the critical level α such that the corresponding critical-level generalized-utilitarian principles avoid each of the above conclusions.

Theorem 3: *(i) A critical-level generalized-utilitarian principle implies the repugnant conclusion if and only if the critical level α is non-positive.*

(ii) A critical-level generalized-utilitarian principle implies the sadistic conclusion if and only if the critical level α is non-zero.

(iii) A critical-level generalized-utilitarian principle implies the strong sadistic conclusion if and only if the critical level α is positive.

Proof. (i) Suppose α is non-positive. Let n be any population size, let ξ be any positive level of utility, and let ε be a utility level strictly between zero and ξ. Let m be a population size such that

$$m > n\left[\frac{g(\xi) - g(\alpha)}{g(\varepsilon) - g(\alpha)}\right]. \tag{2}$$

Because $\xi > \varepsilon$, α is non-positive and g is increasing, the ratio $[g(\xi) - g(\alpha)]/[g(\varepsilon) - g(\alpha)]$ is greater than one and, by (2), m is greater than n. Multiplying both sides of (2) by the positive difference $g(\varepsilon) - g(\alpha)$, we obtain

$$m[g(\varepsilon) - g(\alpha)] > n[g(\xi) - g(\alpha)]$$

and the utility distribution where m people each have utility ε is better than the distribution where n people each have utility ξ. Thus, the repugnant conclusion is implied.

Conversely, suppose α is positive. Let $n = 1$, $\xi = 2\alpha$ and $\varepsilon = \alpha/2$. Substituting these values, for any population size $m > n$, an alternative where m people have utility ε is better than an alternative where n people have utility ξ according to critical-level generalized utilitarianism if and only if

$$m[g(\varepsilon) - g(\alpha)] = m[g(\alpha/2) - g(\alpha)] > n[g(\xi) - g(\alpha)] = 1[g(2\alpha) - g(\alpha)] \tag{3}$$

which is impossible for $m > n = 1$ and $\alpha > 0$ because, in this case, the left side of (3) is negative and the right side is positive. Therefore, the repugnant conclusion is avoided.

(ii) Suppose α is not equal to zero. Let $u = (u_1) = (\alpha)$. If α is positive, let $v = (v_1) = (-\alpha/4)$ and $w = (w_1, w_2) = (\alpha/4, \alpha/4)$. If α is negative, let $v = (v_1, v_2) = (\alpha/4, \alpha/4)$ and $w = (w_1) = (-\alpha/4)$. In both cases, v contains negative utilities only and w contains positive utilities only but (u, v) is better than (u, w) according to critical-level generalized utilitarianism, which shows that the sadistic conclusion is implied.

Now suppose α is equal to zero. Let $u = (u_1, \ldots, u_n)$, $v = (v_1, \ldots, v_m)$ and $w = (w_1, \ldots, w_r)$ be utility distributions such that v contains negative components only, w contains positive components only and there are no restrictions on the utilities in u. According to critical-level generalized utilitarianism with a zero critical level, (u, v) is better than (u, w) if and only if

$$g(u_1) + \ldots + g(u_n) + g(v_1) + \ldots + g(v_m) > g(u_1) + \ldots + g(u_n) + g(w_1) + \ldots + g(w_r)$$

which is equivalent to

$$g(v_1) + \ldots + g(v_m) > g(w_1) + \ldots + g(w_r).$$

Because all the values on the left side of this inequality are negative and all values on the right side are positive, this is impossible and the sadistic conclusion is avoided.

(iii) Suppose α is positive and let $u = (u_1, \ldots, u_n)$ contain negative utilities only. Let

$$m > \frac{[g(u_1) - g(\alpha)] + \ldots + [g(u_n) - g(\alpha)]}{g(\alpha/2) - g(\alpha)} \tag{4}$$

and $v = (v_1, \ldots, v_m) = (\alpha/2, \ldots, \alpha/2)$. Note that both numerator and denominator on the right side of (4) are negative and, as a consequence, the quotient is positive. Then, multiplying both sides by the negative difference $g(\alpha/2) - g(\alpha)$, we obtain

$$[g(u_1)-g(\alpha)]+\ldots+[g(u_n)-g(\alpha)] > m[g(\alpha/2)-g(\alpha)] = [g(v_1)-g(\alpha)]+\ldots+[g(v_m)-g(\alpha)]$$

and u is better than v according to critical-level generalized utilitarianism. Consequently, the strong sadistic conclusion is implied.

Now suppose that α is non-positive and the strong sadistic conclusion is implied. Let $u = (u_1, \ldots, u_n)$ contain utilities that are less than α and, therefore, negative. By the strong sadistic conclusion, there exists a utility distribution $v = (v_1, \ldots, v_m)$ which contains positive utilities such that u is better than v. Thus, according to critical-level generalized utilitarianism,

$$[g(u_1) - g(\alpha)] + \ldots + [g(u_n) - g(\alpha)] > [g(v_1) - g(\alpha)] + \ldots + [g(v_m) - g(\alpha)]. \tag{5}$$

By construction, each term on the left side of (5) is negative and each term on the right side is positive, and a contradiction is obtained. Consequently, the strong sadistic conclusion is not implied when α is non-positive. ■

Theorem 3 implies that it is not possible for a critical-level generalized-utilitarian principle to avoid both the strong sadistic and repugnant conclusions: avoidance of the repugnant conclusion requires the critical level to be positive but the strong sadistic conclusion is avoided only if the critical level is non-positive.

That result is related to another concerning the repugnant and sadistic conclusions. Any same-number utilitarian principle which ranks no one-person alternative above all those with larger populations cannot avoid both the sadistic and repugnant conclusions (Blackorby, Bossert and Donaldson [2001, Theorem 1]). The condition on one-person alternatives is implied by existence of critical levels. Consequently, all of those principles that avoid the repugnant conclusion necessarily imply the sadistic conclusion. This occurs because avoidance of the sadistic conclusion requires the addition of any number of people at a positive but arbitrarily small utility level to be ranked as no worse than the addition of a single person at an arbitrarily small negative utility level. Because we consider the repugnant conclusion unacceptable, we conclude that avoidance of the sadistic conclusion is an axiom that must be discarded.

4. Restricted critical-level principles

Although the axiom avoidance of the sadistic conclusion can be rejected, it may be argued that avoidance of the strong sadistic conclusion should not. That requires distributions with positive utilities only to be ranked as no worse than distributions with negative utilities only. Together with Theorem 1, Theorem 3 implies that there is no population principle that satisfies anonymity, strong Pareto, continuity, weak existence of critical levels and existence independence that avoids the repugnant conclusion and the strong sadistic conclusion. If avoidance of the strong sadistic conclusion is regarded as desirable, therefore, one of the other requirements must be dropped. Given the fundamental nature of anonymity, strong Pareto, continuity and weak existence of critical levels, the obvious candidate is existence independence.

There are principles that are closely related to the critical-level generalized-utilitarian principles with positive critical levels. They are the restricted critical-level generalized-utilitarian principles (Blackorby, Bossert and Donaldson [2001]) and they satisfy anonymity, strong Pareto, continuity and weak existence of critical levels and, furthermore, they avoid both the repugnant conclusion and the strong sadistic conclusion. The positive critical level for a critical-level generalized-utilitarian principle becomes the critical-level parameter for the corresponding restricted principle. However, this parameter is no longer equal to the critical level for all utility distributions.

Each of the restricted critical-level generalized-utilitarian principles employs a value function which is equal to the left side of equation (1) when average transformed utilities are greater than $g(\alpha)$, equal to the percentage shortfall of average transformed utility from $g(\alpha)$ when average transformed utility is positive and less than or equal to $g(\alpha)$, and equal to total transformed utility less one when average transformed utility is non-positive. Consequently, all utility distributions whose average transformed utilities are above $g(\alpha)$ are better than all whose average transformed utilities are positive and no greater than it and these utility distributions are, in turn, better than all whose average transformed utilities are non-positive. Critical levels are equal to α for all utility distributions in the first group. In the second group, the transformed critical level is equal to average transformed utility. For the third group, critical levels are equal to zero.

The restricted critical-level generalized-utilitarian principles satisfy anonymity, strong Pareto, continuity, weak existence of critical levels and they avoid both the repugnant and the strong sadistic conclusions. They are not the only ones that satisfy the axioms on the above list: restricted number-dampened generalized utilitarianism (Hurka [2000]), which is a modification of number-dampened generalized utilitarianism (Ng [1986]) has those properties as well.

All of the restricted principles fail to satisfy existence independence. This is, in our view, a significant problem which is best illustrated by an example. For simplicity, we employ same-number utilitarian principles; the example is easily adapted to their generalized counterparts.

Suppose that a single parent has a handicapped child whose lifetime utility would be zero (neutrality) without the expenditure of additional resources. Two alternatives are available. In the first, which we call x, resources are devoted to improving the child's well-being, resulting in utilities of 50 for the child and 60 for the parent. In the second, which we call y, no additional resources are used to raise the level of well-being of the disabled child, but a second child is born and the same resources are devoted to it resulting in lifetime utility levels of 60 for the second child and the parent and zero for the first child. The parent and his or her children are not the only people who ever live, however. There are ten billion ($10b$) others who have the same utility levels in both alternatives. This example, illustrated in Table 1, is due to Parfit [1976, 1982].

	Parent	First Child	Second Child	Utility Distribution of Others
x	60	50		$(u_1, \ldots, u_{10b})$
y	60	0	60	$(u_1, \ldots, u_{10b})$

Table 1

The parent wants to know which alternative is better. Parfit assumes that utility levels other than those of people who are potentially affected are irrelevant. That assumption is satisfied if critical-level generalized-utilitarian principles are used to rank the alternatives. Their ranking of x and y is independent of the utility levels of the unconcerned and even of their existence.

Classical utilitarianism ranks y as better than x, and this contradicts the moral intuition of many. Critical-level utilitarianism agrees with this ranking if the critical level is less than ten but, if the critical level is greater than ten, x is ranked as better. It is interesting to note that the positive critical level that ensures this ranking also ensures that the repugnant conclusion is avoided.

Suppose, by contrast, that restricted critical-level utilitarianism with a critical-level parameter of 15 is used to rank the alternatives. In that case, the utility levels of the unaffected other people make a difference. If their average utility is equal to 20, the two alternatives are ranked with critical-level utilitarianism with a critical level of 15 and x is better than y. If the average utility of the others is 10, then average utility in both alternatives is between zero and 15, x and y are ranked with average utilitarianism and, again, x is better than y. But, if the average utility of the others is -5, average utilities in both alternatives are negative. In that case, the alternatives are ranked with classical utilitarianism and y is better than x. Restricted number-dampened utilitarianism (see Blackorby, Bossert and Donaldson [2001] and Hurka [2000]) suffers from the same problem because it fails to satisfy existence independence.

5. Conclusion

Parfit [1976, 1982, 1984] argues that the repugnant conclusion should be avoided and we concur. Arrhenius [2000] argues that the sadistic conclusion should also be avoided. Because all same-number utilitarian principles that avoid the repugnant conclusion lead to the sadistic conclusion, we reject it as an axiom.

Avoidance of the strong sadistic conclusion is an axiom that, at first glance, has some ethical appeal. There are, however, no principles that satisfy anonymity, strong Pareto, continuity, weak existence of critical levels and existence independence that avoid both the repugnant conclusion and the sadistic conclusion. It is tempting, therefore, to drop existence independence from the list of axioms. However, in that case, rankings of alternatives may depend on the utilities of unaffected people such as the long dead. Our intuitions tell us that such a dependence is ethically inappropriate and, for that reason, we are prepared to drop avoidance of the strong sadistic conclusion. That leaves the critical-level generalized-utilitarian principles with positive critical levels as the ones that avoid the repugnant conclusion in the most ethically appropriate way.

REFERENCES

Arrhenius, G., 2000, An impossibility theorem for welfarist axiologies, *Economics and Philosophy* **16**, 247–266.

Blackorby, C., W. Bossert and D. Donaldson, 1995, Intertemporal population ethics: critical-level utilitarian principles, *Econometrica* **63**, 1303–1320.

Blackorby, C., W. Bossert and D. Donaldson, 1997, Critical-level utilitarianism and the population ethics dilemma, *Economics and Philosophy* **13**, 197–230.

Blackorby, C., W. Bossert and D. Donaldson, 1998, Uncertainty and critical-level population principles, *Journal of Population Economics* **11**, 1–20.

Blackorby, C., W. Bossert and D. Donaldson, 2001, The axiomatic approach to population ethics, *Discussion Paper* **01-03**, University of British Columbia, Department of Economics.

Blackorby, C., W. Bossert and D. Donaldson, 2002, Utilitarianism and the theory of justice, in *Handbook of Social Choice and Welfare, Vol. 1,* K. Arrow, A. Sen, and K. Suzumura, eds., Elsevier, Amsterdam, 543–596.

Blackorby, C., W. Bossert, D. Donaldson and M. Fleurbaey, 1998, Critical levels and the (reverse) repugnant conclusion, *Journal of Economics* **7**, 1–15.

Blackorby, C. and D. Donaldson, 1984, Social criteria for evaluating population change, *Journal of Public Economics* **25**, 13–33.

Blackorby, C. and D. Donaldson, 1991, Normative population theory: a comment, *Social Choice and Welfare* **8**, 261–267.

Broome, J., 1993, Goodness is reducible to betterness: the evil of death is the value of life, in *The Good and the Economical: Ethical Choices in Economics and Management,* P. Koslowski, ed., Springer-Verlag, Berlin, 69–83.

Broome, J., 2002, Equality or priority: a useful distinction, in *Fairness and Goodness in Health,* D. Wikler and C. Murray, eds., World Health Organization, Geneva, forthcoming.

Carlson, E., 1998, Mere addition and the two trilemmas of population ethics, *Economics and Philosophy* **14**, 283–306.

Fleurbaey, M., 2002, Equality versus priority: how relevant is the distinction?, in *Fairness and Goodness in Health,* D. Wikler and C. Murray, eds., World Health Organization, Geneva, forthcoming.

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

Hurka, T., 2000, Comment on 'Population principles with number-dependent critical levels,' unpublished manuscript.

Ng, Y.-K., 1986, Social criteria for evaluating population change: an alternative to the Blackorby-Donaldson criterion, *Journal of Public Economics* **29**, 375–381.

Ng, Y.-K., 1989, What should we do about future generations? Impossibility of Parfit's Theory X, *Economics and Philosophy* **5**, 235–253.

Parfit, D., 1976, On doing the best for our children, in *Ethics and Population,* M. Bayles, ed., Schenkman, Cambridge, 100–102.

Parfit, D., 1982, Future generations, further problems, *Philosophy and Public Affairs* **11**, 113–172.

Parfit, D., 1984, *Reasons and Persons,* Oxford University Press, Oxford.

Parfit, D., 1997, Equality or priority?, *Ratio* **10**, 202–221.

Sider, T., 1991, Might Theory X be a theory of diminishing marginal value? *Analysis* **51**, 202–213.

Sumner, L., 1996, *Welfare, Happiness, and Ethics,* Clarendon, Oxford.