WHAT DOES "OPTIMUM POPULATION" MEAN?

David D. Friedman

I. INTRODUCTION

One weakness in the discussion of population policy that has raged for the last decade has been the almost unanimous failure of the participants to consider seriously by what criterion alternative futures containing different numbers of people ought to be judged—what, in other words, is meant by the term "optimum population."[1] A striking example is the collection of essays entitled *Is There An Optimum Level of Population?* published in 1971 by the Population Council (Singer, 1971). Most of the contributors use, explicitly or implicitly, "per capita welfare" as their criterion, apparently blissfully unaware of the arguments with which Meade, more than 20 years ago, demonstrated its implausibility (Meade, 1955:82-88). Of the two who actually discuss whether a per capita criterion is appropriate, one dismisses the question with the remark that "most of us would prefer" a per capita measure, and the other with the assertion "surely, what we want to do is to maximize the per capita share." Neither provides any arguments for his position. Of the three contributors who clearly do not see per capita welfare as the appropriate criterion, two assert, again without argument, that the appropriate criterion is the survival (for one author short-term and for one long-term) of the human race; the third wishes to maximize world brain and individual potentiality, both (fortunately) undefined.

The purpose of this article is to repair this omission in the literature, not by demonstrating that there is one clear and unambiguous criterion for optimum population to which all reasonable men must agree, but by showing the inadequacy of the criterion of per capita welfare and discussing some alternatives.

In discussing optima in ordinary economic questions, where the population is treated as given, there are three common approaches. The utilitarian approach assumes that individual utility functions are in principle knowable and comparable across individuals so that one can say whether a proposed change injures one set of people by more or less than it benefits another. A second approach uses the Pareto criterion; because interpersonal utility comparisons are assumed impossible, one alternative is said to be superior to another only if it is preferred by some and opposed by none; this has the unfortunate difficulty of providing only a very partial ordering, hence leaving many pairs of alternatives incomparable. The third approach attempts to impose reasonable conditions on an (unknown) social welfare function which is to be maximized in order to permit arguments that do not depend on the precise form of the function but do depend on its having certain characteristics. This may also assist the search for the "true" social welfare function by eliminating some, most, or (in the case of the Arrow Theorem) all of the alternatives.

Since my purpose is to illuminate those problems peculiar to population decisions, I prefer to assume away, at least initially, all other problems of defining optimality. I therefore begin by assuming, with Meade, that there exist individual utility functions that are interpersonally comparable, and discuss, in that context, the two alternatives he considered: maximizing per capita utility (in his terms, "welfare") and maximizing total utility. Having done so, I will then discuss the possibility of some Pareto-like criterion for population decisions, and lastly consider what general conditions one might impose on possible social welfare functions designed to permit comparisons of alternative futures with different populations, and then examine how well the

criteria I have considered meet those conditions.

II. PER CAPITA UTILITY AND TOTAL UTILITY

Consider two alternative futures with different populations. If our criterion is per capita utility, we compare the two futures by taking, for each, the sum of the utilities received by its inhabitants divided by their number; the higher figure defines the better society. Assuming that our utility function is defined in the sense described by Von Neumann and Morgenstern (1944:15-30), this corresponds to saying that the better society is that in which a person would prefer to be randomly placed; a "lottery" consisting of an equal chance of living any one of the society's lives will itself have a utility equal to the average of the utilities of those lives. This seems, at first glance, an unobjectionable criterion.

It involves, however, a fallacy of composition. Consider two alternative futures, each with a population of one hundred; in future A, the average utility is 80 utiles per person and in future B, it is 100. Suppose we wish to compare, not A and B, but A and A', and B and B', where A' and B' are created from A and B, respectively, by adding, in each case, one more person with a utility of 90 utiles. We assume that the additional person is in precise utility balance with the rest of his population; nobody else is helped or hurt by his existence. Imagine, if you like, that he is born, lives, and dies on a desert island without ever seeing another human being.

If our criterion is average utility, it follows that A' is superior to A, but B' is inferior to B. Whether it is desirable or undesirable for a particular life to be lived then depends, not merely on what that life is like, but on what the lives of a set of people totally uninfluenced by the additional person are like. This is, surely, an unsatisfactory result.

It may be objected that in practice people are affected by each other—an additional person living a below average life will make other people miserable, either because they feel obliged to help him at a cost to themselves or because they receive negative utility from the knowledge that someone else is less happy than themselves. If this is so, then that fact should be included in defining the utility functions of our hypothetical populations. There still remains the question of how, in principle, we choose among such populations; if our criterion is to be one of average utility then even if, after taking account of all such effects, the additional person imposes no harm on the rest of the population (perhaps because they receive benefits from his existence that balance the costs) we must still judge his existence to be undesirable because it brings down the average, even though it does so without injuring anyone.

Meade made the same argument in a somewhat different form:

> Suppose two communities A and B to exist. Suppose that neither has any appreciable economic dependence on the other so that the disappearance of A would not appreciably affect the standard of living in B nor the disappearance of B the standard of living in A. Suppose, further, that the standard of living in B is somewhat lower than in A, though both communities are prosperous and enjoy high standards. The strict application of the objective of maximizing welfare per head would lead to the conclusion that the world would be a better place if community B ceased to exist, since output per head for all citizens of A and B would certainly be increased if that section of the community with the somewhat lower standard were to cease to exist (Meade, 1955:87).

Following Meade's line of argument, it is worth noting that a strict application of the per capita

criterion implies that everyone who is less happy than the average ought to be painlessly killed (or at least, to avoid the question of means versus ends, that it would be a good thing if they all dropped dead) unless he not only benefits other and happier people, but benefits them by enough to outweigh the "injury" he imposes on the average level of utility by his existence. It also implies, given the belief of most writers on population that people in poor countries are on average much less happy than people in rich countries, that an epidemic that depopulated the poorer parts of the world would be an unambiguously good thing. While few of those who support such a criterion would be willing to carry it that far, some do seem to accept similar conclusions of a less drastic sort—in particular, the conclusion that holding down the reproduction of the poor is a good thing in itself, independent of whether the existence of poor people harms or helps the not-poor.

The obvious alternative to maximizing per capita utility is to maximize total utility; this is the alternative Meade chooses. In order to define what this means, we must first define zero utility. This was not necessary when we were concerned with per capita utility because the addition of a constant to everyone's utility function[2] results in adding the same constant to the average, hence that population that has the higher average utility before the transformation will have a higher utility after as well. This is not true for total utility; the lower the level that we define as corresponding to zero utility, the more favorable the total utility criterion is to larger populations. Going back to our previous example of A and A', the question now is whether the additional person brings with him positive or negative utility—or, if we drop the assumption that his existence does not influence others, whether the difference between his own utility and the net reduction he causes in the utility of others is positive or negative. This depends crucially on what standard of life corresponds to zero utility.

Meade recognized the problem and so defined what he called a "welfare subsistence level":

> A man may be above the basic "physical subsistence" level and yet his existence may be considered so wretched as to count as a minus quantity from the point of view of economic welfare. He must attain something appreciably above the bare physical subsistence level before he can be said to be counted as a positive contribution to economic welfare (Meade, 1955:88).

This states the problem but does not answer it; how, in principle, does one decide where the welfare subsistence level is? The answer, I think, is that since utility functions are observable in the form of choices, and since negative utility is by definition that utility below which existence is worse than nonexistence, the zero point of the utility function is that point below which a person, given the choice, would prefer not to exist. But people have that choice; for many of us, suicide is not only possible but inexpensive—all it costs is our life. For others, it is possible but expensive—those, for example, who believe that suicide, being a sin, imposes serious post mortem costs or those who value the welfare of others who would be injured by their death (dependent children, for instance).[3] Following the principle of consumer sovereignty to its somewhat grisly conclusion, it seems natural to say that if a person, having access to means for killing himself, refrains from doing so, it must be because his utility for living is greater than for dying. Making allowance for costs of dying (other than the opportunity cost of not living) such as pain, injury to others, and the price of the bullet, we than have a definition of zero utility; a person's utility is zero when he is exactly indifferent between committing a costless suicide and not doing so or when he would be indifferent to committing a costly suicide or not doing so, were his utility from living to be lowered by an amount corresponding to the costs of suicide (or were he to be offered additional benefits

from suicide, such as someone else taking care of his orphaned children, which canceled the costs).[4]

This definition of zero utility, which seems to be the only one consistent with the usual economic approach to human behavior, makes the criterion of maximum *total* utility very favorable to large populations; if zero utility is defined as the suicide point then even in a very poor society additional children bring with them into the world substantial positive utility. More precisely, if we consider two societies D and E with populations $P_0 > P_E$, E will be more attractive only if a person choosing between a lottery that gives him a $1/P_0$ chance of living each of the lives in D and one that gives him a $1/P_0$ chance of living each of the lives in E and a $(P_0 - P_E)/P_0$ chance of dying, would prefer the latter. Introspection suggests that if P_0 is substantially larger than P_E, E must be very much more attractive than D for that to happen.

We have seen that the criterion of maximizing per capita utility leads to the highly counterintuitive conclusion that the desirability of a particular life existing depends on the accident of whether the other people who happen to exist at the same time are better or worse off then the person who lives the life, so that the existence of a particular life will be judged desirable in one future and undesirable in another even if it has no interaction at all with those to whom it is being compared. It also leads, if taken seriously, to some rather unattractive recommendations for action. We now see that the alternative criterion of maximizing total utility is likely to lead to the conclusion that a world of 24 billion people living at the edge of subsistence is superior to a world of 4 billion living in prosperity—it seeming unlikely that an Indian peasant would be willing to play a game of Russian roulette with five chambers loaded, even if offered the opportunity of emigrating to the United States if he survived. The true utilitarian may reply that the result seems wrong only because we, having been brought up in affluence, fail to appreciate that the difference in utility between an Indian peasant and an American suburbanite is small compared to the difference between an Indian peasant and a corpse. Those of us who remain unconvinced may want to look for yet another criterion.

III. EXPANDED PARETO CRITERION

A. *Principles*

To do so, I replace the assumption that utility functions exist and are interpersonally comparable with the weaker assumption than an individual can in principle compare the attractiveness to him of his life with that of being another person living another life in some alternative future. While this may seem somewhat strained, it is difficult to see how we can make any statements at all about the relative attractiveness of different futures inhabited by different people without something of the sort.

Our new criterion should be chosen to avoid the problems generated by both the per capita utility and total utility criteria. It should avoid the fallacy of composition by which below average lives are treated as if, by bringing down the average, they inflicted a positive injury on the rest of the population. And it should reflect at least a decent agnosticism concerning the superiority of a future with large numbers of less happy people to one with fewer numbers of happier people. The cost of this avoidance, by analogy with the ordinary Pareto criterion, will be the failure to provide a complete ordering; some, perhaps many, pairs of alternative futures will be incomparable.

As a first try, based on the Pareto criterion and my previous argument about how zero utility should be defined, let us say that a larger society D with population P_D is superior to a smaller society E

with population P_E if D contains a subset d of size P_E for which there exists a one-to-one mapping between D and E, such that each person in E is mapped into a person in d whose life he regards as at least as attractive as his own,[5] provided that the individuals in (D - d) have a utility of at least zero in the sense defined earlier. In other words, a larger society is more attractive if it contains enough "spaces" for the population of the smaller society which the members of that society consider at least as attractive as the spaces they presently occupy, and if the remaining members of the larger society are at least sufficiently happy to prefer life to costless suicide.

Applying the same principle in the opposite direction, we could say that the smaller society is superior if it contains "spaces" for all the members of the larger society that they regard as at least as attractive as those they now occupy, where the "excess spaces" corresponding to the difference in the two populations each consist of a "life" of not living.

There are some difficulties with this. To begin with, it implies that a smaller society will never be found superior to a larger, however attractive the smaller may be, unless the larger contains a number of people, equal to the difference in the populations, who would rather be dead than alive. This seems, to put it mildly, unlikely. It further implies that a larger society will never be found superior as long as it contains at least $P_0 - P_E + 1$ lives that nobody in the smaller society prefers to his own. In eliminating our previous problems, we have come up with an ordering so partial as to be virtually nonexistent.

The situation can be improved by altering the criterion in a way that allows the attractiveness of one life to balance the unattractiveness of another. Consider two alternative societies, each of which has only two members. The better-off member of society B (call him b_1) is much better off than the better-off member of society A (a_1); the worse-off member (b_2) is slightly worse off (than a_2). Under our criterion as so far given, A and B are incomparable. They are equally incomparable under the conventional Pareto criterion, if we think of a_2 and b_2 as corresponding to two different alternatives for the same person, and a_1 and b_1 similarly. "Going from" A to B makes one person better off and one worse off; hence neither it nor the reverse change is a Pareto improvement.

But in the context of population we are comparing, not two different futures for the same set of people (as in the conventional Pareto case), but two different futures for two different populations—different in who is in them even if the two populations happen to be of the same size. *A priori*, there is no more reason to compare a_2 to b_2 than to compare him to b_1 — or to a mixture of the two. Suppose the relative attractiveness of different roles in the different populations is such that a_1 would prefer a lottery made up of a .9 chance of being b_1 and a .1 chance of being b_2 to his present life, and that a_2 would similarly prefer a .1 chance of being b_1 plus a .9 chance of being b_2. There then exists a mapping, not of people into people but of people into probability mixes of people, which uses up all of the "places" in both (equal-sized) populations, and which maps each person in A into a preferred alternative (a lottery among possible places) in B. It seems reasonable to say that if this is the case, then B is preferable to A, in a sense analogous but not identical to the normal Pareto criterion . By extension, we can say that a larger population future is superior to a smaller population future if there exists some mapping from the latter to the former that maps each individual into a probability mix of lives (a lottery with probabilities for living each of one or more of the lives lived in the larger society) such that the total probabilities in each lottery add up to one, and the sum of the probabilities with which different people are mapped into the same life adds up to no more than one, where each individual regards his lottery as at least as attractive as his present life, and where any life in the larger society for which the summed probabilities do not add up to one corresponds to a person in the larger society whose utility is not

less than zero, in the sense defined earlier. A smaller society is superior to a larger society in the same sense, with the "missing spaces" corresponding, as earlier, to nonexistence. Mathematically, we require that:

> E is preferred to D iff $\exists\ p_{ij}, (i = 1, \ldots, P_D; j = 1, \ldots, P_E)$ such that $\forall_i, \Sigma_j p_{ij} U_{ij} \geq U_{i0}$, *and* for some $i, \Sigma_j p_{ij} U_{ij} > U_{i0}$, *and* $\forall_j \Sigma_i p_{ij} \leq 1$, *and* $\forall_i \Sigma_j p_{ij} \leq 1$, *and* $\forall_j$ such that $\Sigma_i p_{ij} < 1$, $U_{0j} \geq 0$.

Here p_{ij} is the probability with which person i in D is mapped into life j in E, U_{ij} is the utility to i in D of living life j in E, U_{i0} is the utility to i in D of living his own life, U_{0j} similarly for j in E. Note that these utilities need not be interpersonally comparable.

If, however, we assume that every life has the same utility to everyone, this reduces to

> E is preferred to D iff for every $d \subset D\ \exists\ e \subset E$ such that $\Sigma_{i \in e} U_{0i} \geq \Sigma_{i \in d} U_{i0}$ *and* $\text{Min}_{i \in (E-e)} U_{0i} \geq 0$ *and* $|e| \leq |d|$ (i.e., set of lives e in E contains no more lives than set of lives d in D).

I have now at least reduced the difficulties presented by the first attempt at a Pareto-like criterion. In order for a smaller society to be judged superior, the "surplus" members of the larger society need not all prefer nonexistence to existence; it is sufficient if enough members of that society are willing to accept some risk of nonexistence (in exchange for a probability of a much more attractive life in the smaller society if they win their gamble), so that the summed probabilities of nonexistence add up to the difference between the two populations. This is a difficult requirement, but not an impossible one, if the smaller society is sufficiently attractive. I call this the expanded Pareto criterion.[6]

In some ways, this criterion seems very similar to Meade's preferred alternative of total welfare, since by the definitions of Von Neumann-Morgenstern utility, the utility of a lottery is the sum of the utility of the outcomes weighted by their probabilities. The difference is that under my Pareto-like criterion, lives may be mixed but not added; there is no way that two less happy people can "add up to" one happy person.

B. Application

Of the three criteria for optimal population that I have discussed, two (per capita and total utility) have been discussed elsewhere in this literature. The third (the expanded Pareto criterion) is not only (to the best of my knowledge) novel, it is also less easy to understand intuitively. The reader may find it useful to consider some hypothetical alternative futures and see how they would compare under the various criteria. For purposes of simplicity, I will assume that each life can be assigned a utility, such that a person choosing between two alternative lives will always prefer the one with the higher utility. I will further assume that these are Von Neumann-Morgenstern utilities—that is to say, they are so defined that a person choosing between two lotteries will choose that lottery for which the expected value of the utilities of the outcomes is higher. These assumptions are not essential for the criterion—it is sufficient that individuals can compare their lives to alternative lives and to lotteries among alternative lives—but they greatly simplify exposition.

Consider the alternative futures *f*, *g*, and *h* shown in Table 1. Comparing *f* and *g*, each of which has

a population of 3, we observe that *g* is higher in total (hence also per capita) utility. Under a conventional Pareto criterion the two are incomparable, since person 6 in *g* is worse off and person 4 is better off than anyone in *f*. To apply the expanded Pareto criterion, note that a lottery involving a 50 percent chance of being 4 and a 50 percent chance of being 6 has an expected utility of (0.5 x 7) + (0.5 x 2) = 4.5. Future *g* can be thought of as made up of two such lotteries (giving a total of 1 chance of being 4 and 1 chance of being 6) plus a certainty of being 5. Persons 1 and 2 would each prefer the lottery to his present state; person 3 is indifferent between his present state and becoming person 5, hence by the expanded Pareto criterion *g* is superior to *f*.

Table 1. Comparison of Futures

Future	*Person*	*Utility*	*Total utility*	*Per capita utility*
f	1	3	9	3
	2	3		
	3	3		
g	4	7	12	4
	5	3		
	6	2		
h	7	6	14	2.8
	8	3		
	9	2		
	10	2		
	11			

What about *h*? In terms of total utility, it is the best of all; in terms of per capita utility, it is the worst. To apply the expanded Pareto criterion for comparison of *h* with *f,* note that a lottery consisting of a 50 percent chance of being 7 and a 50 percent chance of being 9 has an expected utility of (0.5 x 6) + (0.5 x 2) = 4. Persons 1 and 2 would each prefer one such lottery to his present condition; person 3 is indifferent between his present position and "being" person 8. We hence have a mapping that maps each person in *f* into a life or lottery of lives that he regards as at least as good as his own, and at least one person (actually two) into a life or lottery of lives that he prefers to his own. The remaining lives in *h* (10 and 11), which nobody from *f* is being mapped into, have utility greater than zero, hence persons 10 and 11 "are better off alive than dead" and their presence in *h* cannot be held to make it a worse future than if they did not exist. So, by the expanded Pareto criterion, *h* is better than *f*.

In comparing *h* to *g,* on the other hand, we note that there is no way to construct three lives or lotteries of lives in *h* (remembering that each life can be used only once) that 4, 5, and 6 would

prefer to their own. There is also no way to transfer the population of *h* into the lives of *g* with satisfaction for all concerned. Persons 8 and 9 would be willing to take lives 5 and 6, respectively, and person 7 would prefer life 4 to his own. But that would leave 10 and 11 with no choice but nonexistence, which they do *not* prefer to their present circumstances. Nor is there any way of combining the lives of *g* (including two "empty spaces"—call them 6' and 6", each a life of nonexistence with utility zero) to give a set of lotteries that the inhabitants of *h* would all be willing to exchange for their present lives. Hence *g* and *h* are incomparable under the expanded Pareto criterion.

C. Multiple Generations

So far this discussion has been put in terms of comparing two futures as if they were future instants; while we have considered people living lives (which presumably takes time to do), we have not explicitly considered that the population that is being optimized consists not of those alive at some instant but of the entire path of population from the present to the end of time, i.e., the entire set of people who will ever live.[7] If we do so, we see another disadvantage of the per capita criterion in comparison with either of the other alternatives. If we are to average over the people alive at any one time, surely we should also average over those who live at different times, and it is then that grand average that is to be maximized. If we include in our average all who have ever lived, there is a strong case for trying to make future populations large; even if their members are not very well off, they are probably better off than the approximately 55 billion people who have already lived and died,[8] so increasing their numbers is likely to pull up the average. If we do not include those already dead, we have a criterion that changes at every instant; we may correctly maximize the average today and be told, 50 years hence, that we acted wrongly; by *their* criterion, the welfare of those who lived in the intervening period should be given no weight at all.

It is also interesting to note that if our criterion is the average welfare of everyone who will ever live, another objective suggested by some—the survival of our species—is not obviously desirable. If future survival must be purchased at the cost of present abstention, why should we want it? Why not burn up our resources in one burst of glory, providing a high level of utility for ourselves and our children and arranging, by appropriate contraceptive measures, that they will have no children to pay the bill? If, as many now argue, the reduction of the spatial extension of our species is entirely unobjectionable—hence desirable if it implies any increase at all in the average welfare of its members—the same ought to be true of its temporal extension as well.

IV. GENERAL CONDITIONS ON OPTIMALITY CRITERIA

Having discussed some specific optimality criteria for population, it is worth asking, in the spirit of the "social welfare function" approach, what general conditions any such criterion ought to meet. I begin by defining a "future" as a set of lives to be lived, including a complete description of all facts relevant to those lives. An optimality criterion is then an ordering (partial or complete) of futures, or in other words a set of ordered pairs (*f*,*g*) where *f* and *g* are different futures. If such an ordering includes the pair (*f*,*g*) we will say that f "is at least as desirable as" *g*. If it also includes (*g*,*f*) we will say that f "is equivalent to" *g*; if it does not include (*g*,*f*) we will say that *f* "is preferred to" *g*. If the ordering contains neither (*f*,*g*) (*g*,*f*) we will say that *f* and *g* "are incomparable" (under that ordering).

The first condition one might expect such an ordering to meet is transitivity; if *f* is at least as desirable as *g* and *g* is at least as desirable as *h*, then *f* is at least as desirable as *h*. This is necessary

in order to make the ordering correspond to our normal intuitive ideas about "as good as," "better than," and so forth. It has the further useful consequence that cycles are impossible: a set of futures *f*, *g*, and *h* such that *f* is preferred to *g*, *g* is preferred to *h*, and *h* is preferred to *f* must violate transitivity.

A second condition that seems reasonable is that the ordering be what Nozick (1974:209) calls "aggregative" and Sen (1973:39-41) "additively separable." To see what this means, consider decomposing a future (containing, say, 5 lives) into two subfutures (say, 2 lives and 3 lives). Since each subfuture is itself a future, it must include a description of all facts relevant to the lives it contains, hence a future containing, say, lives *a* and *b*, must also include a *description* of lives *c*, *d*, and *e*, insofar as they affect *a* and *b* . One may perhaps think of such a subfuture as a future in which *a* and *b* exist as people, while *c*, *d*, and *e* are robots, identical to people insofar as any effect they have on *a* and *b*, but of no normative significance in themselves. A "partitioning" of a future is then a set of subfutures in which each life of that future appears once and only once. A "natural partitioning" is one none of whose subfutures contains a description of any life included in any other of its subfutures; in other words, it partitions the future into sets of lives that in fact do not interact with each other - as in the desert island example given earlier.

An ordering is aggregative if and only if, for any two futures *f* and *g*, if *f* and *g* can each be partitioned (into $f_1, f_2, \ldots, f_n; g_1, g_2, \ldots, g_n$) in such a way that for each i, f_i is at least as attractive as g_i, then f is at least as attractive as g. And if for each i, f_i is at least as attractive as g_i, and for some i f_i is preferred to g_i, then f is preferred to g.

These two conditions correspond reasonably well to some of those conventionally applied to social welfare function. I would like to add one more condition of a rather different nature, which I call "microrationality." It is that the criterion that is to be applied to judging entire futures be one that people in fact apply, at least approximately, in judging that part of the future most relevant to themselves, i.e. the lives of themselves, those near and dear to them, and their descendants. There are two arguments for requiring this. The first is that in constructing such a criterion we are trying to generalize our normative intuitions; we feel that certain principles are appropriate in dealing with those we know and care about, and since we have no reason to believe that those we know and care about are in fact any different from other people, we believe that ideally those principles should somehow be broadened to include everyone.

The second argument is methodological rather than philosophical. In constructing optimality criteria (whether for different populations, or more conventionally for different alternatives for the same population) we usually find ourselves trying to maximize something (total utility, for instance) that we have no possible way of measuring. This seems at first impossible; if we cannot measure what total utility would be under each of several alternatives and compare them in order to choose the best, how can we make any use of the criterion "maximize total utility" (or any other such criterion)? The conventional solution is to choose a criterion that in some approximation (no externalities or zero transaction costs for the conventional Pareto criterion, for instance) is automatically maximized by the separate decisions of the individual actors. We then observe in what respects the real world deviates from that approximation and take actions that either remove the deviations or compensate for them (effluent taxes to compensate for externality effects of pollution, for example). In this way, we can hope to move the world closer to optimality without ever having to measure the quantity we are trying to optimize.[9] It is possible, of course, that there is a "morally correct" criterion for optimum population and that this criterion is not microrational. What this argument suggests is that if this is so, this criterion is unlikely to be very useful for

generating policy recommendations. It may therefore be prudent, in trying to construct a criterion, to limit ourselves to those that are microrational as well as transitive and aggregative.

Of the criteria we have considered, both per capita and total utility are obviously transitive. In order for the expanded Pareto criterion to be transitive, we must assume that in deciding whether or not I prefer my present life to some life (or lottery of lives) in a hypothetical alternative future, I can abstract away from the particular tastes I actually have and make the decision in some "objective" fashion. Otherwise it would be possible for person *i* in future *f* to prefer life *j* in future *g* to his own, for person *j* in *g* to prefer life *k* in future *h* to his own, and for person *k* in future *h* to prefer life *i* in future *f* to his own—each of the three having, of course , different tastes. Supposing that each of the three alternative futures has a population of one, *h* is then preferred to *g*, *g* is preferred to *f*, and *f* is preferred to *h*, which contradicts transitivity. We must therefore assume, in order to guarantee transitivity, that the comparison among lives can be done in some way independent of which person is doing it.

Are the criteria we have considered aggregative? Per capita utility is not (this is the point of Meade's criticism). Total utility and the expanded Pareto criterion are. Are they microrational? Casual examination of individual reproductive behavior suggests that it is not aimed at maximizing the total utility of parents plus descendants (my previous discussion of that criterion suggests that doing so would probably involve raising the reproduction rate to its biological maximum), hence that criterion is not microrational; it is at least arguable that both maximization of per capita utility (remembering that up to some point additional children may raise the utility of their parents even if by diluting the pool of resources available to the family, they lower that of their siblings) and the expanded Pareto criterion are.

APOLOGIA

To many readers, especially those who are not economists, this article may seem both irrelevant and flippant, with its discussions of zero utility, suicide points, and the desirability or otherwise of killing off most of the human race. I can only protest that ideas do have consequences, and different ideas have different consequences. If we accept, as I believe we should not, a criterion that views every below average life as presumptively evil, entitled to exist only if it pays its way by providing some benefit to offset the "harm" it does by lowering the general average, we will be led to conclusions that are, I think, both morally and intellectually indefensible.

REFERENCES

Becker, Gary. 1976. *The Economic Approach to Human Behavior.* Chicago: University of Chicago Press.

Friedman, David. 1972. *Laissez-Faire in Population: The Least Bad Solution.* New York: An Occasional Paper of the Population Council.

Meade, J.E. 1955. *Trade and Welfare.* London: Oxford University Press.

Meade, J.E. 1967. Population explosion: the standard of living and social conflict. *The Economic Journal* 77:233-256.

Ng, Yew-Kwang. 1975. Bentham or Bergson? Finite sensibility, utility functions and social welfare functions. *Review of Economic Studies 42:545 - 569.*

Nozick, Robert. 1974. *Anarchy, State and Utopia.* New York: Basic Books, Inc.

Sen, Amartya. 1973. *On Economic Inequality.* Oxford: Clarendon Press.

Simon, Julian. 1970. The per capita income criterion and natality policies in poor countries. *Demography* 7:369-378.

Simon, Julian. 1975. The welfare effect of an additional child cannot be stated simply and unequivocally. *Demography* 12:89-105.

Simon, Julian. 1975. *The Economics of Population Growth.* Princeton: Princeton University Press.

Singer, S. Fred (ed.). 1971. *ls There An Optimum Level of Population?* New York: McGraw-Hill.

Von Neumann , John and Oskar Morgenstern. 1944. *The Theory of Games and Economic Behavior.* Princeton: Princeton University Press.

---

[1] The only exceptions I know of are two articles by Julian Simon (1970) and (1975, revised in 1977, Chapter 18). In the latter, he refers to a previous discussion of mine that was in a circulated draft of Friedman (1972) but not, as he incorrectly assumed, in the published version. There is also a brief mention of the problem in Meade (1967:236).

[2] The Von Neumann - Morgenstern utility function for a single person is arbitrary with regard to linear transformations; the assumption of interpersonal comparability makes the set of utility functions arbitrary with regard to a single linear transformation applied to all utility functions simultaneously.

[3] This argument applies only to those who are dependent on the potential suicide, not those who merely care about him; in the context of the economic theory of altruism (Becker, 1976:282-294) the latter should be benefited by the suicide of someone whose utility is negative, since it raises his utility and hence theirs.

[4] A very similar definition of zero utility is given by Ng (1975:561- 562).

[5] More precisely, it is necessary that the alternative be as attractive to everyone and more attractive to at least one person; for purposes of simplicity I omit the latter qualification throughout the discussion, but it should be considered implicit in both of my Pareto-like criteria.

[6] This term seems to have been first used by Simon (1975) to describe a special case of my "first try" at a Pareto-like criterion—the case where one of the two societies being described is created from the other by adding an additional person.

[7] This point is discussed in Simon (1970).

[8] This estimate is from unpublished work by Gerald Feinberg.

[9] For an attempt at treating population decisions in this way, see Friedman (1972).