INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
RICE UNIVERSITY

THE PARADOX OF FUTURE GENERATIONS

by

EVELYN KEYES

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

MASTER OF ARTS

APPROVED, THESIS COMMITTEE

Professor Baruch A. Brody
Chairman

Professor Richard E. Grandy

Assistant Professor Larry Temkin

HOUSTON, TEXAS

MAY, 1984
© 1984

EVELYN KEYES

All Rights Reserved
Every choice that influences the future helps determine who is actually born; and different choices cause the existence of different future people. The duty of beneficence requires us to benefit people as much as possible. But, since different choices cause different future people to live, every choice benefits those future people who actually live as much as possible. Yet not all choices that influence the future seem maximally beneficent. Many proposed principles of beneficence fail to solve this "Paradox of Future Generations." The average and total principles of utility solve it by weighing beneficence to those who do live against beneficence to those who would have lived had another alternative been chosen. But both have counter-intuitive implications. We can avoid these by applying the principle of utility rationally, i.e., under a model for moral decision-making derived from four formal postulates of practical reason: universality, impartiality, respect for persons, and similarity.
ACKNOWLEDGMENTS

This paper began as a brief response to the Paradox of Future Generations as Derek Parfit developed it during the course of the Tsanoff Lectures at Rice University in March, 1982. Since then it has gone through many metamorphoses; and I now conceive of it as offering a potentially general solution to the problems of maximal beneficence as a necessary and sufficient notion of morality where human welfare is concerned, problems which the Paradox of Future Generations points up. The answers I have reached are entirely my own; but I have benefited greatly from the published and unpublished work of Derek Parfit in this area, without which I would have remained unaware of the nature, seriousness, and extent of the problems, as well as of important possible approaches to them. I have also benefited greatly from the criticisms and, above all, the encouragement of Larry Temkin; the patience to read my inchoate drafts and the unerring ability to put his finger on their weakest spots of Baruch Brody; and the unfailing good advice and kind consideration of Richard Grandy. But perhaps above all I owe a debt of gratitude to Rachel Hoffman, who typed successive drafts of the evolving manuscript, and to my husband David, upon whom I tried out the arguments that made it into the paper and the many that did not.

February, 1984

Evelyn Keyes
TABLE OF CONTENTS

I. The Paradox of Future Generations....................... 1
II. Possible Principles of Beneficence..................... 11
III. The Strengths and Weaknesses of the Impersonal Principles of Beneficence.......................... 41
IV. A Solution to the Problems: The Rational Approach to Moral Decision-Making......................... 51
V. Defense of the Model for Rational Decision-Making.............................................. 93
VI. Resolution of the Paradox of Future Generations....106
VII. Parfit's Strategy and the Mere Addition Paradox....132
I. The Paradox of Future Generations

Many moral theories hold that we have a duty to benefit people. And any moral theory that has a duty of beneficence requires a general principle of beneficence to tell us what that duty requires. Often our duty is held to be to maximize beneficence. In that case, the principle of beneficence should tell us to make that choice among our alternatives which benefits people most, or makes people as well off as possible. But it is extremely difficult to find a satisfactory version of the required principle. At the heart of the difficulty is the Paradox of Future Generations. This paradox has recently engaged the attention of a number of philosophers, among them Jan Narveson, Robert Merrihew Adams, Gregory S. Kavka, and Derek Parfit. Among these men, Parfit has done the most to explore the ramifications of the problem, but without developing a principle of beneficence that he thinks offers a satisfactory solution to the Paradox of Future Generations. In this paper I will push further in the directions explored by Parfit. I will show that we can reach a satisfactory understanding of moral beneficence which allows us to solve the Paradox of Future Generations and other problems of beneficence as well.

What is the Paradox of Future Generations, and why does it present such a problem for the theory of beneficence? The paradox is generated when a simple fact whose truth most of us will readily acknowledge is introduced into moral theory. This is the fact that if things had been other than they are I (or any given person) would not have been born. If my father and mother had not met
each other, married, and conceived me when they did, I would never have existed. If my mother had married someone other than my father, other children might have been born, but I would not have been. Even if my parents had waited a month—or an hour—to conceive a child beyond the time at which I was conceived the sperm and egg that joined to create a new human being would not have been the sperm and egg that joined to create me. And of course if this is true for me it is true for all of us. Had things been other than they are, we who exist today would not have been born.

Now we must ask how things could have been other than they are. Things would have been other than they are if they had simply fallen out differently by chance. But they would also have been different if we had chosen differently. And here is where moral theory enters the picture. For choice is the subject matter of morality.

If, instead of deciding to start their family during World War II, my parents had decided to wait until the war was over I would never have been born. So my parents' personal choice to have children when they did influenced my coming into being. But so did matters beyond their control. Some of these were matters of chance. If my father had been in perfect physical condition, rather than unfit for military service, he would have served in the American armed forces and his opportunities for conceiving children with my mother would have been severely reduced. It seems very unlikely that I would have been born. And if he had been killed in early 1942, neither I nor my siblings would ever have existed. But some of the factors that influenced my coming into being were neither matters
of personal reproductive choice, nor matters of chance, but rather policy choices.

Had the United Kingdom chosen to surrender to Hitler rather than resisting, the course of world history would have been changed—perhaps affecting my existence. The same would have been true if America had never entered the war. Or, if it seems that these events are too close in time to the time of my conception to have affected my coming into being, we can go further back. My father's maternal grandfather came from Virginia to Mississippi to serve as a surgeon for the Confederate forces at the siege of Vicksburg. He met and married my great-grandmother; and my grandmother was born in Vicksburg. My father's father immigrated from South Carolina to Mississippi after the draining of the Mississippi River Delta had opened the land to cotton planting. Had it not been for the Civil War or the draining of the Mississippi Delta my father would not have been born and neither would I. Any one of us can find parallels to this personal history in his own life. And that is precisely the point: for all of us, not only personal choices but policy choices, as well as chance, influence the future and determine who is born.

Both personal and policy choices, then, are essential parts of the cause of the existence of future people. Why does this fact matter? Well, let us assume that causing to exist can benefit. If we have lives worth living which we would not otherwise have had, it seems that causing people to exist and have lives worth living does benefit them. Parfit argues that at least it is not obviously a mistake to believe that people benefit from being caused to exist with lives worth living. And if this is a
benefit, the value of this benefit should be the full quality of those worthwhile lives which are lived that otherwise would not have been. The first of these claims is called the claim that "causing to exist can benefit" (Cf. FG, p.116) and the second follows from the Full Comparative View, which says that the value of a choice is the full difference it makes. (FG, pp.124-25) For if the value of a choice is the full difference it makes and that choice is an essential part of the cause of someone's existing and having a certain quality of life, then the full difference the choice makes is the full quality of life that is lived that otherwise would not have been. Thus the full difference a choice makes must be the outcome of that choice in terms of beneficence or quality of life. (Or welfare or utility - these are all equivalent terms.) I shall assume that causing to exist does benefit and that the value of this benefit is the outcome of the choice in terms of the quality of life lived.

But then comes the problem. Had our choices been different, those who do exist in the future would not have existed. Thus every choice that influences the future not only benefits every future person who ever lives as a result of that choice but benefits him as much as possible. Therefore all choices that influence the future are maximally beneficent to whoever actually lives. And no one who would have lived given a different choice is harmed, since he never exists and non-existence is not a harm. Consequently, every choice that gives future people lives worth living—no matter what the alternatives—seems to fulfill the duty of beneficence. But this is equivalent to saying that any choice that influences the future
is as good as any alternative, at least so long as both give future people lives worth living. And this conclusion seems obviously false.

Suppose we have a choice between a policy of Depletion and one of Conservation. (The example is Parfit's, FG, p. 118.) If we choose Depletion the quality of life will be slightly higher for the next two hundred years than if we choose conservation, but it will then fall dramatically lower. Everyone affected by either policy will nevertheless be at least as well off as we are now. If we choose Conservation, the quality of life will gradually rise to a high level and stay there. The quality of life of everyone affected will be higher on the average and in sum than if we choose Depletion. Conservation seems to be more beneficent than Depletion— but it is not more beneficent for anyone who ever actually lives. For if we choose Depletion the C-people (those who would have lived had we chosen Conservation) never exist. Hence the C-people are not harmed and the D-people (those who live if we choose Depletion) benefit from our choice as much it is possible for them to benefit. Everyone who ever lives benefits maximally from our choice of Depletion and no one is harmed. There seems to be no objection to my choosing Depletion over Conservation.

Similarly in the personal sphere, I can decide either to have a child now who will predictably be handicapped or to wait and have a predictably normal child later. (This Handicapped Child case is also Parfit's, FG, P. 118.) We can assume that the Handicapped Child will have a life worth living. And if he is born, the child who would have been conceived later will never exist, hence will not be
harmed by my choosing to conceive the Handicapped Child instead. Everyone who ever lives as a result of my decision—whichever way it goes—benefits as much as possible. There seems to be no objection from beneficence towards actual people to my choosing to conceive a predictably Handicapped Child now rather than a predictably normal one later.

We can consider the Depletion and Handicapped Child cases to be Same Number Choices. That is, we can assume that the same number of people will exist no matter which alternative we choose. Many choices are like this. The number of people who live given an alternative is an irrelevant factor in these cases. But some choices affect the number of people who live, possibly with consequences for either the average or the total quality of life. We can call these Different Number Choices.

Population policy decisions are a simple example of Different Number Choices. Whether we choose to have a policy of unlimited population expansion instead of one of limited population growth seems to make no difference if all we are morally obliged to do is to benefit as much as possible those people who actually do live, since any of these policies benefits those who actually do live under it as much as possible.

In the personal sphere, different numbers of people will live according to whether or not I decide to conceive a child. And different numbers will live if I decide to abort a fetus and not replace him than if I go through with my pregnancy. In both those cases whatever I decide, no alternative decision could have been better for those
who are actually born. But these decisions have repercussions beyond the personal sphere. For either could conceivably count as an effect of a population policy and could have an effect on the average and total quality of life welfare lived. And the latter could count as an effect of an abortion policy. But the fact that no one who ever actually lives would have benefited if I had chosen differently seems to imply that it does not matter which of two alternative personal decisions I may make or which of two contradictory population or abortion policies I may instantiate by my decision.

The conclusion that every choice that causes future people to exist and have lives worth living is as good as any other alternative choice seems plainly false. But it is equally plainly true that, whichever alternative I choose in each of the cases I have presented, everyone who ever actually lives benefits, and benefits as much as it is possible for him to benefit. Thus it seems that choices that influence the future both fulfill and do not fulfill the duty of beneficence. This is the Paradox of Future Generations. It tells us that, even though no one who ever actually lives as a result of a choice of ours would be made better off if we chose another alternative, we still need a principle of beneficence that does not call every choice which influences the future maximally beneficent.

Parfit's strategy for solving the paradox—one which I shall adopt—is to play on an ambiguity in the notion of "benefiting people." While admitting that no other choice could have been better for the people who actually do exist as a result of a choice, he attempts to find a
principle of beneficence on which it is false that no other choice could have been better for the people who would have existed had an alternative choice been made. Thus he understands "people," and "beneficence" is a wide sense. He grants that in the narrow sense it is true that no choice can be more beneficent towards future people than any other. That is, no choice could benefit those who actually do live under it more than any other (since on any other choice they would not exist). But he denies that no choice could be better for people in the wide sense--i.e., better for whoever lives, given our alternatives--than any other. And since that is the case, it is not true that every choice that influences the future fulfills the duty of beneficence.

The Paradox of Future Generations presents an insurmountable obstacle to philosophers who think that morality is concerned only with effects on actual people. But for those who take the wide view of morality it does not seem difficult to find a principle of beneficence that will solve the Paradox of Future Generations. The classical principle of utility is one. Parfit calls it the Impersonal Total Principle of Beneficence. So is the average principle of utility. Parfit calls this the Impersonal Average Principle of Beneficence. And Parfit also discusses two principles of his own, which he calls the Wide Total and Wide Average Person-affecting Principles of Beneficence. The difficulty, according to Parfit, is not that there are no possible principles of beneficence that solve the paradox but rather that all available principles have undesirable implications. If we accept any one of them we must accept it with all its implications; and
this, Parfit suggests, we will not be willing to do. He concludes that we still need to find a satisfactory principle of beneficence.

In the next chapter I will show that the Person-affecting Principles of Beneficence Parfit formulates do not meet the requirements of the problem. Something is wrong with them, quite apart from their implications. I will argue that we must therefore reject them on internal grounds. But the Impersonal Principles of Beneficence do meet the actual requirements of the problem. In my view the Impersonal Principles of Beneficence are satisfactory so far as they go. That is, they satisfactorily respond to the requirements of the problem. And they solve the Paradox of Future Generations, for on these principles it is not the case that every choice that influences the future is maximally beneficent, hence equally morally satisfactory. Thus they can solve cases like Conservation versus Depletion or the Handicapped Child versus the Normal Child. But our troubles will still not be over. For the Impersonal Principles do have the unsatisfactory implications that Parfit astutely attributes to them.

At this point I part ways with Parfit. I think that the satisfactory solution to the Paradox of Future Generations, and through it to other problems of beneficence, is not to be found by devising a new principle of beneficence. The reason for this, I will argue, is that the implications of pursuing maximal beneficence without restraint cannot be solved by any principle of maximal beneficence. Rather, the solution to the implications of unrestrained beneficence is to be found in the restriction
of the chosen principle of beneficence to rational applications. I will consider what it means to apply a principle of beneficence rationally; and I will defend my concept of moral rationality in the area of beneficence against that of other philosophers. I will then show that all of the problematic cases attributable to the unrestrained application of the principle of beneficence can be solved by applying either Impersonal Principle of Beneficence under a model for rational moral decision-making. Finally, I will apply the Impersonal Principles of Beneficence to another problem given by Parfit, the Mere Addition Paradox. I will show that here we need not look to the model for rational moral decision-making. The principles of beneficence by themselves give us a solution to this puzzle, with the Impersonal Average Principle emerging as clearly superior to the Impersonal Total Principle. This result suggests that that principle is sound and should be adopted as the correct interpretation of the duty of beneficence. If I am right, the solution to the Paradox of Future Generations confirms both a sound principle of beneficence and a valid rational method for applying that principle.
II. Possible Principles of Beneficence

Parfit begins his search for the satisfactory general principle of beneficence that will solve the Paradox of Future Generations with four versions of such a principle. All four solve the Paradox of Future Generations. That is, all four say that it is not the case that every choice that influences the future is as good as every alternative choice. But all four have undesirable implications. Therefore Parfit argues that all fail.

What are these possible principles of beneficence?

One is the Impersonal Total Principle of Beneficence. As Parfit formulates it, this principle says that

it is wrong if other things are equal, not to increase the net sum of happiness.

(§F, p. 152)

Here the principle is given its hedonistic form by substituting "happiness" for "beneficence." We might just as well substitute "pleasure" or "preference satisfaction" or "quality of life" or "welfare" without changing the essential import of the principle.

This principle is identical in content with the classical total principle of utility. But a certain difference should be noted. A principle of utility has traditionally been put forward by utilitarians not only as a principle capable of justifying decisions in the area of morality that deals with beneficence but also as the supreme moral principle and arbiter of all other moral principles. We can, however, accept the claim that beneficence should take the form classical utilitarianism says it does without accepting the further claim that the
principle of beneficence is the supreme moral principle. So in that sense it is useful to distinguish the Impersonal Total Principle of Beneficence from the Total Principle of Utility, even if they both say the same thing.

The Impersonal Total Principle of Beneficence, then, says that we should increase the net total amount of happiness in the world. This seems like a laudable goal. What is wrong with the principle? Parfit argues that there is nothing obviously wrong with it except that we reject what it implies. (FG, p. 153) But what is this objectionable implication? Parfit argues forcefully that the Impersonal Total Principle of Beneficence implies what he calls the Repugnant Conclusion.

Parfit asks us to imagine a set of alternative states of affairs, A through Z, that we can bring about as a result of our choice. If we choose A, a population with a very high quality of life or welfare level will be created. If we choose B, twice as many people will live as if we choose A. All these B-people will have lives of worse quality than if we choose A, but all will have a quality of life that is more than half as good as that in A. (We assume equal quality of life among the affected population within each state of affairs.) If we choose C, twice as many C-people as B-people will exist, at a lower quality of life, but again they will have a quality of life that is more than half as good as the quality of life in B. And so on to Z, which has an enormous population whose lives are barely worth living. The result is shown by the following graph.
With each move from A to Z the number of people who live doubles and the quality of life falls but remains better than half as good as in the previous state of affairs. Thus the loss in the average quality of life is more than made up by the gain in the total quantity of worthwhile lives lived, and the total quality of life rises. The Impersonal Total Principle of Beneficence endorses each move down the scale from A to Z. Hence the

**Repugnant Conclusion:**

For any possible and large population, say of eight billion, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, and be what we ought to bring about, even though its members have lives that are barely worth living.

(FG, P. 142)

This conclusion is indeed repugnant. And it is indeed implied by the Impersonal Total Principle of Beneficence. Therefore, its presence suggests that that principle is unsatisfactory, as Parfit says it is.

What are the alternatives?

One alternative principle is the **Impersonal Average Principle of Beneficence**. This says that

it is best, other things being equal, if there is the greatest average net sum of happiness per life lived, or if on average the lives that are lived go as well as possible, or have the highest average quality.

(FG, P. 155)

As Parfit points out, this principle is accepted by many economists. What is wrong with it? Consider a Happy
Child, one whose life will be well worth living and will not affect for either better or worse the lives of other people. The Impersonal Average Principle tells his potential parents to have this child if and only if his existence will increase the average net sum of happiness in the world. Since his existence will have neither a negative nor a positive effect on the happiness of others, it will increase the average net quality of life if and only if he is happier than average. Parfit rejects this principle with a question: "What is wrong with a lower average, if it involves no loss of any kind but only mere addition?" (FG, p. 159)

Parfit's question seems to me insufficient grounds for rejecting the Impersonal Average Principle. Suppose I could choose to conceive the predictably Handicapped Child. His life would be worth living but of less than average quality. Let us further assume that his existence would not affect for better or worse the quality of life of others. That is, any griefs to others caused by his existence would be offset by the pleasures his existence would bring them. Parfit's question implies that on grounds of beneficence there is no objection to conceiving the Handicapped Child. But, if we believe that average beneficence is what matters morally, there is an objection: the existence of the Handicapped Child would lower the average quality of life. Accepting the Impersonal Average Principle of Beneficence would offer us an objection from beneficence to the deliberate conception of the Handicapped Child. This might give us a reason to reject Parfit's stance and accept the average principle.
The problem is that if we accept the Impersonal Average Principle in order to be able to justify certain claims we want to justify we will also find ourselves justifying claims we do not want to justify. For if we accept the average principle we must also accept an implication which Parfit does not make explicit. I will call it the Draconian Conclusion:

If we have a choice between realizing a person or people whose existence will lower the average quality of life (or happiness) in the world and realizing no one at all we should realize no one.

This seems to me to be a forceful objection to the Impersonal Average Principle of Beneficence. We are thus left to choose between the Scylla of the Repugnant Conclusion and the Charybdis of the Draconian Conclusion.

Is there no other alternative? Parfit suggests that we might also consider what he calls the Person-affecting Principles of Beneficence, i.e., those principles of beneficence that accept a Person-affecting Restriction. The Person-affecting Restriction says merely that

this part of morality, the part concerned with human well-being, should be explained entirely in terms of what is good or bad for those whom our acts affect.

(Principles of Biomedical Ethics, p. 149)

This is all the restriction says, but this not all it means. For, as it is stated, this restriction could apply equally well to the Impersonal Principles of Beneficence. In that case it could distinguish no separate principles of beneficence. Nor can it mean that we are to take into
account when making moral judgments only the effects of our decisions on those people who actually live. If it meant this we could never solve the Paradox of Future Generations with a principle of beneficence that used this restriction. For we saw at the beginning that in order to solve the paradox we must compare the effects of alternatives—even though those people who would exist on rejected alternatives will never be realized—in order to say that any alternative is better than any other.

In practice the Person-affecting Restriction tells us to count among the real effects of our alternatives the opportunity costs to already existing people of foregoing their other alternatives. Thus it tells us to include in the net effect of a choice the gain or loss to already actual people relative to their other alternatives. As Parfit actually uses it, the Person-affecting Restriction says that

the part of morality concerned with human well-being should be explained entirely in terms of the difference to actual people between the alternatives that affect them.

We will see below how this restriction works in practice.

First we should set down the three Person-affecting Principles of Beneficence. (FG, pp. 136-138) The first of these is the Narrow Person-affecting Principle of Beneficence. Essentially as Parfit describes it, this principle tells us that

if the choice of P is worse for the P-people (those people who exist if we choose P) than Q is for the P-people, we should choose Q.
But this could equally well be said of a Narrow Impersonal Principle of Beneficence. Therefore I will switch to an interpreted version of the principle and give only interpreted versions of the Person-affecting Principles of Beneficence from this point on. The interpreted Narrow Person-affecting Principle of Beneficence says that

if \( P \) would give the \( P \)-people a smaller net benefit than \( Q \) would give the \( P \)-people, taking into account the effect of alternative choices on presently existing people, we should choose \( Q \).

The problem with this principle where future people are concerned is that if we choose \( Q \) those \( P \)-people who would live in the future never exist: only the \( Q \)-people do. Thus \( Q \) cannot be worse for the \( P \)-people than \( P \) is. This principle might work where all the people affected by a choice already exist but not where some are caused to exist by the choice itself. Therefore, we must reject the Narrow Person-affecting Principle of Beneficence as a solution to the Paradox of Future Generations, hence as a satisfactory general principle of beneficence. Instead we must concentrate on the two Wide Person-affecting Principles of Beneficence.

The Wide Average Person-affecting Principle of Beneficence tells us that

if \( P \) would give the \( P \)-people a smaller average net benefit than \( Q \) would give the \( Q \)-people, taking into account the effect of alternative choices on presently existing people, we should choose \( Q \).
And the Wide Total Person-affecting Principle of Beneficence tells us that

if \( P \) would give the \( P \)-people a smaller total net benefit than \( Q \) would give the \( Q \)-people, taking into account the effect of alternative choices on presently existing people, we should choose \( Q \).

Both principles assume that \( P \) and \( Q \) are alternative choices.

Both Wide Person-affecting Principles of Beneficence, like the two Impersonal Principles of Beneficence, offer potential solutions to the Paradox of Future Generations. Both accord with the notion that the objective of morality in the area of morality that deals with the well-being of future people is the maximization of beneficence; and both employ a wide—as opposed to a narrow—notion of beneficence. That is, both Wide Person-affecting Principles of Beneficence compare the benefits to those people who do exist under a choice with the benefits to those who would otherwise have existed in order to determine the morality of a choice that realizes future people. The trouble is that both imply the Repugnant Conclusion. To see that this is so we must consider in detail how the Wide Person-affecting Principles of Beneficence work.

In general, given alternative choices on a Person-affecting Principle of Beneficence, we must compute the net benefits or gains to individuals who exist under each choice relative to the net benefits or gains to those who exist under the other. If a person affected by a choice exists in the future, he exists on only one alternative. Therefore he foregoes no opportunities. He neither gains
nor loses anything if we choose an alternative on which he does not exist, since non-existent people can be neither benefited nor harmed. The net benefit to a future person of the choice on which he lives is, therefore, the full quality of life he enjoys. But if a person exists on two (or more) alternatives, we must calculate the effects of each choice on him relative to the other choices we can make. This restriction applies to any and all people who exist no matter what we choose. Therefore, given the kind of choices with which we are concerned, it applies to all people who are presently alive and only those people. If a presently existing person loses any of his quality of life on a given alternative we subtract the quality of life he enjoys on the more beneficent alternative from that he ends up enjoying to determine his net loss or opportunity cost relative to the alternative at hand. This in turn is subtracted from the net gain anyone else enjoys on that alternative to determine the net benefit of the choice. If a presently existing person gains on a choice relative to an alternative we count his benefit as the gain over his outcome on the alternative. We then add that gain to the gains of all other people who did not previously exist to determine the net gain or net benefit of that alternative. The objective is to maximize net benefits, i.e., the gains of those who exist on a choice over those who exist on an alternative choice when opportunity costs to those who already existed are subtracted. In this way our principles are "person-affecting." They count the effects on persons of alternatives that would affect them.
Suppose we could make either of two alternative choices, (1) or (2). We wish to maximize total net benefits to those who exist under a Wide Total Person-affecting Principle of Beneficence. We can first use the figures Parfit supplies on an "Adam and Eve scenario." (Cf. FG, fn.36, pp. 155-56.) We may call this the Eden scenario. It may help to visualize the situation if we refer to the following graphs. The first graph represents what happens if we choose (1). Eve already exists. Adam does not.

Eden (1)

\[
\begin{array}{c}
\text{Eve} \\
49 \\
51 \\
\end{array}
\text{V} \\
100
\]

Total net gain, or benefit, on Eden (1) vis à vis Eden (2):

\[
49
\]

If we choose (1), Eve, who already exists, will continue to exist at a quality of life of 100 utility points, or utils, and no one else will exist. Eve's total
benefit is 100. What is her net benefit? In order to know this we must subtract from her total benefit on (1) the total benefit she would have had had we chosen (2). Now, if we had chosen (2), Eve would have continued to exist, but her quality of life would have dropped to 51. Adam would also have existed on (2) at a quality of life of 51, but since he does not exist on (1) he does not affect the calculations if we choose (1). Therefore, if we wish to know Eve's net benefit on (1), we must subtract 51, Eve's total benefit on (2), from 100, her total benefit on (1). Eve's net benefit on (1) is 49. And as she is the only person who exists on (1), her net benefit is the total net benefit to all the people who exist if we choose (1) over (2).

Now let us see what happens if we choose (2). Again, Eve already exists and Adam does not.

Eden (2)

Eve's net loss on Eden (2) vis à vis Eden (1): -49
Adam's net gain on Eden(2): 51
Total net gain, or benefit, 2 on Eden (2) vis à vis Eden (1)
If we choose (2), we must subtract Eve's total benefit on (1) from her total benefit on (2), again in order to determine the opportunity costs of realizing Eve on (2). This gives us a negative number. Eve loses 49 (-110+51). But Adam, who exists only on (2), gains 51 over (1) because he does not exist on (1) and consequently has no opportunity costs. His total gain on (2) is also his net gain on (2). Adam's net gain plus Eve's net loss gives us a total net gain or total net benefit on (2) of 2 (51-49).

If we wish to maximize total net benefits to everyone who lives we should choose (1): those who live, namely Eve, have a total net benefit of 49, whereas those who live on (2), namely Adam and Eve, have a total net benefit of only 2.

But if we are directed by the Wide Total Person-affecting Principle to keep just Eve rather than realizing Adam as well, it seems that we should not reach the Repugnant Conclusion using this principle. For it has just said that, if we add extra happy people to those who already exist, the cost to those who already exist may outweigh the gain to the newcomers, and in that case there will be more total beneficence by having fewer people.

Let us see if this result holds when we use different figures. We can use the figures from Parfit's explanation of the Repugnant Conclusion, the RC scenario, substituting the names of "Eve" for A and "Adam and Eve" for B in order to make the parallel with the scenario I have just given. In effect, using the Wide Total Person-affecting Principle of Beneficence, we are trying to justify generating B out of A, where population A presently exists and all the
additional people are future people. For convenience we conceptualize all the presently existing people in A as one person (Eve) and all the future people in the next generation to which A gives birth as one person (Adam).

On the RC scenario we already have Eve with a total benefit of 100 utility points as before. We can leave Eve alone. Call this choice (1):

\[ \text{RC (1)} \]

\[
\begin{array}{c|c|c}
20 & 0 & 100 \\
80 & 0 & 100 \\
\end{array}
\]

Total net gain, or benefit, on RC (1) vis à vis RC (2): 20

Or we can create Adam in addition to Eve. Call this choice (2).

\[ \text{RC (2)} \]

\[
\begin{array}{c|c|c|c}
20 & 0 & 100 & 80 \\
80 & 0 & 100 & 80 \\
\end{array}
\]

Eve's net loss on RC (2) vis à vis RC (1): -20
Adam's net gain, or benefit, on RC (2): 80
Total net gain, or benefit, on RC (2) vis à vis RC (1): 60
If we choose (2), Eve's total benefit will drop to 80 but Adam will gain 80. Thus choosing (2) costs Eve 20 (100-80) but causes Adam to gain 80. Adam pays no opportunity costs since he foregoes no opportunity on (1), where he does not exist. We subtract Eve's 20 loss from Adam's 80 gain for a total net gain to all the people who live in (2) of 60. But if we don't create Adam, i.e., if we choose (1), Eve will continue at a total benefit of 100. This is a net benefit, however, of only 20 when we subtract what she would have had on (2). Adam does not exist and so does not affect the calculations.

The total gain to everyone who lives on (1) minus opportunity costs equals a total net gain of 20. But the total gain to everyone who lives on (2) minus opportunity costs equals a total net gain of 60. We should choose (2), which shows the greater total net gain. But in this way we reach the Repugnant Conclusion.

But if this is the case the Wide Person-affecting Principle gives inconsistent results. There is something wrong with it. Can it be amended? Parfit acknowledges that both the Wide Person-affecting Principles of Beneficence can lead to inconsistent judgments, but he claims that "we could avoid this problem by giving the principles a more complicated form." (FG, fn.36, p.156) He predicts that the principles would still endorse the Repugnant Conclusion. We want to see if he is right.

What is the difference between the original Adam and Eve scenario and the RC scenario? On both, Adam's gain is greater than Eve's loss (51 to 49 v. 80 to 20). But on the Adam and Eve scenario there is a very little profit in creating Adam (2) and a very great loss to Eve (49). On
the RC scenario, there is a great gain in creating Adam (60) while there is only a minor loss to Eve (20). Therefore, we might try formulating a Revised Wide Total Person-affecting Principle of Beneficence which says that

where P is a choice on which present people continue to exist alone and Q is an alternative on which future people are added, and the difference Q makes to the P-people is greater than the net total gain Q gives to the Q-people, we should choose P.

This would work on the scenario because the difference Q makes to the P-people (Eve) is 49 and the net gain to the Q-people (Adam and Eve) is only 2; so we should choose (1). But the new principle will not work on the RC scenario, for here the difference to the P-people (Eve) is only 20 and the net gain to the Q-people (Adam and Eve) is 60. Thus the net gain to the Q-people is greater than the difference to the P-people. Therefore, we cannot conform the facts to the principle. And if we extend the principle to say,

but where the total net gain to the Q-people is greater than the difference to the P-people we should choose Q,

we are led to the Repugnant Conclusion. Parfit was right. A consistent Revised Wide Total Person-affecting Principle leads to the Repugnant Conclusion.

What about the Wide Average Person-affecting Principle of Beneficence? Does it give the same inconsistent results as the total principle? Parfit works out the calculations in terms of the RC scenario (FG, fn.30, p.146). Again I shall use "Eve" to stand for population A
and "Adam and Eve" to stand for population B. We may use the same graphs as before, RC(1) and (2); only this time we want to calculate the average rather than the total benefits of our alternatives.

Here again Eve already exists alone in A with a total benefit or quality of life of 100. On choice (1) we allow Eve to continue living alone and on choice (2) we create Adam in addition to Eve. If we create Adam he again gains 80 utility points, or utils and Eve loses 20, so that each ends up with 80 utils. The difference to Eve of either choice is 20. If we choose (2) she loses 20 vis à vis (1); and if we choose (1) she gains 20 vis à vis (2). Therefore on (1) the total net gain to everyone who lives is 20. And the average net gain is the same thing because there is only one person, Eve. (2) has twice as many people as (1). And these people have a total net benefit or net gain of 60 (i.e., a net gain of 80 to Adam and a net loss to Eve of 20 for a total net gain of 60). We divide the gain of 60 by 2 (Adam and Eve), and the average net gain turns out to be 30. 30 is greater than the average net gain of 20 on (1). Therefore under the Wide Average Person-affecting Principle we should choose (2), and we are led to the Repugnant Conclusion.

How does the Wide Average Person-affecting Principle work out on the Eden scenario? Again, referring to our previous graphs, Eden (1) and (2), Eve already exists at 100. If we choose (1), the gain to Eve over what she would have had on (2), namely 51, is 49. Therefore, the average net gain of all the people who live on (1), as well as the total net gain, is 49, for there is only one person. But if we choose (2), the total gain is only 2,
i.e., Adam's 51 gain minus Eve's 49 loss. Therefore, the average net gain is 2 divided by 2, or 1. We should choose (1) for an average net gain of 49 over (2) for an average net gain of 1. We avoid the Repugnant Conclusion, but our results are inconsistent with those of the RC scenario under the same principle—exactly the same situation as with the Wide Total Person-affecting Principle. And just as with that principle, we can derive consistent results from a Revised Wide Average Person-affecting Principle of Beneficence, but we end up endorsing the Repugnant Conclusion.

It seems, then, that both the Wide Person-affecting Principles of Beneficence either give inconsistent results or, if reformulated to give consistent results, direct us to the Repugnant Conclusion. Therefore, as Parfit says, they have unsatisfactory implications. Nevertheless, before abandoning them, we remind ourselves what it is that the Wide Person-affecting Principles of Beneficence are actually measuring.

On the Wide Person-affecting Principles of Beneficence the real quality of life that already actual people enjoy as a result of our choice is not the measure of the value of the choice to them. What does matter is their theoretical gain or loss relative to other choices we might have made. The measure of utility for already existing people is thus the real quality of life they end up enjoying minus theoretical opportunity costs. It is therefore, not their real benefit enjoyed or outcome in terms of beneficence. But the measure of utility for people who exist in the future is their real gain or benefit enjoyed. This seems inconsistent. But if we
figure in the opportunity costs to future people at 0, since they have no other alternatives, the standard of measurement is consistent but inelegant.

Every principle of beneficence we can propose will correspond to a different way of calculating benefits and harms. We will want to see what calculations the Impersonal Principles of Beneficence use. But before we turn to them, we can consider yet another way to calculate beneficence.

We might want to say that the benefit of a choice to an already existing person is the difference that choice makes in the quality of life he ends up enjoying over that he previously enjoyed. And the cost of a choice to an already existing person is the quality of life or number of utils he loses \textit{vis à vis} his previous position. As on both the Wide Person-affecting Principle of Beneficence and the Impersonal Principles, the benefit of a choice to future people is the full quality of life they enjoy, since otherwise they would not exist.

If Eve, for example, already exists and enjoys a quality of life of 100 utils and I choose Eden (1) over Eden (2) Eve's quality of life does not change at all. She remains at 100. Therefore, the net benefit, or difference in quality of life for everyone who ever lives, caused by my choosing (1) is 0. There is no difference in Eve's status before and after my choosing (1). If I choose Eden (2), on the other hand, Adam will gain 51 and Eve will lose 49. The real difference, or net benefit, they both end up enjoying is 2: Adam gains 51 and Eve loses 49. This way of calculating benefits is represented on the following graph.
And surely we want to say that I can benefit from a passive choice which leaves me alone rather than taking away something I already had. The benefit to me may be passive, but it is real nonetheless.

The trouble is more than a quibble over point of view. Notice that with the new method of calculating benefits, anyone who is already enjoying a benefit does not get to count that benefit for the purpose of calculating the effects. Thus he will usually benefit very little from choices, since only specific changes in his quality of life caused by those choices will count. But someone who is not already enjoying a benefit will stand to gain enormously from choices which benefit him, since he literally has nothing to lose and everything to gain. This should alert us that something has gone wrong.

The problem is this. It applies equally as well to the Wide Person-affecting Principles of Beneficence as to the new method. That is, it applies to any method of determining the benefits and losses of choices that counts for presently actual people only the difference the present choice makes to them and not the outcome they enjoy.

Regardless of how we calculate benefits and costs, both present and future people really end up enjoying not only the specific benefit of the single choice under consideration but also the benefit of all other choices which contribute to the total outcome. But on the new method of calculating benefits presently existing people count as gaining only the real benefit of the choice in question (or as suffering only the real loss of that
choice); and on the Person-affecting Principle of Beneficence, presently existing people take only the benefit or loss of the present alternatives calculated relative to each other. Neither counts the real total outcome (or real average outcome) as the real gain or loss of a choice for presently existing people. Yet the Full Comparative View, which we are using, requires that we measure the benefit of a choice that is an essential cause, or part of the cause, of a person's enjoying a certain quality of life as the real outcome or real benefit he ends up enjoying once the choice is made. (Cf. pp.3-4) This should hold whether the person affected by the choice lives in the present or in the future.

Calculating the benefits to already existing people as the difference between where they were before and where they end up, rather than as their real outcome in terms of quality of life, does not meet the criterion of the Full Comparative View. Nor does counting their benefits as the surplus benefits they enjoy relative to their other opportunities. These methods of benefit calculation only appear to meet that criterion because of an ambiguity in the expression of the Full Comparative View. The Full Comparative View requires that the value of a choice be the full difference a choice makes; but this necessarily means the full difference as measured by the outcome. Here we can quote Parfit:

In deciding which act would benefit people most, we should use the Full Comparative View. Suppose that I can do either P or Q. In deciding which would benefit people more, we should compare all of the benefits and losses that people would later receive if and only if
I do P rather than Q, and all of the benefits and losses that people would receive if and only if I do Q rather than P. The act which benefits people more is the one which, in this comparison, would be followed by the greater net sum of benefits—that is, the greater sum of benefits minus losses. It is irrelevant that many other acts will also be necessary causal antecedents for the receiving of these benefits.

(FG, p. 125; italics mine.)

The new method of calculating the full difference a choice makes to presently existing people and the Person-affecting Principles of Beneficence fail to meet the criterion of the Full Comparative View in just the same way: they do not count an already existing person's benefit from a choice as the full quality of life he ends up enjoying and which otherwise would have been different. But we agreed at the start to count the value of a choice as the full outcome in terms of beneficence of that choice (plus all other factors required to reach the result).

True, we agreed to accept the Full Comparative View in conjunction with the claim that causing to exist can benefit. But there are strong reasons for applying it to the effects of our choices on presently existing people as well as to the effects of our choices on future people. (1) Parfit himself justifies the Full Comparative View by examples of beneficent outcomes involving only existing people. In each case the argument is that the value of choice is the full outcome, not its Share of the Total outcome vis-à-vis the previous position (Cf. FG, pp.124-25) (2) In the absence of some reason for not employing the same standard with respect to present people as we employ with respect to future people and (3) in the.
absence of some extra benefit from being caused to exist, over and above the benefit in terms of quality of life—an extra benefit which might justify our taking a different approach to future people than to present people—we should, to be consistent, rely on the Full Comparative View in both cases. And if there is any such extra benefit of existence to future people it is not obvious and has not been argued for. Therefore we will disregard it. But if we do rely on the Full Comparative View, we must reject both the Person-affecting Principles of Beneficence and the new method of calculating benefits, for neither accords with it.

We are left with only one alternative, the Impersonal Principles of Beneficence. The benefit of a choice on the Impersonal Principles of Beneficence is the full difference that choice makes as measured by the full result—the outcome in terms of benefits—for everyone.

The value of our choices on the Eden scenario, using this method of calculation, is shown on the following graph.

Eden (1)
Real total net benefit: 100

Real average net benefit: 100
Eden (2)

Real net benefit to Eve: 51
Real net benefit to Adam: 51
Real total net benefit: 102
Real average net benefit: 51

Eve's total benefit from our choosing Eden (1) is the quality of life she enjoys as a result of that choice, namely, the quality of life she already had, 100 beneficence points. The average quality of life of everyone affected by the choice is also 100. The outcome of a passive choice counts just as much as the outcome of an active choice. And if a (passive or active) choice is an essential part of the cause of someone's ending up with a certain quality of life, he gets the full benefit of that choice, where "benefit" is understood as "quality of life" or "difference over non-existence." Thus a presently actual person enjoys the full difference of a choice in the same sense as a future actual person.

If we choose Eden (2), Eve, who already exists, loses 49, just as we recognized on the Wide Person-affecting Principles of Beneficence. We do not ignore this cost,
which is not only the opportunity cost to her of choosing (2) but the real cost. Her real benefit, however, in terms of quality of life, or outcome, or gain over non-existence, is 51, the same as Adam's. Together Adam and Eve enjoy a real total quality of life of 102 and a real average quality of life of 51.

We then compare (1) and (2). The real total benefit enjoyed is greater on (2), since (2) more than makes up in quantity of beneficence the loss in quality. But the real average benefit is greater on (1).

The value of our choices on the RC scenario using the Impersonal Principles of Beneficence, is as follows.

RC (1)

\[
\begin{array}{c}
\text{Eve} \\
\checkmark
\end{array}
\]

\[
\begin{array}{c}
100 \\
\text{Real total net benefit: 100} \\
\text{Real average net benefit: 100}
\end{array}
\]

RC (2)

\[
\begin{array}{c}
\text{Eve} \\
20 \\
\checkmark 80
\end{array}
\]

\[
\begin{array}{c}
\text{Adam} \\
\checkmark 80
\end{array}
\]

\[
\begin{array}{c}
\text{Real net benefit to Eve: 80} \\
\text{Real net benefit to Adam: 80} \\
\text{Real total net benefit: 160} \\
\text{Real average net benefit: 80}
\end{array}
\]

Again the real average benefit is greater on (1) and the total benefit on (2). Quantity more than compensates for the loss in quality of life in terms of total beneficence.
But the greater number of people decreases average benefi-
cence. The results of the RC scenario are straightforwardly consistent with those of the Eden scenario.

We can now fully express the requirements of the Impersonal Principles of Beneficence as the maximization of real utility (benefits, quality of life, or welfare).

The fully explicated **Impersonal Total Principle of Beneficence** says that

given alternative choices P and Q, if the total quality of life of everyone who lives on P is greater for the P-people than the total quality of life on Q is for the Q-people, taking into account the effect of each choice on presently existing people, we should choose P.

And the fully explicated **Impersonal Average Principle of Beneficence** says that

given alternative choices P and Q, if the average quality of life of everyone who lives on P is greater for the P-people than the average quality of life on Q is for the Q-people, taking into account the effect of each choice on presently existing people, we should choose P.

Here we should notice that where no one who will be affected by a choice currently exists there would be no difference between the requirements of the Impersonal Principles of Beneficence and the corresponding Wide Person-affecting Principles. The difference to everyone who ever exists under a choice which takes effect only in the future is the full quality of life he has, starting from 0. Therefore, in such cases the Person-affecting Principles and the Impersonal Principles give the same directive: if the P-people would benefit more from P than the Q-people would benefit from Q, we should choose P.
The difference arises where some of the people affected by a choice presently exist. Here the Person-affecting Principles of Beneficence calculate benefits in terms of opportunity costs and the Impersonal Principles in terms of outcomes. But this difference is enough to cause us to reject the Person-affecting Principles in light of the Full Comparative View and to endorse the Impersonal Principles. Only the latter correspond to the givens of the problem.

Notice that the fully explicated Impersonal Principles of Beneficence are complex principles. The total principle does not tell us to add every life worth living that we possibly can to the world. It tells us to add extra happy people if and only if doing so will not decrease the total happiness in the world. Admittedly the possibility of decreasing the total happiness in the world by adding happy people is far-fetched, but it is theoretically possible. For example, a desert sheikh might choose to have 100 happy sons whose extravagant lifestyle would be so burdensome to his subjects that the total happiness of the affected population would be decreased by the existence of the princes. The Impersonal Total Principle of Beneficence would not justify the sheikh's choice. Similarly, the Impersonal Average Principle of Beneficence does not tell us to add only those extra lives that will be of more than average quality. It tells us to add extra lives if and only if they will not reduce the average quality of life of all. Whether or not they do this depends on the effect of their existence on others as well as the quality of their own lives. This is a very important point. I will return to it in Chapter VI.
The fully explicated Impersonal Principles of Beneficence correspond to the dictates of the problem before us in a way the Person-affecting Principles of Beneficence do not. Therefore they seem to remain as candidates for a satisfactory general principle of beneficence when we reject the Person-affecting Principles. The trouble is that fully expressing the Impersonal Principles of Beneficence has no effect on their implications. The Impersonal Total Principle still implies the Repugnant Conclusion, and the Impersonal Average Principle still implies the Draconian Conclusion.

Recall that the explicated Impersonal Total Principle tells us that if extra people would add more utility to the world than presently existing people lose by their addition, we should realize those extra people. Thus if an already existing Eve loses 50 utility points by the creation of Adam and Adam gains 50, the Impersonal Total Principle is indifferent as to whether we should keep Eve alone or realize Adam in addition to Eve. But if Eve loses only 49 points and Adam gains 50, we should add Adam: this is more beneficent than merely keeping Eve alone. In reality future people would almost always gain more from existence than previously existing people would lose by their addition. So the Impersonal Total Principle would almost always tell us to add extra people. This is the case on both the Eden scenario and the RC scenario. The Impersonal Total Principle of Beneficence thus leads to the Repugnant Conclusion.

The explicated Impersonal Average Principle of Beneficence, on the other hand, tells us not to add extra lives if doing so will reduce the average quality of life
or average welfare. On the Eden scenario we can choose either Eden (1), to leave Eve alone at an average quality of life of 100, or Eden (2), to create Adam in addition to Eve, with an average quality of life of 51. And on the RC scenario we can choose either RC (1), to leave Eve alone at an average quality of life of 100, or RC (2), to create Adam in addition to Eve at an average quality of life of 80. In both cases the Impersonal Principle of Beneficence advises us to leave Eve alone even though this means forbidding the generation of future people. The Impersonal Average Principle thus implies the Draconian Conclusion.

What are we to do? Since the Impersonal Average Principle of Beneficence implies the Draconian Conclusion it seems we should reject that principle. But the Impersonal Total Principle of Beneficence implies the Repugnant Conclusion. Thus it seems we should reject that principle also. But there are no other obvious alternative principles of maximal beneficence possible, without going totally outside the realm of considerations with which we have been dealing and seeking, say, a principle that will maximize median beneficence, i.e., the number of people enjoying the middle position on a continuum of beneficence, or a Rawlsian maximin principle, i.e., one that maximizes welfare only so long as the worst off benefit. Before we take so drastic a step we should consider whether the principles we have been working with seem worth preserving. Certainly both the Impersonal Total Principle and the Impersonal Average Principle have enjoyed great acceptance historically in their guise as principles of utility, both by moral philosophers and by
economists. It seems rash to reject them altogether. We should first attempt to solve our problems, either by modifying the essential principles without rejecting them (Parfit's approach) or by suggesting a method for the evaluation of choices under the Impersonal Principles that will allow only intuitively good choices to be made (our own preference).
III. The Strengths and Weaknesses of the Impersonal Principles of Beneficence

Let us now assess our position.

Parfit says that we must reject both the Wide Person-affecting Principles of Beneficence because they are inconsistent in their directives and when made consistent they imply the Repugnant Conclusion. We agreed with this, but we have also shown that these principles define the net benefits of our choices as people's gains on those choices over their starting points, when the difference to presently existing people of their unrealized alternatives is figured in. But net maximal beneficence can be interpreted in other ways as well, notably as requiring the maximization of benefits, where benefits are defined as real outcomes in terms of beneficence or quality of life. Indeed, this way of calculating benefits is a requirement of employing the Full Comparative View, as we agreed at the beginning to do. We saw that the Impersonal Principles of Beneficence were consistent with the Full Comparative View where the Person-affecting Principles were not. Moreover, these principles seem to be the only workable general principles of beneficence there are that accept the maximization of real beneficence as the goal of morality towards future people. The Impersonal Total Principle of Beneficence implies, however, the Repugnant Conclusion, just as Parfit says it does; and the Impersonal Average Principle of Beneficence implies the Draconian Conclusion. Therefore, it appears that these principles too are unsatisfactory.
At this point it would be good to consider whether the Impersonal Principles of Beneficence have any virtues other than avoiding the difficulties of the Person-affecting Principles. And it appears that both new principles have considerable strength.

Both Impersonal Principles of Beneficence require us to choose that available alternative whose over-all outcome is most beneficent in the wide sense. Therefore both principles solve the Paradox of Future Generations, and both imply and explain Claim (A),\(^6\) as Parfit requires of a satisfactory principle of beneficence. This is the claim that,

If the same number of lives would be lived either way, it would be intrinsically worse if those who live are worse off than those who would have lived.

\((\text{FG, p.122})\)

Claim (A), that is, follows from either Impersonal Principle of Beneficence plus the claim that the objective of morality in the area of morality which deals with the well-being of future people is to maximize beneficence (in the wide sense) through our choices. (Cf. Chapter I.). For if the objective of morality is to maximize beneficence, then any choice which satisfies that objective is intrinsically better than any alternative choice which does not. And, by the same token, any choice which does not satisfy that objective is intrinsically worse than any alternative which does. Therefore, if our moral objective is to maximize beneficence and if the principle of beneficence tells us that we should always choose that available alternative which is maximally beneficent, then any less beneficent alternative is intrinsically worse on that
principle, i.e., worse with respect to the objective of morality in the area of beneficence. Moreover, it would appear that the Impersonal Principles of Beneficence are necessary and sufficient principles of moral beneficence. For if the objective of morality in this area is to maximize beneficence and if any choice that maximizes beneficence is intrinsically better than one that does not, and the Impersonal Principles of Beneficence prescribe maximal beneficence, as they do, then by employing either Impersonal Principle of Beneficence we can ensure that the objective of morality in this area is met.

As we would expect of necessary and sufficient principles of beneficence, both Impersonal Principles of Beneficence in most cases justify the intuitively moral choice. They do this in precisely those cases that are problematic for a narrow view of beneficence but can be resolved by taking a wide view. That is, they are effective in precisely those cases that are problematic with respect to the Paradox of Future Generations but are solved by any principle of beneficence that solves the paradox.

Consider, for example, Parfit's Depletion case. (Cf. p. 5.) Recall that in the Depletion case we have a choice between following a policy of Depletion (D) and following one of Conservation (C). If we choose D, the quality of life over the next 200 years will be slightly higher than if we choose C, but then it will fall dramatically below the level that C would give. Life, however, will still remain well worth living; indeed, it will be no worse than life today. Therefore, D benefits the people who live under it, since they have lives worth living and
they would not have lived had D not been chosen. Moreover, no one who would have lived had C been chosen instead exists to be harmed by the choice of D. Thus if we choose D, everyone who ever lives benefits, and there is no one for whom the decision is worse. D is maximally beneficent in the narrow sense. That is, it is as beneficent as possible towards everyone who ever actually lives. But if we choose C, everyone who lives benefits more on the average and in sum than if we had chosen D.

No traditional narrow principle of beneficence can direct us to choose C over D. But under the Impersonal Total and Average Principles of Beneficence we should choose C, because those who live, the C-people, will be better off than those who would have lived had we chosen D. Thus in this case both Impersonal Principles of Beneficence justify the intuitively moral choice.

This is not the only kind of case the Impersonal Principles of Beneficence solve. They also tell us why we should make the intuitively satisfactory choice in Parfit's cases of the 14-Year-Old Girl (FG, p.118) and of the Handicapped Child (FG, p. 118). In the 14-Year-Old Girl case, a girl has the chance of having a baby now or waiting and having one later. Both Impersonal Principles of Beneficence tell her that if the child born later would predictably have a better quality of life than the child born now she should wait. In the Handicapped Child case, as we saw, a mother has the opportunity of having a child now who will predictably be born handicapped but with a life worth living, or waiting and having a normal child later. Our principles of beneficence tell her she should wait, since the child born later will predictably have a
better quality of life. In both cases the intuitively satisfactory choice is dictated.  

Nevertheless, both Impersonal Principles of Beneficence face an array of problems, including, but going beyond, the Repugnant Conclusion and the Draconian Conclusion. In the remainder of this chapter I will present those problems, and in the following chapters I will suggest a comprehensive solution to them.

i. Different Number Choices

Different Number Choices--choices on which different numbers of people live according to which alternative we choose--are the ones that lead to the Repugnant Conclusion (on the Impersonal Total Principle of Beneficence) and the Draconian Conclusion (on the Impersonal Average Principle of Beneficence). Thus they present a problem for both Impersonal Principles. We might be tempted to argue, however, that these unsatisfactory conclusions will never be reached in real life; therefore they present no substantial objection to the Impersonal Principles of Beneficence. This argument would be mistaken. Many people seem to believe today that having children would interfere with their own lifestyle and that fulfilling their own desires is preferable to parenthood. What has been called the "New Narcissism" can be justified by the Impersonal Average Principle of Beneficence. Indeed, Narcissism emerges as morally better than parenthood. (I will consider the implications of this fact below, pp. 119ff.) On the other hand, a policy of unrestrained human reproduction may well ultimately result in a population like Z, where an alternative policy of birth control eventually might result in a population like A. Mexico today appears to be approaching Z. A major reason that population
policies are undertaken is to prevent societies' reaching Z. But the Impersonal Total Principle of Beneficence approves of reaching Z. Moreover, many of our other choices tend to confirm the implications of the Impersonal Principles of Beneficence. To see this we need only consider some of our original cases as Different Number Choices.

We saw that either the Impersonal Total Principle or the Impersonal Average Principle of Beneficence will allow us to solve the Handicapped Child case or the 14-Year-Old Girl case, as they are given. But in each case there is actually a further option: the potential mother can choose to have no child at all. Both these cases are thus, at least potentially, Different Number Choices, as opposed to Same Number Choices, on which the same number of people will live whatever we choose.

Suppose the choices are between a woman's having the Handicapped Child now or no child at all and between the 14-Year old Girl's having a child now or never. The Impersonal Total Principle of Beneficence tells us that the potential mother of the Handicapped Child should have him and that the 14-Year Old Girl should also have a child now. This is the case because any increase in numbers of people with lives worth living increases the total beneficence in the world just so long as those who already existed do not lose more from the presence of the newcomers than the newcomers gain. The newcomers will almost always gain more than those already alive lose by their creation, and we can assume that this is the case here. Therefore, if the potential mother of the Handicapped Child could only have the Handicapped Child she should have him, and so should every similarly situated mother.
And every 14-year-old girl should have as good a child as possible under her options even if his life will be pretty miserable, so long as it is worth living and contributes more beneficence to the world than it subtracts.

It may seem, therefore, that Different Number Choices give us a reason to prefer the Impersonal Average Principle of Beneficence over the total one. But this is not the case, for our average principle implies the Draconian Conclusion. Whenever people can be caused to exist who will maintain or increase the average quality of life, the average principle says that they should be caused to exist. And whenever people can be caused to exist who will decrease the average quality of life, the average principle says they should not be caused to exist. Suppose that not even the better-off child on the Handicapped Child case or the 14-Year-Old Girl case will have a life of average quality. And bringing either into the world will not significantly increase or decrease the happiness of others. Bringing either into the world will therefore decrease the average quality of life lived. The average principle then must tell the potential mother not to have this child. This seems indeed Draconian.

Nor is this the only problem Different Number Choices present for the Impersonal Average Principle of Beneficence. Suppose a mother has a choice of having a Happy Child or no child at all. The Happy Child will have a life of above average quality. The Impersonal Average Principle says the mother should have this child. But if we endorse a mother's having a Happy Child in every case, we may create a prisoners' dilemma. That is, it may increase the happiness in the world if I have a happy
child. But if everyone who can have a Happy Child does so, we can envision the resulting strain on resources' leading to a reduction in the average quality of life of everyone who lives. ⁸ (For the Happy Child case cf. FG, p.138. Parfit does not mention the problem of the prisoners' dilemma, but it is a Parfitian notion.) Thus Different Number Choices create a dilemma for the average principle just as they do for the total principle.

ii. The Exploitation Policy

There are still other problems with the principles of beneficence. Neither the Impersonal Total Principle nor the Impersonal Average Principle can tell us which of two alternatives to choose where the total or average beneficence under each is the same but is achieved in widely different ways.

For example, let us assume that we have two alternatives. On each alternative there will be two affected populations, each with the same number of people, a present population A, viz., ourselves, and a future population B, and a finite quantity of some available resource, say coal. Our alternatives are to adopt either an Exploitation Policy (P) or an Equalization Policy (Q). If we adopt the Exploitation Policy, we in A will use up this resource as fast as possible, leaving only what we cannot consume for future people. If we adopt the Equalization Policy, we will share the use of this resource equally with future people. Say that on P the Aₚ-peoples have a total benefit of 140 utils and the Bₚ-peoples have a benefit of only 60, for a total net benefit of 200 utils and an average net benefit of 100. On Q the Aₗ-peoples and the Bₗ-peoples both have benefits of 100 utils, for a total
benefit of 200 and an average benefit of 100. The size of each affected population is the same. Neither the average nor the total principle of beneficence can distinguish morally between P and Q. But either is supposed to be able to serve as a necessary and sufficient principle of beneficence. This implies that it is morally unobjectionable to indulge ourselves at the expense of the future, so long as total and/or average beneficence remain the same.⁹ But this conclusion seems just as objectionable as either the Repugnant Conclusion or the Draconian Conclusion. The case is my own.

iii. The Risky Policy

The Impersonal Principles of Beneficence also cannot solve the problem posed for morality by cases like Parfit's Risky Policy, where maximal beneficence is assured by an intuitively immoral policy. (FG, pp.114ff.) On the Risky Policy, an energy policy, the quality of life rises for two hundred years and then levels off. Moreover, it rises higher than it would have risen had an alternative policy, the Safe Policy, been chosen. But in some generation after two hundred years thousands of people are killed and injured by radiation leakage as a predictable effect of the Risky Policy. The Safe Policy has no such bad effect. According to both principles of beneficence we should choose the Risky Policy. The people who live, even the dead and injured, are better off than if they had never existed, since they have lives worth living. And those who live are better off in sum and on the average than those who would otherwise have lived. The Risky Policy is therefore maximally beneficent in accordance with the terms of either principle of beneficence. How can it not
be the better alternative? This is a great problem for beneficence as a sufficient moral theory.

Thus our two Impersonal Principles of Beneficence have these problems: (1) on Different Number Choices the Impersonal Total Principle implies the Repugnant Conclusion and the Impersonal Average Principle implies the Draconian Conclusion; and with respect to happy children the average principle implies a significant prisoners' dilemma; (2) neither principle can handle the Exploitation Policy; and (3) neither principle can solve the problem posed for morality by the Risky Policy. If either of our two principles of beneficence is to serve as the principle of choice in the area of morality that deals with well-being, either it must be modified in a way that will solve all these problems, or restrictions must be placed on its application. Otherwise the principles will have been defeated by their implications.

From this point on, Parfit's and my paths to a solution to the problems of beneficence diverge. I will first pursue my own path through to the end, and I will then argue that the path Parfit takes leads to a dead end.
IV. A Solution to the Problems: A Rational Approach to Moral Decision-making

What is "beneficence"? We have so far not settled upon an interpretation of this term, although we have at times substituted for it the terms "well-being," "welfare," "quality of life," and even "happiness" and "utility." But it seems clear that a certain interpretation of "beneficence" is implicit both in the Impersonal Principles of Beneficence and in the cases we have brought to bear against them. On both sides there is an implicit assumption that beneficence equals the satisfaction of actual preferences, i.e., of those personal preferences that people actually have. Thus the directive to maximize beneficence to whoever is affected under either Impersonal Principle of Beneficence is a command to maximize the satisfaction of the actual personal preferences of whoever is affected by our choice vis a vis the actual personal preference satisfaction of those who would have been affected had we chosen differently. From this point on I will give the Impersonal Principles of Beneficence this preference utilitarian interpretation.

The presumed objective of morality in the area of morality that deals with the well-being of present and future people, then, is to maximize the satisfaction of the actual personal preferences of those who are affected in the wide sense; and the Impersonal Principles of Beneficence fulfill this objective. Therefore they should offer a necessary and sufficient condition for the goodness of choices in the area of morality that deals with well-being. In light of the problems of the Impersonal
Principles of Beneficence, however, it appears that
neither Impersonal Principle as just interpreted can serve
as a sufficient principle of morality where well-being is
concerned. For the Risky Policy, the Exploitation Policy,
and the problematic Different Number Choices all meet this
criterion for the goodness of choices. Yet we still want
to say that there is something wrong with these choices.
Thus it must be that maximal actual personal preference
satisfaction is an insufficient notion of morality in this
area after all. Either the principle of beneficence is a
bad principle or it requires restriction by something
outside itself. And since the claim that maximal actual
personal preference satisfaction is the object of moral
welfare is itself the central claim of preference utili-
tarianism, if this claim fails in this area of morality so
must preference utilitarianism fail as a moral theory, at
least in a pure or unrestricted form.

We can confirm this conclusion by noting that what-
ever we appeal to to correct the shortcomings of the
Impersonal Principles of Beneficence, we cannot appeal to
more actual personal preference satisfaction. Consider
the Risky Policy. No matter how bad we make the nuclear
catastrophe on that policy we can also make the Risky
Policy more beneficent than the alternative. Thus,
whatever our objection is to the Risky Policy, it is not
that that policy does not satisfy actual personal prefer-
ences enough. Similarly with the Exploitation Policy:
since maximal beneficence cannot distinguish between P and
Q, if we find that one of these policies is morally better
than the other, our standard for determining this cannot
be that our preferred choice satisfies more actual personal preferences. Nor can the lack of preference satisfaction be our complaint against the Repugnant and Draconian Conclusions. Beneficence itself cannot solve its own problems as a sufficient criterion of morality. If we are to solve the problems posed for morality by choices that affect future people, we are going to have to widen our conception of morality beyond unrestricted preference utilitarian concerns.

I suggest, however, that the interpretation of the Impersonal Principles of Beneficence as preference utilitarian principles is not the source of the problem. Nor is the notion that our duty in this area of morality is to maximize beneficence. What is missing is a method for evaluating choices that places rational restrictions on the application of whichever principle of beneficence we choose. In the following sections I will argue for a specific method for the rational evaluation of moral choices. When the Impersonal Principles of Beneficence are applied rationally--i.e., within the confines of the proposed method of choice evaluation--they will give us intuitively satisfactory answers to all the problems raised by the Paradox of Future Generations.

A. Rational Morality: The Kantian Background

The method of rational moral decision-making that I am going to propose is Kantian in that it can be derived essentially--although not entirely--from the categorical imperative, Kant's version of the moral law, or law of practical reason. But I wish to be careful here. I use the term "Kantian" to indicate the origins of the method
and to distinguish it from similar models for rational moral decision-making put forward by others. (The models will be compared in Section D.) But if the name draws me into a general exegesis or general defense of Kantian doctrine, I will repudiate the Kantian designation. Kant is a major inspiration for the model. But no study of Kant will provide all parts of the theory. And any study of Kant will reveal much more than I have taken or than I wish to defend as my own view. Nor should anyone take the authority of Kant as justification of the rational system I will put forth. The derivation of a model of rational moral decision-making is justified by the fact of moral reasoning. And the form I attribute to rational decision-making is confirmed as the correct one by the results its application gives, results that confirm our moral intuitions.

Further caveats are in order here. It is conceivable that rational morality could take different forms. I will assume throughout this paper that in fact our morality has one underlying rational or formal structure and that we can discover what this structure is and what strictures it places on moral decision-making. I do not claim that we consciously reason in accordance with it whenever we make moral decisions, but I do claim that fully rational moral judgments must accord with it. Also, I will speak throughout as if "moral rationality" and "practical reason" were synonymous phrases, since Kant interprets morality as practical reason. In reality "practical reason" is a broader term than "morality" or "moral rationality" and applies to the rationality of all voluntary action and not just moral action.
The thesis is, then, that neither the Impersonal Total Principle of Beneficence nor the Impersonal Average Principle gives us a complete criterion of morality towards future individuals if the principle of beneficence is simply given an unrestricted preference utilitarian interpretation. But either principle can be interpreted as a preference utilitarian principle and its shortcomings corrected if it is applied rationally, i.e., under the conditions imposed by moral rationality.

In suggesting that we can look to Kant to correct the deficiencies of utilitarianism I follow a path pointed to by Gregory S. Kavka. Kavka argues that an appeal to a "modified categorical imperative," systematically carried out, together with an independent maximizing principle and a limitation on creating restricted lives, might solve the Paradox of Future Generations. He does not claim, however, to have solved that paradox himself. (Cf. PFI, p. 112)

Kavka's modified categorical imperative is called the **Extended Obligation Principle**, and it says that

One can have an obligation to choose act or policy A rather than alternative B only if it is the case that either (i) if one chose B, some particular person would exist and be worse off than if one had chosen A, or (ii) if one chose A, some particular person would exist and be better off than if one had chosen B.

(PFI, pp. 98-99)

Kavka suggests that this principle needs to be supplemented by a maximizing principle that can weigh the benefits to different potential future people and direct us to choose that which will produce the greatest net
benefit. (PFI, p. 99) He suggests, however, that this principle would have counterintuitive results in certain cases. To meet some of the problems, he adds an injunction against creating restricted lives. (PFI, p.112)

Parfit argues at length against the Extended Obligation Principle. (FG, pp.119ff.) My own immediate objection to it is that it takes a narrow rather than a wide approach to the problem of morality towards future people; and we have already seen that the Paradox of Future Generations can never be solved by an approach which addresses itself to particular individuals. (Cf. Chapter I.) Moreover, we have superseded Kavka's undeveloped maximizing principle with our own.

Nevertheless, Kavka's point is well-taken. The maximizing principle by itself cannot satisfactorily solve the Paradox of Future Generations. Something else is required as well. I have claimed that a rational method for applying the principle will correct the failures of mere beneficence. As we shall see, the Kantian model of moral rationality incorporates features that fulfill our notion of obligation.

Parfit tries to combine maximization of beneficence and obligation in one principle rather than in a maximizing principle together with an obligation principle or in a maximizing principle together with a method for applying that principle. He amends Kavka's obligation principle to a Revised Maximizing Principle, which says that,

If our choice is a necessary part of the cause of the existence of a person with a life worth living, our choice thereby benefits this person. Other things being equal, we ought to
do what would benefit people most. In deciding how to do this, we must compare the benefits to different people who, if we make different choices, would exist. We have no duty to benefit others when this would require from us too great a sacrifice. Most of us therefore have no duty to have unwanted children. But if we do decide to have a child, and other things are equal, we ought to do so in a way that would benefit people most, unless this would require...too great a sacrifice.

(FG, pp. 127-28)

Parfit later abandons the Revised Maximizing Principle in favor of the Wide Total Person-affecting Principle of Beneficence, which he correctly claims is "very similar." (FG, p. 137) Later he adopts Kavka's injunction against restricted lives to attempt to solve the problems of a pure beneficence principle, but without success. (We shall take up Parfit's strategy in Chapter VII.) From Parfit's point of view, as well as from Kavka's, the paradox remains unsolved.

Our own notion of obligation is deeper than either Parfit's or Kavka's. It is that we are obligated always to act within the confines of moral rationality as expressed by the categorical imperative (insofar as possible). Any choice is intrinsically good only so long as it does not violate practical reason. Thus the requirements of practical reason place a limiting condition on any moral choice. To put the same point another way: good choices are rational choices; and the law of practical reason prescribes the conditions for the rationality of choices.
B. The Categorical Imperative and the Postulates of Moral Rationality

In this section I will consider the formal properties of moral reason.

What criteria of rationality does the categorical imperative impose on moral choices? Or, what are the postulates of practical reason?

i. The universality postulate

Act as if the maxim of your action were to become through your will a universal law of nature.10

(FMM [421])

This is the Universality formula. It requires that we treat all similar cases similarly, i.e., that we universalize our prescriptions, prescriptions being the maxims upon which we act when we make a choice. It says that if it is right for me to act in a certain way in certain circumstances, it is right for any similarly situated person to act in the same way in similar circumstances. Many philosophers distinguish between acting rightly and achieving effects. But to take this approach when an action is a choice may be to break the link between that which is chosen (a set of future effects) and the realization of that choice (the same set of effects); and it is not at all certain that a potentially breakable "link" of this type exists: a choice and its effects may be one and the same thing. For those disposed to argue the point I will merely say here that I will interpret the universality postulate as applying to choices conceived as sets of effects. In this way we can judge the universality of our prescriptions by the effects of the choices they
prescribe. Every effect of a choice should be one we would be willing for anyone in similar circumstances to prescribe.

ii. The postulate of respect for persons

Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only.

(FNM [429])

This is the formula of the End-in-itself. It requires that we treat all persons as worthy of respect and not as mere means to an end. And when it is universalized in accordance with the command of i., it requires that we treat all persons as equally worthy of respect. This means that in order to be moral we must treat all people, present and future, with equal respect. Parfit concurs.11 This is important because the belief that we should have equal concern for future people—hence should treat them with equal respect—seems to be a requirement of the Paradox of Future Generations itself, at least implicitly: if we do not have equal concern for future people and if they do not have the right to equal respect, we need not be troubled if we treat them unequally, as we apparently do on the Risky Policy and the Exploitation Policy and the Repugnant Conclusion scenario. But if to treat people morally we must treat all people, including future people, with equal respect, it follows that when we do not treat people with equal respect we do not treat them morally. Treating people with less than equal respect is sufficient to make a choice immoral.

But what is it to treat people as equally worthy of respect? Whatever else respect for persons means it means that we must treat equal interests equally; otherwise we cannot treat people as equally worthy of respect. But I
think it means more than this. I think it means that we cannot make up for harms we do to some people by benefits we give others. If we do this we are using some people as mere means in Kant's terminology. And if we treat some people as means and others not we cannot be treating all with equal respect. Thus I interpret the second postulate of practical reason as calling for parity of interests, where by "parity" we understand both that equal interests count equally and that no one's interests can be subordinated to anyone else's in the sense that we can atone for harm to the one by benefits to the other. The importance of the interpretation of respect for persons as parity in this particular sense will emerge below. (I use the term "interests" here; but where the interest in question is an interest in beneficence we can understand "interests" to mean "preferences"; equal good preferences count equally and equal bad preferences count equally; and we cannot make up for thwarting someone's good preferences by fulfilling someone else's good preferences. The criteria for good and bad preferences and the relationship between them are explained below in Section C.)

We know, then, that our moral prescriptions must be universalizable and that they must respect the interests of people equally. But we can have no complete idea of the rational requirements of morality without the third formulation of the categorical imperative.

iii. The impartiality postulate

Act according to maxims which can at the same time have themselves as universal laws of nature as their object.

(FM [437])
Formula iii looks like i., but Kant clearly intends it to be different. For he says that the third formula follows from the first two as the "supreme condition" of the will's "harmony with universal practical reason, viz. the idea of the will of every rational being as making universal law." (FMM [431]) The difference is that in iii. the will is regarded as legislating for all rational beings, i.e., as autonomously creating the law for them not merely as prescribing in conformity with the principle of universalization. Hence iii. is called the Autonomy formula. It is also called the Realm-of-ends formula, since in another expression of it Kant requires us to prescribe only those laws to which we ourselves would be willing to be subject in a potential universal realm of ends, i.e., a realm where everyone treated everyone rationally. (Cf. FMM [438-439]) Formula iii., then, requires not only that our maxims conform to universal rational law but also that we regard ourselves as legislators of laws for everyone, ourselves included.

What does it mean in practical terms to legislate autonomously universal practical laws of nature to which we ourselves would be willing to be subject? Such laws must be laws which treat ourselves and others as ends and not as means only, i.e., with equal respect for human dignity, by counting equal interests equally and not subordinating the interest people have in not being harmed to the interest they have in being benefited, or, even worse, subordinating the interests some people have in not being harmed to the interests of others in being benefited. Moreover, they will prescribe the same actions for everyone in similar circumstances. But how can we make
the determination that a prescription is universalizable and respects persons? In order to judge whether a prescription is moral, hence fit to realize a universal practical law, we must determine whether or not we ourselves would be willing to bear each effect of the choice if that effect were to be the only effect we ourselves received. In order to do this, we must project each of the outcomes of each of our choices onto ourselves or project ourselves into each of the final positions, which is the same thing, and ask ourselves if we would be willing to take that position. That is, we must make our decision as if we might end up occupying any one of the final positions of the choice. In that event we would want to ensure that no final position violates the canons of rationality. If one does, given that we ourselves might be the unlucky person to occupy that final position, we should reject the choice that entails it.

The postulates of practical reason together give the formal criteria for the rationality of moral choices. But we should note one point about them. Merely by consulting these rational criteria we cannot know whether or not any given choice is irrational. We must submit the choice to an empirical test using the criteria. Postulate i. requires that the effects of a choice be such that we as rational beings would be willing to universalize them. We can make this determination only by deciding whether or not we would indeed be willing to universalize our own prescriptions. Postulate ii. requires that we judge the rationality of choices by determining whether or not each of their final effects treats persons equally, again in our own estimation. And postulate iii. requires that
every effect of a choice be such that we would be willing to bear it ourselves in the circumstances into which we project ourselves. We can make the determination of the rationality or intrinsic goodness, of an effect of a choice only by deciding whether or not the effect violates universality, parity of interests, or impartiality towards those affected. These must all be empirical determinations.

In the following section I will show how we can judge the rationality of choices made under the principles of beneficence by considering the requirements of practical reason as conditions for the rationality of choices.

C. The Kantian Model for Rational Moral Decision-Making

Once we have defined the three postulates of moral rationality as a Kantian understands them we are in a position to develop a method for rational moral decision-making.

On the Kantian model the term "moral reason," or, for Kant, "practical reason," refers on three levels—four if we distinguish between the prescription to take a certain action and the action itself as realized in its effects.

First, "moral reason" refers to the form of moral reason itself, i.e., the principles, or postulates, of universality, respect for persons, and impartiality. When this form is taken to give a general directive to human action, moral reason become a "categorical imperative," i.e., an imperative, or command, dictated by practical reason itself, or the form of duty in general.
Second, "moral reason" refers to what Kant calls "practical laws of nature," i.e., laws describing specific human duties. A genuine practical law will have the form of moral reason. That is, it will respect universality, human dignity, and impartiality. A purported practical law that says, for example, "Push down anyone who gets in your way," obviously violates the form of moral reason. Therefore this purported practical law cannot be a genuine moral law. But a law that says, "Maximize beneficence," could be, and is, a genuine practical law. This law does not intrinsically offend any of the criteria of moral reason; rather, it conforms to the law of reason. For this reason genuine practical laws of nature are derivatively called "categorical imperatives."

Third, "moral reason" refers to the form a choice takes. Whenever anyone proposes a potentially moral decision to himself he prescribes a certain action in certain circumstances. The prescription, or maxim, must not violate the canons of moral reason, or the action will be irrational. Subjectively, the prescription is a directive for action. An irrational prescription might be, "I will push down anyone who comes through this door." A rational one might be, "I will maximize beneficence by choosing this given alternative over the other." Objectively, a prescription is a directive to instantiate a practical law of nature. The irrational "law" instantiated by the first prescription is, "Push down anyone who gets in your way." The rational law instantiated by the
second prescription is the principle of beneficence. It is only through such prescriptions that practical laws are instantiated and actually exist in the world. (But they exist in the abstract as laws of reason whether they are instantiated or not.) Thus when a person prescribes a certain action he legislates or brings into actual being a practical law. If the law conforms to reason it is a genuine practical law, and if not it is irrational and no true law. At the same time, when a person prescribes a certain action he makes a rational or irrational choice.

A prescription, then, prescribes or commands a choice. And a rational prescription prescribes or commands rationally. The criteria of rational moral choice are formal; that is, they are given by the form of rationality. But, just as with the specific ends of a duty (the maximization of beneficence, for example), the only way to determine whether or not formal, or rational, moral ends are realized in action is by consulting the predictable effects of our choices. If a choice is made in the area of morality that deals with beneficence, it will have certain effects which we can translate into terms of actual personal preference satisfaction. (For example, if we choose Eden (1), Eve will exist at a utility level of 100.) And if all we look at is utility information, the total outcome of a choice in the area of beneficence will consist in a set of effects, or final positions, in terms of utility, or preference satisfaction. But if we look at the choice both in terms of utility and in terms of rationality, the outcome of the choice will be understood as a state of affairs having both utility properties and formal properties. And each of the final positions of
that total state of affairs will have both utility and formal characteristics, since the total outcome is nothing but the set of the effects of a choice. An intrinsically rational moral choice made under the principle of beneficence will create an over-all state of affairs, or outcome, characterized not only by utility information but also by its conformity to the criteria of moral rationality. And an intrinsically irrational decision in this area of morality will contain utility information and will violate the postulates of moral reason in at least one of its final positions.

From this characterization of moral or rational choices it follows that if we take a rational view of morality we will believe that utility information alone insufficiently characterizes the outcome of a choice. It also follows that we will judge a choice both by its utility information and by its formal properties and that we will judge to be intrinsically irrational any choice that has any intrinsically irrational effect.

Prescriptions, then, prescribe choices. And choices may have irrational effects. If a choice has any predictable irrational effect the choice is irrational and so is the prescription that directs the moral agent to make that choice. This may be the case even though there is nothing irrational about the practical law under which the action is taken. For this reason a prescription may violate reason while the law it instantiates does not. Thus a practical law may be rational and its application irrational. All that is required for this to be the case is that something in the choice we propose to make--hence in the prescription through which we propose to realize the
law--violate reason. This is a crucial point. We shall see its import below.

Another point. Although these three levels of rationality exhaust the possible references of the term "moral reason," they do not exhaust the possible meanings of "reason" itself. "Reason" is a broader term than "moral practical reason." It is also rational to work to achieve one's own purpose or goals and irrational to defeat them. In this sense it is rational--having set oneself the intrinsically rational goal of maximizing beneficence, to maximize beneficence.

But if it is rational to maximize beneficence through one's actions when that is one's specific goal, and it is also rational to follow the dictates of practical reason in moral choice, and a choice that maximizes beneficence (hence is rational in the first sense) also violates the postulates of practical reason (hence is irrational in the second sense), how can we reconcile our two applicable rational objectives? Clearly one must take precedence over the other; and Kant cannot help us here because he does not deal in helpful detail with applied ethics (or what he would call "moral anthropology"). Yet this is a problem of applied ethics since it arises when we invoke a moral practical law to justify an immoral or bad choice. The answer to the problem is to accord the general formal ends of practical reason lexicographical priority over the specific ends of any practical law. In this way conformity to practical rationality becomes a condition for the moral application of any practical law.

Let us see what difference it makes to the Impersonal Principles of Beneficence to condition their application
on the conformity of the application to practical reason. We wish to interpret these principles as preference utilitarian principles. We can express a Comprehensive Impersonal Principle of Beneficence, using this interpretation, in the following way:

Given alternative choices, P and Q, if P would give the P-people greater (average/total) preference satisfaction than Q would give the Q-people, when the effects of both alternatives on presently existing people are taken into account, choose P, subject to the condition that P conform to the postulates of practical reason. If any consequence of either alternative violates the postulates of practical reason the alternative containing that violation should be removed from consideration.

The problem, of course, is that all our alternatives may violate the postulates of practical reason. We need to extend the rational condition of application of the principles of beneficence to say that,

If all alternatives violate the postulates of practical reason, choose that alternative which violates them least.

This condition of moral rationality is the traditional Principle of the Least Evil. When we add it to the Comprehensive Impersonal Principle of Beneficence a complete rational model for the moral evaluation of choices emerges. We shall see how it works in more detail below. First some further explanation of the system is required.

On our view choices (or prescriptions for action) that meet all the requirements of practical reason are
intrinsically good and those that violate them are intrinsically bad. A choice is therefore intrinsically good if and only if none of its effects violates the postulates of practical reason and intrinsically bad if-and only if any of its effects violates those postulates. But an intrinsically bad choice may nevertheless be better than the alternatives. In this case it is intrinsically bad but relatively better, i.e., less bad, and we should choose it.

The Principle of the Least Evil requires elaboration.

What do we mean when we say, Choose that alternative whose effects least violate the postulates of practical reason? We mean that each violation should be assigned a degree of harm. Determining the degree of harm of an effect is similar to determining the quantity of beneficence of an effect in that offenses against practical reason can be translated into intrinsically bad preference satisfaction. Say that on a scale of -1 through -100 we would assign a certain harm a -50. (The harm might, for example, consist in torturing people by tearing out their fingernails.) Then assigning this harm a -50 is identical to assigning the satisfaction of this intrinsically bad preference a -50. We then must ask how many people will be affected by this harm as the effect of a certain choice. Call the choice A- and the effect in question A-1. Suppose the number of people affected by A-1 is 100. We then multiply the assigned degree of harm by the number of people afflicted by it in order to determine the total harm of the effect, or the total intrinsically bad preference satisfaction. In this case the total harm is -5000. If more than one effect, or final position, of a choice is
harmful, we then add, or sum, the products obtained in this manner to determine the total harm, or total bad preference satisfaction, of the choice. This method of calculating harms enables us to compare the relative harms of different effects of the same choice and of different choices. It provides a standard of commensurability of harm. This operation is performed prior to the calculation of the benefits—or intrinsically good preference satisfaction—of the same effects and is independent of it.

We can now see how we can determine which of two choices entails the least evil. Suppose \( A_{-1} \) is the only harmful effect of \( A_- \). And we have one other alternative choice we can make, \( B_- \). Like \( A_- \), \( B_- \) has only one harmful effect, \( B_{-1} \). The degree of harm of this effect is slightly less bad, however, than the degree of harm of \( A_{-1} \). We assign it a \(-48\). (We can suppose that \( B_- \) consists in putting a lighted match between people's toes rather than tearing out their fingernails.) But \( B_{-1} \) affects twice as many people as \( A_{-1} \), 200 people. We must multiply the degree of intrinsic badness of \( B_{-1} \) (-48) by the number of people affected (200). The total intrinsic badness of \( B_{-1} \) is \(-9600\). We should choose \( A_- \) over \( B_- \). The degree of harm on \( A_- \) is slightly more, but the number of afflicted people is much smaller. On the Principle of the Least Evil we choose the lesser total harm over the greater.

Obviously the case is an artificial one in that it assumes that these are the only available choices and that we are actually trying to do the least immoral thing. In reality we also always have the option not to torture
people; and when we do torture them we are usually not trying to harm people as little as possible. But we could consider torturing the few to be a lesser evil than, for example, leaving them or their confederates, to kill many other people. The Principle of the Least Evil would then direct us to torture as little as necessary to prevent the greater evil, if this were the only way to prevent it and the evil were certain.

Here we must make an essential distinction. Harms, as we have been using the term, are violations of the canons of practical reason. They are intrinsically bad. They are thus costs of a particular gravity. As such they must be distinguished from mere costs, or disutilities, or merely negative—as opposed to intrinsically bad—preference satisfaction. Mere costs are those negative effects of our choices that do not violate practical reason. These are costs that we would be willing to bear ourselves and to have anyone prescribe in the given circumstances. In general these will be costs we think are compensated for by benefits to the same persons. The reason for this is that it is usually irrational—and unacceptable or inappropriate—to compensate for a cost to me by a benefit to you. For example, you can compensate me for your running cattle on my land by paying me for grazing rights. But it would be irrational for me to allow you to run cattle on my land and for you to compensate yourself. For me to accept this state of affairs would be to violate my legitimate self-interest. This relation between costs and benefits can be expressed by indexing costs to benefits. It is irrational to accept for oneself and prescribe for others costs that can be indexed and are not or are
inappropriately indexed. (Assuming the costs warrant indexing. Some are so minor as to be entirely discounted. We can make this determination empirically.) We figure costs in when we calculate benefits. Harms are calculated separately. These are costs that either are not compensated for appropriately on a given choice, with respect to the people they affect, or cannot be compensated for with respect to those people. This incompensatoriness makes their acceptance for oneself—hence their prescription for others—irrational.

This distinction between mere costs and harms is crucial because without it any cost whatsoever might be construed as a harm. If this were done, any cost would be sufficient to make a choice intrinsically bad. And, since almost all choices have some costs, almost all choices would be intrinsically bad. When we say that harmful choices are intrinsically bad, we mean something more than this. We mean that costs that violate practical reason are intrinsically bad and so are all choices that entail such a cost. We can assume that mere costs have already been subtracted from benefits to arrive at total benefits, or total good preference satisfaction, in the discussion of benefits throughout this paper. Our concern is with the remaining relationship between benefits and harms.

Notice that the Principle of the Least Evil does not contradict the postulate of respect for persons. It does not say that we can make up for harms to some by benefits to others. Moreover, it does not say that we can make up for harms to ourselves by benefits to ourselves. Rather, we cannot. In one sense this last claim is trivial since we, in fact, will not impose on ourselves what we consider
to be uncompensated harms, and any cost we impose on ourselves we can therefore assume to be compensated for by benefits to ourselves (or possibly to others, e.g., to our children). But in another sense the claim is not trivial. For it says, in effect, that it is moral to impose on people costs that are compensated for by benefits to them, but it is not moral to impose uncompensated harms on people and say it is for their own good.

The distinction between costs and harms is particularly useful here. And it is perfectly general. Once we allow for indexing to maintain the appropriate relation between costs and benefits we can say that harms to some cannot be made up for by benefits to the same person or to anyone. In this way we maintain the impersonality of morality. But at the same time we acknowledge the rational necessity of maintaining indexical relations between costs and benefits, the harmful character of costs that are not indexed to benefits, and the benign character of costs that are. Indexing also allows us to say why similar costs are harms in some cases and mere costs in others. It is a mere (although great!) cost if I give up my life to save my child's life, but a great harm if my life is taken to benefit someone to whom I do not bear the cost/benefit relation.

Notice, however, that although a harm to someone cannot be compensated, hence justified by a benefit, a harm can be justified by a greater harm, so long, again, as the harms are indexed to each other, as in fact they were in our torture versus terrorism example. In this way we maintain the parity of interests, i.e., of harms
vis à vis harms and of net benefits vis à vis net benefits. And I claim that this concept of parity fully captures in an objective way the intuitive notion of respect for persons.

There is another crucial point to note. Our exposition assumes that the harms in question are predictably certain. If they are not certain they should be treated like uncertainties under Bayesian decision theory. A harm should be diminished for the purposes of moral calculation in proportion to the decrease in the likelihood of its occurrence. And a harm of whatever magnitude whose probability of occurrence is so small that a rational person would take the risk of its occurring should be counted at its diminished force as a mere cost and not as a moral harm. In this way we can avoid calling intrinsically immoral those choices which have only a very small chance of entraining harmful effects.

If we distinguish the calculation of harms from the calculation of benefits and take conformity to the Principle of the Least Evil as a condition on the satisfaction of preferences, we avoid a severe problem of unrestricted average preference utilitarianism.

Suppose that what matters to us morally is the maximization of average beneficence and we do not distinguish harms, as violations of practical reason, from mere costs. All we acknowledge are benefits, or utilities, and costs, or disutilities. Therefore we do not place the avoidance of total harm above the satisfaction of "good" preferences. All that matters is average preference satisfaction when costs are subtracted from benefits. Now suppose that all our alternatives are certain to be strongly negative in terms of preference satisfaction.
That is, whatever we choose, the affected people will have to endure some sort of hell. If our standard of morality is simply average beneficence, it seems that we should carry over to negative cases the same measure of the value of choices that we use in positive cases. Thus if we employ the Impersonal Average Principle of Beneficence to determine the goodness of beneficent choices, we should use it to determine the relative "goodness"--i.e., lack of badness--of harmful choices.

Let us take the example we just used, comparing A- and B--; only, for simplicity, let us say that A- and B- are not only the only negative effects of our alternatives but the only effects there are. It is necessary to make this qualification because on our own method of calculation harmful effects --i.e., not merely negative effects but intrinsically bad ones--take priority over intrinsically good ones. Therefore we can judge two alternative choices merely by comparing their harmful effects, if there are any. A utilitarian does not distinguish harmful effects from mere costs. Therefore he has to compare all the effects of each of his alternatives in order to determine which choice has the most benefits when costs are subtracted. This is the only way he can compare alternatives. Thus we could compare A- and B- merely by comparing their harmful effects, but a utilitarian could compare them only by comparing all their effects. So we will say that A- and B- have only one effect each and that effect is harmful.

We have already seen that on our own system we should choose A- over B-, since B- entails more total harm. But in our guise as utilitarians we get a different result.
First we determine the disutility on each choice. A- comes out with an average utility quotient of -50; and B- comes out with an average utility quotient of -48. The level of average negative beneficence in B- is not as low as the level of average negative beneficence in A-. Therefore, we should choose B-, even though twice as many people suffer hellish lives as if we had chosen A-, and those lives are only slightly less bad than the lives in A-.

It is easy to see that extending the provisions of an unrestricted average principle of beneficence to cover negative cases leads to a kind of **Negative Repugnant Conclusion**:

For any possible and large population, A-, with a hellish average quality of life, there is some much larger population, Z-, with an average quality of life only slightly less bad, whose existence, other things being equal, would be better and would be what we ought to bring about, even though the total suffering in the world would be greatly increased.

(The conclusion is mine; I frame it in Parfit's terms.) A case of this type would replace the decrease in positive beneficence for each succeeding population on the RC scenario with a decrease in negative beneficence on the Negative RC scenario. This would give us the following graph:

![Graph](image-url)
If we judge this scenario by the Impersonal Average Principle of Beneficence as required by a utilitarian system, we must choose Z- over A-. If we judge it by the Principle of the Least Evil, as required on our own system, we must choose A- over Z-. We do not reach the Negative Repugnant Conclusion if we distinguish between the calculation of harms and that of merely negative and positive preference satisfaction and give priority to avoiding harms over satisfying good preferences in determining whether a choice is fit to be made.

Let us see how our method for the rational evaluation of choices works.

This model for rational decision-making in the area of beneficence entails that where one alternative is moral or intrinsically good—i.e., respects the postulates of practical reason and maximizes beneficence— in the wide sense, i.e., maximizes only intrinsically good preferences — and the other respects the postulates of practical reason but is less beneficent — or maximizes fewer good preferences and no intrinsically bad ones—we should choose the more beneficent alternative. That is, we should choose the more beneficent of two good alternatives. It also entails that where one choice is maximally beneficent but intrinsically bad—i.e., includes intrinsically bad preferences, or harms, among those preferences satisfied—and the alternative is less beneficent but intrinsically good—i.e., includes no intrinsically bad preferences among those satisfied—we should choose the less beneficent alternative. We should choose a good alternative over a bad one. It entails that where two choices are intrinsically bad—both satisfy intrinsically
bad preferences—we should choose that which is less bad in degree.

Finally, the model entails that choices with equal outcomes in terms of beneficence can be evaluated differently. This is the case if one has an outcome which offends practical reason and the other does not. This happens if two choices are equal in total or average good preference satisfaction, but one satisfies bad preferences in addition to the good ones and the other does not, or if both satisfy good preferences equally and bad preferences as well, but one includes a greater total amount of bad preference satisfaction than the other. Thus where two choices are equal in beneficence but one is intrinsically bad and the other is not, or both are intrinsically bad but one is worse than the other, we should reject the intrinsically worse alternative. This is possible because the maximization of beneficence is not our only, or even our primary, moral concern.

Here we should notice an important implication of our theory. Once we say that a choice is intrinsically bad when it violates the postulates of practical reason, hence should be rejected in favor of an intrinsically good but less beneficent choice, our principles of beneficence as restricted by reason no longer imply Parfit’s Claim (A), the claim that,

"If the same number of lives would be lived either way, it would be intrinsically worse if those who live are worse off than those who would have lived." (Cf. p.29)

This is not surprising since in connection with developing the rational method for the evaluation of choices with
respect to beneficence we have redefined intrinsic goodness and badness. Parfit, however, claims that a satisfactory principle of beneficence should not only imply Claim (A) but extend its provision to Different Number Choices as well. The unrestricted Impersonal Principles of Beneficence did extend Claim (A) to Different Number Choices. That was one reason they proved unsatisfactory.

On our view Claim (A) is too strong. We are justified in rejecting it because, as we shall see, our system gives intuitively better results than an unrestricted actual preference utilitarian system, which would endorse it. Not all utilitarian systems are unrestricted actual preference satisfaction systems, however. Several eminent utilitarians have put forward utilitarianism as a rational system, indeed as the objective rational moral system. Moreover, our own system has very strong affinities with these rational utilitarian systems. Therefore, in the next section I will compare our version of rational morality with the utilitarian versions.

D. Other Models for Rational Moral Decision-making

Certain leading utilitarians, like R. M. Hare,14 John C. Harsanyi,15 and Richard Brandt,16 have been concerned with the possibilities of utilitarianism as the unique system of morality dictated by practical reason.

These utilitarians acknowledge postulates of practical reason almost identical to those we have acknowledged. They believe that equal interests should count equally. (But they do not believe that respect for persons implies that we cannot make up for harms to some people by benefits to others.) They believe that moral
prescriptions should be universalizable. And they endorse the impartiality that results when the moral agent puts himself into all the final positions of his own choices. Yet they conclude that the three concepts of respect for persons, universality, and impartiality entail a utilitarian moral theory.

Why do they get a different result from concepts of moral reason very similar to ours?

There are two essential differences between a rational utilitarian theory and our own theory of rational morality.

First, utilitarians assume that the supreme moral duty or practical law—and the only one to which the postulates of practical reason apply—is the principle of beneficence. This assumption allows for a certain economy of theory in that it eliminates the steps between the law of reason itself and the realization of that law in the outcomes of choices. But it means that the postulates of reason are subordinated to the maximization of beneficence. On our system it is the other way around: the maximization of beneficence is subordinated to reason.

The second difference derives from the first. On a utilitarian system all that finally matters morally is the maximization of beneficence over all. The agent is viewed as that abstract individual who bears all of the effects of his own choice. He will therefore want to maximize "his" utility, i.e., the utility level of all the final positions of the choice taken together. And any harms "he" suffers in any final position will necessarily be subtracted from "his" over-all utility as mere costs of maximizing "his" utility. On our system equal respect for
persons is seen as entailing not merely that equal interests count equally but that harms to some people cannot be made up by benefits to others. Hence the agent is viewed as the abstract individual who will end up bearing any one of the final positions of his own choice. He will therefore want to make sure that each final position of the choice is intrinsically good. Thus he will put each final position to the test of rationality: Would he prescribe it for himself if it were the only final position he were to occupy? Would he be willing to prescribe it for everyone? Could he will all other people in circumstances similar to his own to prescribe it? The answers to these questions will be determined empirically by the agent's consulting his own intuitions.17

Utilitarianism is, then, very different from our system. We can best understand how rational utilitarianism works by starting from the ground up, not from the top down, as we did in analyzing our own system.

The utilitarian has no way of extending the requirements of practical reason to each final position of a choice: practical reason requires only the maximization of beneficence over all the final positions. The problem with this is that many preferences that people actually do have are notoriously bad or "irrational," and utilitarians, like us, want to maximize preferences in a rational manner. The utilitarian solves this problem by restricting the maximization of preferences to the maximization of preferences that are rational in their own right. But what preferences are these? For us rational moral preferences are those that meet the criteria of practical reason. For a utilitarian rational preferences are "true"
preferences. These may be defined as prudential preferences, i.e., the preferences a person would have if he were perfectly prudent and well-informed, or they may be defined as those personal preferences a person would have if he were fully informed and disposed to act rationally, with no mention of prudence.

The trouble is that there is no necessary connection between maximizing one's own "true" preferences as an actual person and maximizing the "true" preferences of everyone as a moral agent. This is the problem on which Sidgwick floundered one hundred years ago. Indeed, in maximizing one's own "true" preferences one may defeat those of others and vice versa. Therefore, for a utilitarian, moral decision-making must be of a different order from ordinary decision-making aimed at the satisfaction of one's personal preferences. This is where the model for rational moral decision-making comes in. Personal preference satisfaction must be submitted to the requirements of practical morality in order to emerge as moral preference satisfaction. That is, in order to act morally a person must universalize the prescription to maximize "true" personal preferences and he must do so impartially, i.e., as if he were to occupy any of the final positions of his own choice himself, hence all the positions at one time or another. As a result of employing this rational method of decision-making where others besides himself are concerned, a moral agent will maximize the prudential preferences of all the people affected by his choice (viewed as different aspects of "himself"--the agent who bears all the effects of the choice), counting equal interests equally. That is, he will express a special kind of
preference, a moral preference. And this is the only way he can make a moral choice.

But there is still no essential connection between the rationality of the preferences maximized (which are rational$_1$ because they are the "true" preferences of those affected) and the rationality of a moral preference (which is rational$_2$ because it has the form dictated by practical reason). Rather, there remains an implicit tension between rationality as the maximization of one's own "true" preferences and rationality as the expression of moral preferences. The two may still conflict if it is not in my "true" interests to maximize the true interests of others. This leads to a particularly embarrassing question for a rational utilitarian: Why should I be moral when my own personal preference to maximize my rational interests conflicts with my moral preference to maximize the rational interests of all? Or, even worse, how can I be rational (rational$_1$) if I do not maximize my own rational interests; and how can I be moral (rational$_2$) if I do? On our own view, by contrast, a moral choice is not one that rationally maximizes preferences whose own rationality is guaranteed by a different standard. A moral choice is one which conforms to the same standard of rationality that characterizes each of its ends. We avoid a potential source of conflict within rationality itself. (Nevertheless, we do not thereby avert the question, Why should I be moral? Moral ends are simply taken as ends we do have; and why we have them and why they have the form they do are separate questions which I will not address in this paper.)
We might ask what justification rational utilitarians offer for their view. When we see their answers we may be better able to judge the status of our own system in comparison. Two utilitarians who have given justifications for the utilitarian version of moral rationality are R. M. Hare and John C. Harsanyi. I will examine the arguments of both.

Hare attempts to derive the principles of moral reason from the meanings of the moral terms. (ETU, P. 24) We need not dispute his designation of the relevant moral terms as universality, impartiality, and respect for persons, since we have derived the postulates of moral reason from the same terms ourselves. We should notice, however, that we have interpreted "respect for persons" differently from Hare. For a utilitarian respect for persons requires only that we count equal interests equally; for us it requires not only this but also that we not compensate for harms to some by benefits to others. Because of this difference in the interpretation of one of the moral terms we were forced to an interpretation of the model for rational moral decision-making in cases of beneficence different from the utilitarian one. That is, we see the moral concepts themselves as requiring us to reject as intrinsically immoral any choice with any effect we would not accept for ourselves. And we claim that we would accept any actual effect which did not violate the canons of practical reason. A utilitarian, however, as we saw, accepts whatever choice maximally satisfies "his" preferences over-all when he is viewed as the recipient of all the effects of each choice. Hare follows the model of utilitarian moral decision-making we have described. And
he argues that this utilitarian calculus should be applied only to maximize the sum of "true" or prudential preferences, i.e., those preferences that the people affected by a choice would have if they "were perfectly prudent - i.e., desired what they would desire if they were fully informed and unconfused," not rational preferences in our sense. (ETU, p. 28)

To defeat Hare's argument it should suffice to point out either that the system as he envisions it is not a necessary deduction from "the moral words" or that it is not a system sufficient to capture our notion of morality. As it happens, both objections can be made.

In the first place, other rational moral systems can claim to be derived from the meanings of the same moral terms. Our own is an example. So it is not clear that utilitarianism is a necessary derivation from the meanings of the moral terms.

In the second place, as the Risky Policy, the Exploitation Policy, and Different Number Choices indicate, rational utilitarianism is insufficient to capture our full notion of rational morality. We can call these choices bad on our system because at least one of the effects of each violates the postulates of practical reason, as I shall show below. Hare can bring no such objection to bear against them. For him a choice is irrational if and only if it fails to consider the utility interests of all affected in the prescribed manner and thereby fails to maximize prudential preferences. And these choices do this. Nor are these the only cases that satisfy prudential preferences without satisfying out moral intuitions. Consider Genghis Kahn. Arguably it was
in the true, or prudential, interests of the Tartar hordes to rob, rape, and kill the villagers of Central Asia and in the true, or prudential, interests of the villagers to resist. But who can say that prudential preference satisfaction would have been greater if the resistance had been successful? Not only does reliance on prudential preference satisfaction as a moral standard allow for counter-intuitive results; in many cases it leaves us with uncertain results, since there is no formal criterion for even determining what "true", or prudential preferences are, much less which should prevail.

Moreover, even if we could be sure of admitting only and all prudential preferences to the calculus, in many cases prudential preferences do not even seem directly relevant to choices. When we look at the Risky Policy and the Safe Policy, for example, we want to know the effects of each choice in terms of actual preference satisfaction. We do not want to know whether these desires are prudent. (Maybe coal is better than nuclear energy from a prudential point of view.) So Hare's system seems neither necessary nor sufficient to explain the concept of moral rationality.  

But there is an even graver objection than these. Hare is claiming that the entire system is derivable from reason; and this is not so. What is derivable from reason is the form of moral choice, i.e., the postulates to which rational choices must conform. The end of maximizing beneficence, i.e., the substance of moral choices, is not given by reason but is a prior assumption. The maximization of utility is assumed to be the aim of morality and
reason is put to its service. We agree that the maximization of utility is an aim of morality; but we argue that this is so because it does not conflict with reason.

Harsanyi also advances a utilitarian theory of moral rationality. But his justification for the system is entirely different from Hare's. Harsanyi believes that utilitarianism can be derived from the fact of practical reason. What he has to say should give support to our own view, because Harsanyi has strong arguments that practical reason or morality exists as a distinct branch of rational decision-making. And we have claimed that rational morality must seek its justification in fact (and the suggested form must seek confirmation in results).

Harsanyi argues that ethics belongs to the same "basic discipline" as all other types of rational decision-making under risk and uncertainty. (Risk is distinguished from uncertainty in that all the possible outcomes of an action are known in advance under the former, although it is not know which of the outcomes will actually come about, whereas under the latter not all the possible outcomes of an action are known in advance). All branches use essentially the same method—Bayesian decision theory—and all employ closely related axioms. Moreover, basic ethical problems can be reduced to problems of rational decision-making. (MTRB, pp. 42-43)

I wish to focus on Harsanyi's derivation of utilitarianism from reason. Harsanyi distinguishes a primary and a secondary definition of rationality in each area of rationality. The primary definition of moral rationality is given by a decision model, the "equiprobability model
for moral value judgements," derived from universalization, impartiality and respect for persons (in the utilitarian sense) applied to preferences. Like Hare, Harsanyi thinks of morality as expressing "preferences of a very special kind," (MTRB, p. 44), namely those preferences we would have if we were to occupy any and all of the final positions of our own choices and wished to maximize the "personal preferences" realized in these final positions. As we shall see below, Harsanyi defines "personal preferences" as "true preferences" in much the same sense as Hare, and for the same reason. The secondary definition of morality -- the maximization of expected utility -- is then derived from the primary definition as "a straightforward mathematical problem...in terms of maximizing the average utility level of all individuals in the society." (MTRB, p. 44)\(^{19}\)

At this point Harsanyi makes a very interesting statement, namely that "discovery of an appropriate primary definition is always essentially a philosophical—that is, a conceptual—problem." (MTRB, p. 44) This implies that getting the notion of moral rationality right in the first place—or getting the best possible decision model—is a conceptual problem. And we may take this to imply that there are conceivably other ways of conceptualizing moral rationality than Harsanyi's equiprobability model; and the proof of the pudding must be in the eating. The only way to know we have got it right is to try it out. (But we can be justified in believing that there is some correct—or at least some best—system of moral reasoning.)
It seems ungrateful to demur from Harsanyi's model after all the work he has done for us. But he himself admits the insufficiency of the system and his own inability to overcome the problem. Harsanyi recognizes that the equiprobability model offers no way to distinguish "bad" actual preferences from "good" ones. Having started out with "personal" preferences as the subject matter on which morality works, Harsanyi finds himself driven to define a person's preferences not as his "manifest" preferences, i.e., his actual personal preferences, but as his "true preferences, i.e., "the preferences he would have if he had all the relevant factual information, always reasoned with the greatest possible care, and were in a state of mind most conducive to rational choice." Therefore, "a person's rational wants are those most consistent with his true preferences...whereas irrational wants are those that fail this test." (MTRB, p. 55)

Like Hare, Harsanyi realizes that if morality is to be thoroughly rational only rational or "true" preferences and not actual preferences - must be realized. But, unlike Hare, Harsanyi himself does not believe that maximizing true preferences as he defines them in moral decision-making will solve the problem of bad preferences. He argues that we must replace "irrational" preferences with rational ones. But, he adds, "I think we have to go even further than this: some preferences, which may very well be their 'true' preferences under my definition, must be altogether excluded from our social-utility function. In particular, we must exclude all clearly antisocial preferences, such as sadism, envy, resentment, and malice." (MTRB, p. 56) But Harsanyi concedes that he cannot offer
"any clear and formal criterion for defining legitimate and illegitimate interests." (MTRB, fn. 24, p. 57) That is precisely the defect we would expect to correct on our own system by extending the criteria of moral rationality to the preferences to be maximized.

Harsanyi is, in effect, acknowledging that his own theory is incomplete. But, in addition to that, the theory, as a utilitarian theory, could not be made complete merely by devising a criterion for distinguishing legitimate and illegitimate preferences. Harsanyi would still have to decide whether simply to exclude bad preferences from figuring in his calculations at all (which would destroy the ordering of preferences) or whether to exclude all choices that include bad preference satisfaction (as I have done). And if he made the latter decision he would have to have a way of deciding what to do when all available alternatives satisfied bad preferences. In other words, Harsanyi cannot solve the problems of his system as a utilitarian, but he could solve the problems of morality as a rational system by moving to ours. He would have to surrender none of the postulates of practical reason. He would merely have to use them differently. He would have to subordinate beneficence to rationality rather than the other way around.

It appears, then, that neither Hare's nor Harsanyi's rational utilitarianism is a sufficient system of rational morality. And it appears that the justifications they offer for their systems--from the meanings of the moral terms, or from the fact of moral reason as a distinct branch of decision-making--can be appropriated by other systems, such as our own. Indeed, our system seems to be
better justified on the same grounds, since it seems complete (with respect to beneficence) and the utilitarian systems are not. Therefore it is better to take our own, less economical, approach to the structure of moral rationality than the utilitarian one.

E. Conclusion

The Kantian model for rational decision-making seems superior to similar models put forward by leading utilitarians. And it fulfills Kavka's suggestion that a Kantian notion of obligation is required to restrain the principle of beneficence. It meets this suggestion by taking as the criteria of rational decision-making notions that flesh out the concept of moral obligation. Our primary obligation to future people, as to all people affected by our choices, is to treat them according to the dictates of universality, impartiality, and respect for persons, insofar as possible. Correspondingly, people have a right to be treated rationally. Rights are therefore, in the first instance, constraints on moral action. Only subordinate to this could we construe a right to maximal beneficence. This is the case because specific duties, like the duty of beneficence, are subordinate to the law of practical reason. Therefore, any rights corresponding to them are subordinate to the right to rational treatment in general. In this sense of "rights" a rational decision made using our model for moral decision-making is one that respects people's rights. And our system is therefore a "goal rights system," i.e., one that respects people's rights in reaching goals, or
incorporates consideration of rights into the judgment of final states of affairs.

As I shall show in the following chapters, the Kantian approach to beneficence allows us to solve problems that are demonstrably irresoluble using the unrestrained actual preference utilitarian method or either of the two rational utilitarian methods we have examined. Nevertheless, our approach is subject to objection. In the next chapter I shall defend the Kantian approach from two lines of attack: first, that our particular theory suffers from internal weaknesses, and, second, that no appeal to rights—specifically the right to rational treatment—can solve the problems of morality where future people are concerned.
V. Defense of the Model for Rational Decision-Making

Parfit has argued that no theory which includes an appeal to rights can solve the Paradox of Future Generations. Our theory includes a fundamental appeal to the right of people to be treated according to the requirements of practical reason and, subordinate to that, to be treated with maximal beneficence. Therefore we must meet the challenge to the rights approach.

We have argued that choices are intrinsically bad when they violate the requirement of practical reason that we treat future people in a way we ourselves would agree to be treated if we were to bear the effects of our own choice. But future people, it might be argued, differ from ourselves in a crucial respect. Those who live under the Risky Policy, to take one example of an intrinsically bad policy on the Kantian approach, have no other chance for existence. So it seems that in projecting these people into our own place; or vice versa, we overlook the fact that because of our choice these people benefit from lives which they would not otherwise have had. It seems natural to ask whether we ourselves, as the Risky Policy people, would not accept the Risky Policy, complete with the nuclear disaster, knowing that this was the only way to exist and that we would have lives worth living. Would we not be better off than we would otherwise have been? Therefore, would we not prefer to exist, even knowing our lives would be shortened by the very terms under which we were given existence? And how can a choice be immoral if it makes people better off than they would otherwise have been? Indeed, we ourselves have acknowledged that if
every one is treated as beneficently as possible and there are no other factors to make the choice bad, the requirements of practical rationality are not infringed.

As Parfit puts the question, "Can it be wrong to harm others, when we know both that the people harmed would not regret what we are doing, and that what we are doing would not be worse for these people than anything else that we could have done?" 22

Let us take first the question, "Can it be wrong to harm others when we know...that what we are doing would not be worse for these people than anything else we could have done?" (Italics mine) Our answer is, of course, yes, it could be wrong. Moreover, we would expect anyone who takes the Paradox of Future Generations seriously to agree with us. For this is a narrow claim. And the point of the Paradox of Future Generations is that it forces us to take a wide view of morality. If we do this, we must agree that what matters in beneficence is not whether or not a choice is maximally beneficent for those who actually are affected by a choice but whether or not a choice is maximally beneficent given our alternatives. So even if we are unrestricted actual preference utilitarians we will deny this claim. We will argue that maximal beneficence to those who actually do live is insufficient to make a choice that causes them to exist moral.

This claim is standing in, however, for a general claim that can be interpreted in either the narrow sense or the wide sense. This is the claim that Wrongs Require Victims:

Our choice cannot be wrong if we know that it will be worse for no one. (FG, p. 118)
This claim follows on unrestricted preference utilitarianism, hence on the unrestricted Impersonal Principles of Beneficence. For if all that matters morally is the maximization of beneficence and we do maximize beneficence, our choice cannot be wrong, by definition. But we have claimed that a choice is wrong if it is irrational, i.e., if it satisfies preferences that violate the canons of reason. Therefore on our system a maximally beneficent but irrational choice is wrong. The rational approach to beneficence seems to be justified as a derivation from the nature of morality itself, morality being a branch of rational decision-making. But can we persuade a utilitarian that the claim that Wrongs Require Victims should be rejected, other than by arguing that on our own justifiable system it is rejected? If we do so we will be forcing him to admit that his own system is insufficient to determine the morality of choices.

The claim that Wrongs Require Victims, then, can be interpreted in two ways, that is, in both a narrow sense and a wide sense. In the narrow sense it coincides with the claim that our choice cannot be wrong if it would not be worse for those people it actually affects than anything else we could have done for them. We just saw that if we take the Paradox of Future Generations seriously we must deny this claim. So we must assume that the claim that Wrongs Require Victims is meant to apply in the wide sense: our choice cannot be wrong if we know it will be worse for none of the people we do realize than for others we might have realized. But notice that if we agree that a choice cannot be wrong if it is worse for no one in the wide sense we mean that it cannot be wrong if it satisfies
actual preferences maximally relative to the other choices we might have made. But in that case there is no objection to a choice like the Risky Policy, which is maximally beneficent and still violates our sense of morality. But we can assume that most people even utilitarians -- do find something objectionable about such choices. If they do, they must agree that we should reject the claim that Wrongs Require Victims. (In that case they should also agree that unrestricted preference utilitarianism is insufficient to satisfy all our intuitions about morality.)

But even if we reject the claim that Wrongs Require Victims, we must still contend with Parfit's question. "Can it be wrong to harm others, when we know...that the people harmed would not regret what we are doing?" Parfit believes that if we respect the autonomy of people we will respect their decisions as to whether or not a choice that affects them is desirable. Suppose we choose the Risky Policy. Then the RP-people will have lives worth living which they would not otherwise have had. Thus it is hard to believe that they would have any regrets over being realized, even given the nuclear disaster that befalls them as a condition of existence. If they were to choose, they would prefer the Risky Policy.

Suppose, then, that we do accept the notion that future people have the right to decide for themselves what they would accept in the way of harms imposed upon them as a condition of existence. We shall determine the morality of the choice that causes them to exist by whether or not they would accept the harms imposed on them.

First we must dispose of the notion that we should count existence itself as an extra benefit over and above
the quality of life future people have. As we argued in
Chapter I, this "extra" benefit has no attributes other
than those of quality of life; therefore the question
reduces to whether future people would prefer their
quality of life, given the benefits and harms they would
receive. (Cf. p.4)

Now we must ask how we might go about deriving the
morality of a choice from the preferences of future people
for existence. Are we to consult future people whose very
existence weighs in the balance? In this case our stan-
dard of morality is absurd. For future people are not
actual people but non-existent beings who have no autonomy
and no preference for existence and are not harmed by not
being realized, since non-existence is not a harm. It is
not possible to base our decisions about the morality of
our choices that will affect future people on the presumed
preference of non-existent beings for existence over
non-existence.

The following diagram may help us to visualize the
argument.

![Diagram](image)

Position 1 is the actual present. Position 2 represents
the present position of future people. Position 3

represents the actual future if we choose $P$. Position $3_q$ represents the actual future if we choose $Q$.

Parfit may be asking us to address the indeterminate "people" in position 2 as autonomous beings and ask them whether or not they would prefer to exist if they were to enjoy all the benefits and endure all the harms of position 3 as either as $3_p$-people or as $3_q$-people. The morality of the choice that realizes them will be determined by their choice. Parfit might be assuming that they would prefer 3 to 2 and $3_p$ to $3_q$. My reply is that there is literally no one in position 2 either to be benefited or to have any preferences. So we cannot determine the moral choice by "their" preferences. The opportunity to exist is not a club we can hold over the heads of non-existent people to persuade them to accede to harms we wish to impose upon them. There is no one there to intimidate. Rather, we must judge our choices by their effects when they actually happen, i.e., when people actually exist.

So perhaps we are expected to determine the morality of a choice that affects future people, not by asking them in advance whether they would accept existence under our terms, but by asking them after they are realized whether our action towards them was moral or right. Perhaps Parfit really means us to consult the actual people in position 3. But then we might very well discover both that these people prefer existence to non-existence and that they consider the choice which realized them immoral or wrong. We cannot be sure that they would prefer their existence over that of other people in the abstract or
wide sense, i.e., believe that it was morally better to realize themselves than to realize others. The actual preference of the people in position 3 for existence is not enough to make the choice that realized them right. Indeed if these people accepted the criterion of morality we have developed they would agree that rationally it is better to realize the Safe Policy people rather than themselves, the Risky Policy people. If we choose the Risky Policy then in the morally relevant circumstances -- i.e., when they actually exist, the RP-people will be subjected to consequences to which we who chose to realize them would not subject ourselves. But if we do not choose the Risky Policy, these people never exist. Hence they are not harmed.

Parfit also implicitly acknowledges that the preference of people who presently exist for existence, now that they have been realized, is not a sufficient argument for the morality of the choice that caused them to exist. He adduces the case of a man actually born to a 14-year-old mother. Parfit's claim is that this man's rights were not violated because of his bad early start. (We disagree on this point: from our point of view they may have been.\(^{24}\)) But--and this is our point--he also says we can agree with a politician who suggested that "it would have been better if the man's mother had waited several years before having children," with its implication that "it would have been better if he had never been born." But if we do accept this implication, we are saying that the preferences of existing people for existence do not constitute a sufficient standard for determining the morality of the choices that brought them into existence. If Parfit
himself agrees with this implication, we can expect him to agree with us that people's preferences for existence do not constitute an argument against the rights approach, where rights are taken to be rational restraints on beneficent action.

But these arguments do not really answer Parfit's objection. For all they do is to repeat the claim that maximal beneficence is an insufficient notion of morality. The problem is not solved so long as we have not answered the questions, Why might a future person not simply waive the right to rational treatment, and what is it exactly he would be waiving?

To take the first thing first: we cannot allow the possibility of waiving rights. In order for rights to be a meaningful concept they must always hold. If they hold for whoever lives, then the fact that one waives them makes no difference. A choice is still irrational even if its irrational effects are discounted by those who are the recipients of them. Here we have to distinguish between what a person in the final position would actually prefer and what he would rationally prefer. In order for a choice to be rational it must conform to all three criteria of practical reason. Therefore, even if an affected person would actually prefer existence even though his rights were violated, he would not necessarily rationally prefer his own existence.

This argument may seem at first glance to solve our problem, but actually it does not. The real difficulty with the rights approach where a person's rights are defined by the three postulates of practical reason lies in the nature of the right to rational treatment itself.
The problem is this. Our postulates of practical reason require us to place ourselves in the final positions of our own choices in order to determine the morality of the choice. Therefore we place ourselves in the position of actually existing future people. We discount existence as an extra benefit; but we determine the morality of the choice by what these future people would rationally accept. What would they accept? Like us, they have certain givens attached to their existence, givens which are different from our own. One of these is their quality of life. How are we to say that these people would not accept for themselves and others and would not prescribe existence such as they are given, or that we would not accept existence on these terms if we were in their place, which is the same thing? It seems that in order for future people to make such a judgment they must have lives not worth living. And in that case they themselves would accept the claim that Wrongs Require Victims, and they would say that a decision no one regrets cannot be bad. This problem, I think, is at the heart of Parfit's objection; and it is a very serious one for the rational approach to beneficence. It indicates that placing rational conditions on preference satisfaction does nothing to solve the problems of beneficence where the affected people by whose reactions we are to judge the morality of our choice exist in circumstances different from our own. How can we determine that they would say that a choice treats them irrationally, or that we would make such a determination in their place?

In order to give a satisfactory answer to this question, let us actually put ourselves in the place of
future people. What do we now know about "ourselves," i.e., what do we know about the future people whose places we have assumed? (1) First, it seems fair to attribute to the people affected by our choices the same type of rationality we have ourselves. Therefore, we can assume that they will employ the same principles of practical reason we would employ. But they are not we, as we exist in our own lives. They will, for example, have a certain quality of life, which will often be different from our own quality of life. If we have projected ourselves into their position and assumed their characteristics, how can we say that this quality of life is unacceptable, i.e., is such that we would not prescribe it for everyone in our own position? (2) It seems that the people in the final position must be capable of being aware of the options. That is, they have to be able to compare their own quality of life with all alternatives available at the time the choice was made. This means they must have the same principles we have and they must employ them in the same way; and, in addition, they must have the same knowledge of relevant empirical factors we have. That is, they must be able to judge rationally that another final position is or is not better than theirs with respect to a principle like the principle of beneficence and with respect to the facts. But this still does not give us a complete enough picture of the people affected by our choices to determine what they would consider rational. In order to know this we have to assume that (3) the affected people are like us in all relevant respects, except where otherwise specified. Therefore they will judge like us. If we would like a certain quality of life, they would like it. If we
would think a certain quality of life bad, they would think it bad. (We can make allowances for empirical differences in preference, where they are specified, without destroying the force of this requirement.)

To make these assumptions is to introduce a *similarity postulate* into our calculations. This postulate constitutes a fourth postulate of practical reason:

iv. The similarity postulate

Those people who are affected by our choices have the same principles of reason, the same values, the same knowledge of the options, and the same preferences we have, except where otherwise specified.

Thus when we place ourselves in the final positions of our choices we are asking, Given that "I"--the affected person--have the same likes and dislikes, the same values, the same knowledge of the alternatives, and the same power of reasoning as "I"--the moral agent--what do I think about the acceptability of the position in which I am placed?

The introduction of a fourth postulate of practical reason into our system is an admission that the earlier characterization of our position is insufficient. But it allows an answer to Parfit's question about the success of the rights approach to beneficence. Our claim is now that with the similarity postulate we can show that people in the future would not accept as morally rational any final position we ourselves would not accept as morally rational unless we have reason to believe they are relevantly different from us in reasoning, values or preferences even though they would *actually* prefer their position to nonexistence.
There are still, however, certain problems intrinsic to the rational approach to morality.

In most cases it is clear whether or not a choice violates the postulates of practical reason. The Risky Policy certainly does. In other cases it is not clear. And in these cases practical reason cannot tell us what its own limits are. We cannot tell what practical reason requires in all cases simply by consulting the notion of treating people as we would treat ourselves in similar circumstances, or by consulting the notion of treating equal interests equally, or by consulting the notion of the universalizability of prescriptions for all moral agents in similar circumstances, assuming in each case the similarity postulate. We must always employ fallible empirical tests of these notions. This is where human judgment enters into morality. This is a true limitation of the system. But so far as I know, every moral system requires us to make our own judgments as to when its precepts are violated. It is doubtful that any moral system can make our moral judgments automatic.

This problem of the indeterminacy of the empirical requirements of the theory is shared by the rational approach with other moral theories. One problem which follows on these limitations is specific to our theory.

Suppose that if I pollute your sky with \( x \) units of sulphur I do not do to you what I would not do to myself. So my pollution is a mere cost and is justified in return for economic benefits. (We choose the more beneficent of two good alternatives.) But if I pollute your sky to \( x+1 \) units of sulphur I treat you badly and nothing can justify my choice except my having only worse alternatives. (We
choose the lesser of two evils.) Such a case tests the limits of tolerance of practical reason. This is where the theory is least decisive. Yet the consequences are great and our choice seemingly arbitrary. Cases like this present hard choices under the theory. We must make the decision as best we can on empirical grounds. So long as the choice does not in our best estimation violate practical reason--no matter how much it costs people or how unequal its effects on people--we should choose the maximally beneficial alternative. Otherwise we should choose the least evil alternative.

These are real limitations of the rational approach to beneficence. They should be borne in mind. But they do not seem to argue radically against acceptance of the theory. It appears, therefore, that neither the arguments against the rights approach in general nor the objections to our theory in particular weigh strongly against the theory, once we add the similarity postulate. And we have claimed that a theory should be tested by its intuitively satisfactory results. If our theory is sound, it should enable us to solve the problems of morality towards future people which merely maximizing actual preference satisfaction or beneficence does not solve.
VI. Resolution of the Paradox of Future Generations

We are now ready to apply the model for rational decision-making in cases of beneficence to those cases which we could not satisfactorily resolve merely by maximizing actual beneficence according to the unrestricted Impersonal Principles of Beneficence. This will require evaluating choices not only in terms of their specific end—beneficence—but also in terms of their form, i.e., their adherence to the dictates of practical reason.

We can begin with Same Number Choices.

i. Same Number Choices

When two alternatives are both intrinsically good, i.e., meet the criteria of practical rationality, we should choose the more beneficent alternative. This explains why the Impersonal Principles of Beneficence give satisfactory results in many cases, like the Depletion case and the case of the 14-Year-Old Girl, where the latter is not considered as a Different Number Choice. These cases were also satisfactorily decidable under the Impersonal Principles of Beneficence interpreted as preference utilitarian principles without rational constraints. The difficulties arise when the notion of unrestrained beneficence is insufficient to capture our intuitions of morality. These cases can be solved only by applying the Impersonal Principles of Beneficence rationally.
When one of two alternatives is intrinsically bad, i.e., violates the criteria of practical rationality, and the other is intrinsically good, we should choose the good alternative. This allows us to answer the problem posed for morality by the Risky Policy. In the case of the Risky Policy we must project ourselves into each affected generation. In each generation we would get the benefits of our quality of life and in one generation we would get those benefits plus the nuclear disaster. Clearly we would not agree to accept the nuclear disaster if it were to fall tomorrow on ourselves, our friends, and our relations. And we have no reason to think the Risky Policy people would like to have such a disaster visited upon them. The nuclear disaster is an intrinsically bad effect of the Risky Policy. The presence of any intrinsically bad effect is sufficient to make a choice intrinsically bad. The Risky Policy cannot meet our standard of morality, but the Safe Policy does. Only the latter is intrinsically good. We should choose it.

Where two choices are both intrinsically bad, we should choose the lesser of two evils. Suppose that instead of the Risky Policy and the Safe Policy (RP and SP) we had a choice between the Risky Policy as it is given and the Safe Policy* (SP*). On SP* there is a nuclear disaster, just as on RP, only even more people are killed or injured by radiation leakage.26 We have said that RP is intrinsically bad. Here SP is worse. We should choose RP. Given two immoral alternatives, we should choose the less immoral.

Finally, because ours is a rational system and therefore does not judge the morality of choices solely by
their utility information, we may evaluate different sets of choices differently even though they fit the same positive utility schema. Therefore, where alternative choices are equal in beneficence, but one builds in an offense to practical reason and the other does not, we should choose the alternative that does not offend practical reason.

We can see this clearly if we revise the Risky Policy case once more. Suppose we keep the facts as they were except that we allow the outcome in terms of beneficence on the Safe Policy to be equal to that on the Risky Policy once the effects of the nuclear disaster on the latter policy are figured in. This changes SF to SP** but leaves RP as it was. Now we have two alternatives with outcomes equal in beneficence. For a utilitarian the policies should be morally indistinguishable. But the Risky Policy still offends our notion of human dignity. It still seems intrinsically bad. The approach is superior to the unrestrained utilitarian one because it recognizes that it is not possible to compensate in terms of good preference satisfaction for offenses against practical reason itself. Such offenses are intrinsically bad. They are therefore prior in the moral scale to utility.

We can make the same point even more clearly in the Exploitation Policy case.

In the Exploitation Policy case everything is equal except that some of the affected people live in the future. Intuitively we can see that where everything is equal the notion of treating all people with equal respect must require assuring them equal outcomes. To act otherwise is to offend the notion of respect for persons, or
treat the equal interests of people equally, as if we were to occupy only one of the final positions of our choice. And we have said that a person's living in the future is not a moral reason to treat him as not worthy of equal respect. Thus, if we have equal respect for all people and all things are equal except that some of the people affected by our choice live in the future, we should choose an alternative with equal outcomes over one with unequal outcomes. On the rational approach we should choose the Equalization Policy over the Exploitation Policy. A policy with unequal outcomes where all things are equal except that some affected people live in the future is immoral because of its form. It builds in an offense to human respect.

The rational approach thus allows us to resolve and explain problems posed for morality by Same Number Choices like Depletion, the Exploitation Policy, and the Risky Policy. We have still to show that it can help us with the problem of Different Number Choices.

ii. Different Number Choices

The subordination of the Impersonal Total Principle of Beneficence to the requirements of practical reason enables us to avoid the Repugnant Conclusion. Thus it gives us a solution to the counter-intuitive implications of Different Number Choices for the total principle of beneficence.

The argument that the rational application of the principle of total beneficence allows us to reach the Repugnant Conclusion might go as follows. Under a certain policy the second generation, B, will have lives of a
quality almost as good as that of the lives of people in
the generation that initiates the policy, A. We can
assume that the B-people will not have to bear burdens the
A-people would not put on themselves. The C-people will
have lives almost as good as the lives of the B-people:
they will not have to bear burdens which the B-people
would not be willing to bear themselves. And so on to Z.
Therefore, even rationally applied, the Impersonal Total
Principle of Beneficence leads to the Repugnant Conclu-
sion.

Our approach does not imply the Repugnant Conclusion
because for a choice to be moral by our standard of
morality none of its effects can violate the dictates of
practical reason. Each final position must be one the
agent, namely A, would be willing to accept for himself
if he were to bear the effects of his own choice. We
assume that we are the A-people and the policy put into
effect remains unchanged. While we would presumably be
willing to accept the drop in our quality of life if we
lived in B and perhaps the drop if we lived in C; at some
point the quality of life would drop below that we would
accept for ourselves. This point is well short of Z, or
there would be nothing repugnant about the Repugnant
Conclusion. And notice that we are not holding existence
itself over the heads of future people and asking them
what they would accept in order to be allowed to live.
(This would justify any choice of ours that gave people
lives worth living.) We are asking what we ourselves, as
actually existing people in the final positions of our
choice, consider a quality of life that does not offend
the criteria of practical reason. The only quality of
life that can meet this standard is one we would accept for ourselves as a result of our own policies given our existence and given our likings. At the point the quality of life we accord others drops below this level our policy becomes intrinsically immoral. The Repugnant Conclusion policy is also immoral on other grounds since it is grossly unequal in its outcomes without morally relevant justification. Thus a choice which entails the Repugnant Conclusion can never be moral under our standard of rational morality.

At this point we should notice an important implication of adopting the Impersonal Total Principle of Beneficence. Z is worse than A, and any policy which leads to Z is intrinsically bad. But C is actually better than A. If we can maximize beneficence by causing greater numbers of people to exist with lives worse than the prior average quality we should do so, just so long as we do not give other people lives we ourselves would not be willing to lead. B has more total beneficence than A, without violating the dictates of practical reason, and C has more total beneficence than B. Therefore, we should choose B over A and C over B. If intuitively we object to maximizing beneficence by bringing people into the world with lives of lower than average quality, we should reject the Impersonal Total Principle of Beneficence. The rational application of the total principles limits its implications; it does not change their thrust.

The rational application of the Impersonal Average Principle of Beneficence, however, does not solve the problem of Different Number Choices for that principle. This is because, even when it is rationally applied, that
principle still implies both the Draconian Conclusion and the prisoners' dilemma caused by having too many happy children. Surely we cannot call it an infringement of anyone's right to rational treatment to require that only children who will maintain or increase the average quality of life be conceived or to require that happy children be conceived. So we still need an answer to the problems of the average principle.

There are two approaches we might take.

(1) We might argue that we do not have a duty to maximize average beneficence on Different Number Choices. If we could plausibly make this argument it would solve two problems. It would allow the conception of children with lives of less than average quality where the alternative was to have no child at all. Thus it would solve the problem of the Draconian Conclusion. And it would release people from the obligation to have happy children, thereby solving the prisoners' dilemma which arises when too many happy children are born. Unfortunately this solution to the problem of Different Number Choices for an average principle of beneficence fails.

If we argue that, so long as no one's right to rational treatment is infringed, it is moral to conceive children whose lives will be of less than average quality, the average principle becomes indistinguishable from the total principle. For if we morally could have children with lives of less than average quality, it seems we should have such children. But then the average versus total issue ceases to matter in the one case where it is important: Different Number Choices. And if we claim that there is no duty to maximize either average or total
beneficence on Different Number Choices our principles of beneficence cease to be sufficient principles. It is not clear how they could be supplemented.

Similarly, suppose we say that where a person has a choice of having no child or having one with a life of better than average quality, i.e., a Happy Child, she may have such a child, but she is not obligated to. (This parallels Parfit's Revised Maximizing Principle.) There are several difficulties with this claim. First, it offers no moral reason to choose one alternative over another. And second, again it does not allow one to distinguish an average principle of beneficence from a total one.

Moreover, if we deny that there is an obligation to have a Happy Child, we will still probably want to claim that there is an obligation not to have a Wretched Child, one whose life will not be worth living. To deny the obligation to have a Happy Child while confirming the obligation not to have a Wretched Child is to endorse the Asymmetry. (C. FG, p.148) But if we claim that having a Wretched Child harms him, it is not clear how we can claim that having a Happy Child does not benefit him. And, furthermore, we wish to retain the idea that causing to exist can benefit, for this is the justification for the principle of beneficence where future people are concerned. It seems to me that we must reject the Asymmetry and instead accept the conclusion that an average principle of beneficence (as well as a total one) requires the conception of a Happy Child.
We are left with no solution to either the Draconian Conclusion or the prisoners' dilemma caused by having happy children.

(2) Alternatively, we can begin by accepting the Draconian Conclusion. If we do this, we must agree that it is immoral to conceive a child whose life will drag down the average quality of life, but it is moral--and even obligatory--to conceive children whose lives will raise or maintain the average quality of life. This alternative is not, however, either as elitist or as grim as it looks. And I shall argue that we should accept it.

It is a mistake to assume that all that matters morally is the quality of life of the future person whose conception is under consideration. For, as we saw when we explicated the Impersonal Principles of Beneficence, what matters in terms of beneficence is the effect of the lives of future people on the average (or total) quality of life of everyone who lives. Prior to this, what matters morally is the conformity of a choice to practical reason. Let us note at the outset that on our average principle of beneficence as modified by reason, it is immoral to conceive a child whose right to a life one would accept for oneself is not met or one whose existence will violate practical reason with respect to others (by, for example, seriously impairing the health of the mother). In this way we respect the intuitive notion that the existence of future generations should not impose "too great a sacrifice," or hardship, on anyone. (Cf. Parfit's Revised Maximizing Principle, p.56 above.) But assuming this condition fulfilled, what matters is the effect of the existence of a future person on everyone who lives,
and this means that future people must be considered not only as vessels of utility and consumers of the utility of others, but also as producers of utility. It is this last crucial consideration which has been overlooked.

A person's own life might be of less than average quality—call him the Pretty Happy Person—but the effect of his existence may nevertheless be to maintain or increase average happiness. Suppose, for example, the Pretty Happy Person grows up to be an arthritic garbage collector. He is in pain; he earns very little money; and, although he has self-respect, he has no great respect in the community. He also has a happy marriage and one of the best voices in the church choir. His life is certainly worth living but of less than average quality; it is not so bad, however, as to violate our rational obligation to him in conceiving him.

The decisive moral factor in deciding whether or not to conceive someone who predictably will meet this profile is whether he will contribute to maintaining the average quality of life in the society. As a garbage collector our Pretty Happy Person will be one of the cogs on the wheel of civilization. And as a mainstay of the church choir he will give happiness and hope to others. Without people whose contribution helps maintain or increase the average quality of life of all, even though their own lives are of less than average quality, the quality of life of all would drop. It is moral under the average principle of beneficence to conceive the Pretty Happy Person. And it is moral even though we retain the Draconian Conclusion. What is not moral is to conceive children who will have a negative effect on the average
welfare of all, or who will be given lives a rational person would not accept, or whose existence will violate practical reason with respect to others.\textsuperscript{28}

Looked at in this way, the Draconian Conclusion is hardly draconian. It merely says that we should conceive only those children who will contribute positively to maintaining the average quality of life in their society. I will hazard the guess that everyone intuitively believes this claim anyway. The alternative — under a total principle of beneficence — is to claim that it is moral, and even obligatory, to realize all people who will make any positive contribution to society in terms of the \textit{quantity} of worthwhile life lived, i.e., who do not actually reduce the total quantity of worthwhile life lived. And I hazard the guess that no one believes this.

But now we come to the problem of the Happy Child, compounded by that of the Pretty Happy Person. For the prisoners' dilemma still remains. If I have the maximum number of happy children—happy in the sense that they make positive contributions to the average quality in life in their society—and everyone else does too, the quality of life of everyone may fall, simply through their consumption of resources. We should have all and only those children who will contribute most to increasing the average quality of life of all. But who is to decide who those children are?

We might try claiming that the mere addition of extra lives that are either Happy or Pretty Happy cannot make a choice worse than the alternative, even if it does not make it better. Indeed, this seems to be an implication
of our position. But we must strenuously deny this implication, for, as I shall show below when we come to the Mere Addition Paradox, it destroys the transitivity of the principle of beneficence. And, in any event, it does not solve the prisoner's dilemma.

We must agree that as a society we have a moral obligation to have just those children whose existence will contribute to maximizing the quality of life of all. And we must understand that we are not therefore required to have only the happiest possible children. The problem is, of course, to determine just which children should be born. Possibly a principle of choice could be developed, but I doubt it. I think, rather, that decisions about whether or not to conceive children must remain as they are now, ad hoc and personal. But I also think that we really do make such decisions according to our own best judgment that our potential child will make a positive contribution to the world (when we think about the morality of our having children at all). And we justify our not having children by the argument that we ourselves will contribute more to average beneficence by not having children than by having them. And, if this is the case, we are intuitively following a principle of average beneficence.

Notice that the obligation to maximize average beneficence rationally prescribes a certain type of population policy. Governments and other institutions concerned with morality may and should develop incentives for people to conceive greater or smaller numbers of children, according to the needs of society, with respect to maximizing the quality of life of all, so long as no
one's rights are violated. The idea of a moral population policy is that people become persuaded that their having any children or having more or less than a certain number of children will probably decrease the quality of life of all, or conversely that their having (only) a certain number of children will increase the quality of life of all. But no one may be treated in any intrinsically bad way in order to achieve this goal.

Given our understanding of the Impersonal Average Principle of Beneficence and of the Impersonal Total Principle of Beneficence, as applied within the confines of practical reason, we can refine our answers to the cases of the 14-Year-Old Girl and the Handicapped Child, conceived as Different Number Choices.

In both cases the potential mother should not have the worse-off child, regardless of whether or not she can later have another child, if the worse-off child is so bad off that, considering his position alone, a rational person would refuse to accept that position for herself, or if having him would ruin her own life. If no one's right to rational treatment will be violated by the conception of a child, however, whether or not the potential mother should conceive him depends on (a) whether she can have a better-off child instead and (b) whether her having either this child or another will increase or decrease the average beneficence in the world.

If she can, the mother should have the better-off child rather than the worse-off child. But there is always the further chance that a woman might have no child at all. Suppose that her options are to have the worse-off child or none. She should have this child only if his
existence would contribute to maintaining or increasing the average quality of life in the world and would make a greater contribution to average beneficence than she herself would make by having no children. (We are assuming now that having him is not intrinsically immoral.) Suppose that she could have the worse-off child, or a better-off child whose life will still not be of average quality, i.e., a Pretty Happy Child, or no child at all. In this case she should have the better-off child if having him will maintain the average quality of life in the world and will contribute more to average beneficence than she herself would contribute if she remained childless. Or suppose that she could conceive the worse-off child, a Happy Child, or no child. Again, she should have the Happy Child if having him will maintain the average quality of life in the world at a higher level than if she contributed to average beneficence in some way other than through motherhood.

In all these cases the potential mother must decide whether or not her having a child will maximally benefit the world under an average principle of beneficence, given her options. She should have all and only those children whose existence will allow the average quality of life to be maximized with respect to her actual alternatives. If repeated decisions of mothers to have children who meet this criterion lead to over-population problems, it is the province of the government, the church, and whatever other moral institutions there are in society to persuade some of the people that their having happy children, or pretty happy children, or more than a certain number of children will contribute to a lower over-all quality of life. In
this way the prisoner's dilemma generated by the mandate
to have happy and pretty happy children is at least
partially correctable.

Thus we can accept the Draconian Conclusion and still
be satisfied with the answers the Impersonal Average Prin-
ciple of Beneficence gives in personal reproductive cases
and in the policy cases which they instantiate. But there
is still a grave problem with the Impersonal Average
Principle of Beneficence.

In policy cases and in cases where we must decide
whether or not to conceive a child, there is a question
whether we should realize future generations at all. The
argument that we should not bring future generations into
being if doing so will lower the average quality of life
of all goes like this. An average principle of bene-
ficence simply directs us to maximize beneficence. As far
as the principle is concerned, it does not matter whether
we do so by realizing future people or not. If we do not
realize future people there will be more beneficence for
us to consume ourselves. Therefore a principle of average
beneficence directs us not to realize future generations
where we can have a policy of consuming like mad our-
selves, since if we share what we have with those who live
in the future average beneficence will be lower. Why
doesn't the Draconian Conclusion apply with a vengeance
here? Or, if we want to look at this issue more posi-
tively, we can ask, What is wrong with a self-regarding
life-style that precludes children? We can call this the
Challenge from Narcissism.

L. W. Summers has developed a version of this chal-
lenge. He argues that if we were presented with an extra
amount of utility of which to dispose and could use it either to benefit ourselves or to create extra happy people, an average principle would tell us to benefit ourselves. Jefferson McMahan has responded to this challenge by pointing out that the future individuals whom Sumners takes to be deprived do not exist to be discriminated against. Thus no one is harmed by "favoritism" to present people. While McMahan is correct, I think Sumner's challenge goes deeper than this and requires a further response.

Suppose that the quality of life in our society of 100 people (A) is such that we enjoy an average beneficence quotient of 50. We are presented with sufficient utility to create 100 extra people with an average beneficence quotient of 60. Alternatively, we can consume the extra utility ourselves. If we create extra people, the average quality of life of everyone who exists will rise to 55. But if we do not create extra people, the quality of life of everyone who exists (ourselves) will rise to 110. The average principle of beneficence tells us we should consume the extra utility ourselves.

But now let us consider three different possible outcomes. We could use the extra utility ourselves, thereby turning our society from A into A*. We could use it to bring into existence a new society in a distant land with which we ourselves would never have contact (A+A'). Or we could save it to benefit our children (A+B). We can represent these outcomes by the following graph:
In terms of total utility all three outcomes are equal. Therefore a total principle has no particular advantage in deciding this question. But on an average principle the best choice is apparently clear: we should choose A*.

But now consider: it may well be that we would not see any moral virtue in using our extra utility merely in order to add extra worthwhile lives. This would be the situation, however, if we chose A+A'. So if there is no moral virtue in the mere addition of people with worthwhile lives, it seems that our average principle gives us the right answer in directing us to choose A* over A+A'.

But how about A+B? If this is one of our alternatives, it immediately becomes apparent why the challenge is so powerful: if we choose A* over A+B, we are choosing a higher quality of life for ourselves at the cost of prohibiting future people from being born. And it is because this choice is intuitively wrong that we find something wrong with the average principle: we value our perpetuation of our kind more than we value extra benefits in our own quality of life.31 (Assuming we have lives we consider worthwhile. There is surely a point below which we think life so grim that we would not want to bring children into existence. In that case we would prefer to ameliorate our own lives. This would be the case if realizing future people would cause people to have an average quality of life we could not rationally prescribe.) That is, the decision to have children generally carries with it extra utility over and above the utility of more consumption on our own part, utility that more than compensates us in our own minds for foregoing the
luxuries we could enjoy if we did not have children. If this is so, the decision to have children is justified from a total utilitarian point of view; but it still appears not to be justified from an average utilitarian point of view in cases like that of A and B. It is, however, justified from the average utilitarian standpoint. For if the decision to have children gives us extra utility over and above the utility offered by the choice of further consumption for ourselves the decision to have children is justified as an increase in our own average utility, regardless of what happens to that utility once our children are conceived. Then if we value having children more than increasing our own quality of life, given worthwhile lives of our own in A, we can assume the existence of B. From this point on we are no longer talking about maximizing our own average utility but about maximizing the average utility of A and B together.

This practical assumption is reinforced when we recall that our children will be producers of utility as well as consumers. It may well be that if we are all Narcissists we will consume our bien, our own good; and without replenishing benefits at their source, i.e., without creating new producers of beneficence, we may well cause our own quality of life to fall. If this happens, by choosing to have no children we will have chosen a policy of lesser average beneficence than if we had chosen to have children. This is not a necessary result of Narcissism, presumably; but it seems a likely one. Narcissism, like the opposite policy, the unrestrained
conception of Happy Children, implies a prisoner's dilemma.

The solution, therefore, to the Challenge from Narcissism lies in the recognition that as a people we place reproduction above extra benefits for ourselves, so long as all concerned have lives whose conditions do not offend practical reason, and that placing reproduction above beneficence to ourselves can, and presumably does, maximize beneficence to all concerned. (An increase in general beneficence through the contribution of the newcomers is not as likely to be the case, however, with mere addition, since in that case the addition of the newcomers does nothing to reinforce the maintenance of average beneficence in the original population. Nor do we presumably derive extract benefits from realizing them. The Impersonal Principle of Beneficence may condemn the mere addition of extra people at the same time that it commends the procreation of an identical number of people in the original population.)

We can illustrate the difference the assumption of the existence of future generations makes by the following scenario. To keep the case pure we assume that there is no gain in total beneficence from procreation.

On the new scenario we have no equivalent for \( A + A' \), because we assume that we would not place the mere addition of extra worthwhile lives above our own self-interest; and, in any event, this is not a realistic alternative (since we are not God). \( A^* \) changes to \( A^* + B^* \). \( A + B \) stays the same. And we can add a third option, \( A'' + B'' \).
On A²+B² we take all the extra utility for ourselves in A and we reserve nothing for future generations in B*. We guarantee the B*-people worthless lives; and since we have posited that they will exist, we violate the most fundamental considerations of practical reason. Moreover, this outcome has exactly the same level of average beneficence as A+B and A"+B". Therefore, it cannot be preferred over these other outcomes on grounds of average beneficence. A+B, on the other hand, guarantees everyone who exists a quality of life that is high and almost the same between generations. It is a far better choice. An even better outcome, however, is the impartial one where A takes 5 beneficence points for each of its own members, thereby guaranteeing an average quality of life in A" of 55, and reserves the rest for B", guaranteeing an average quality of life in B" of 55 also. (This assumes that A+B and A"+B" are both options and there are no other morally relevant factors to distort the picture. The picture would be still different if we could assume that B would have a certain level of quality of life regardless of what we in A did with our utility bonus or that the mere existence of B would increase beneficence in A. But the moral directive would be the same: to share our good fortune, maximizing beneficence in accordance with practical reason among generations.)

Thus, if we assume the existence of B, the moral ranking of our possible choices changes. Given what we can take to be our real alternatives, A"+B" is preferable to A+B, which in turn is preferable to A²+B². Therefore, if we understand the average principle of beneficence as applying in the real world, where the perpetuation of
mankind is valued above increases in the quality of already worthwhile lives, we can meet the Challenge from Narcissism.

The Impersonal Average Principle of Beneficence directs us not to make any choice which will lower the average quality of life of those who actually exist. But where policies that affect the future are concerned, it refers to the average quality of life of all. If we assume the existence of future people, the moral choice will be that which counts them as just as worthy of respect as ourselves. In that case we must maximize the average beneficence of present and future actual people together. This does not return us to the total principle of beneficence, for even though we endorse the existence of future generations despite a drop in quality of life we endorse the existence of future generations at the highest possible average quality of life not at the lowest possible average quality of life consistent with both practical reason and maximal total beneficence.

What has happened to the Draconian Conclusion? It now appears that the objective of the average principle of beneficence is not to prevent any fall in the average quality of life but rather to maximize beneficence impartially over all those generations we can expect to exist and to be affected by our policy. And we can expect future generations to exist if it is not irrational to realize them. Thus it is not immoral for a policy to lower the average quality of life of those who presently exist if in doing so it maximizes the average quality of life of all those who can be expected to live and be affected by it. This requires a reinterpretation of the
phrase "the average quality of life." For up until now we have understood by this phrase the average quality of life of those who previously existed. We have been requiring that the existence of future generations not lower the present average quality of life. Now we understand by this phrase the average quality of life of all those affected by our choice. We require that our choices maximize the average quality of life of all affected people.

But now suppose that we were the A*-people in the previous scenario. Why doesn't this view require us to turn ourselves into the A"-people in order to provide for the B"-people who would otherwise start at 0? This would indeed be the case if things were as they are depicted on that scenario. But in most cases the quantity of available beneficence is not static. The entire well-being of future generations does not depend on our sharing what we already have with them impartially over all affected generations. The well-being of both present and future generations will vary with our policies. We can make the determination of how each policy will affect the well-being of all by studying the expected consequences of the policy and the manner of attaining them. We are then required to choose the maximally beneficent policy which does not violate the right of all affected people to be treated with equal respect. It is not the case, therefore, that if our policy affects ourselves (A) and our children (B) we need share our present resources equally with B (except where present resources are all the resources there are and B will need them equally as much as A). Rather, we must choose that policy which will be
maximally beneficent across generations, so long as it does not violate the requirements of practical reason.

But there is still a problem caused by the requirement that our policies assume the existence of B. B, of course, does not presently exist. I am required to maximize average beneficence in B. But do I maximize average beneficence in B assuming the real projected size of B, or the optimal size of B for the purpose of beneficence, or the size of B as it would be if all and only those lives of equal or better quality than our own were realized?

If I am required to maximize average beneficence in B, I must be required to maximize beneficence at that optimum population level where average beneficence is greatest. In this way I fulfill the requirement of the average principle of beneficence with respect to B.

Here personal and public policy coincide. My obligation as a citizen of my society is to do my part in promoting the optimum inter-generational level of quality of life. This requires that in my capacity as legislator of policies I choose those policies which will be maximally beneficent for all on the average, so long as those policies treat people with equal respect. Where I, as legislator, face a choice among population policies I should choose that policy which will ensure the optimum level of quality of life among affected generations. Such a policy may require that I personally have more or less than a certain number of children. On the individual level, therefore, it is my responsibility to promote that policy. I should have a child or children if and only if
their existence would contribute to maintaining or increasing the optimum population level inter-generationally with respect to the maximum possible average quality of life and no one would be harmed by their existence.

Our assuming the existence of future generations thus requires redefining our duty with respect to beneficence on the Impersonal Average Principle. We redefine it as the maximalization of average beneficence inter-generationally for optimum population levels. It is no doubt no more than an ideal that we will adopt exactly those policies and conceive exactly those children who will contribute best to realizing this goal. But it is an ideal we should espouse. With it the Draconian Conclusion ceases to be draconian.

Let us now summarize.

Both the Impersonal Total Principle of Beneficence and the Impersonal Average Principle of Beneficence when applied rationally offer solutions to the problems posed for morality by the Risky Policy and the Exploitation Policy. And the Impersonal Total Principle of Beneficence, rationally applied, stops short of implying the Repugnant Conclusion. This principle thus emerges as satisfactory so long as we are willing to accept the conclusion that a larger population which is worse off is better than a smaller one which is better off, if the larger population more than makes up in quantity what it loses in quality and no one's right to equal respect is violated in bringing about that population. On the RC scenario the best population for us to seek to bring about is that which has the largest number of people at the lowest quality of life we would accept for ourselves. If
we are not satisfied with this conclusion, we should adopt the Impersonal Average Principle. But the rational application of this principle does not affect the Draconian Conclusion. The solution to the Draconian Conclusion must come from our understanding of average beneficence itself. The Impersonal Average Principle of Beneficence is satisfactory if we believe either that the continuation of mankind is a moral imperative superior to the command to maximize beneficence or that maximal average beneficence in the wide sense will actually be promoted by our having children. If we believe either of these claims, we should assume the existence of future generations. When we do, we must redefine our duty to maximize the average quality of life as a duty to maximize the average quality of life inter-generationally for populations of the optimal size. We do not avert the Draconian Conclusion, but with this interpretation of the requirements of the average principle, the Draconian Conclusion becomes innocuous. The Impersonal Average Principle of Beneficence emerges as intuitively superior to the Impersonal Total Principle of Beneficence. But, whichever of the two Impersonal Principles of Beneficence we choose, we can solve the Paradox of Future Generations. And by applying either principle rationally (and also assuming, in the case of the average principle, that future generations will exist), we can avoid endorsing the intuitively unsatisfactory implications of unrestricted beneficence.

At this point it will be instructive to return to where our path diverged from Parfit's. If our own strategy has been successful, we should be able to show where
Parfit's fails and to solve further problems of beneficence which he professes not to be able to solve.
VII. Parfit's Strategy and the Mere Addition Paradox

Parfit attempts to solve the general problems of beneficence by developing a new principle of beneficence from the old ones, one that does not have the unsatisfactory implications of the old. As we saw, he rejects the Impersonal Average Principle of Beneficence on the grounds that when all else is equal, the mere addition of extra people with lives of below average quality cannot be morally bad if there is no loss to anyone. (Cf. p.159) On this principle the mere addition of such people is bad if their existence lowers the average quality of life of all. (Unless we assume that the extra people are the children of the original people and that their existence should be assumed because procreation is more important than increased beneficence to ourselves. But this is not what is meant by "mere addition.") Parfit also rejects the Wide Person-affecting Total and Average Principles of Beneficence and the Impersonal Total Principle since all imply the Repugnant Conclusion. But he does retain the notion that morality in the area of well-being requires a maximization principle necessary and sufficient to give only satisfactory results. He tries developing a satisfactory principle on the basis of the Impersonal Total Principle of Beneficence. It is important to see what Parfit says, especially as we have endorsed the Impersonal Total Principle as a possibly satisfactory general principle of beneficence.

Parfit begins his search for a satisfactory principle of beneficence by accepting two parts of the Impersonal
Total Principle, namely the views "that it is intrinsically worse if there is more suffering, and that it is intrinsically better if more lives are lived above some Wonderful Level." (FG, p. 157) He adds to the claim that it is intrinsically worse if there is more suffering the claim that it is intrinsically bad if lives are lived below a Restricted Level where life is still worth living but its quality is severely restricted. (FG, pp. 163-164) That is, it would be better if lives below the Restricted Level were never lived. The Restricted Level thus works like our own limitation on causing people to exist whose quality of life we would not accept for ourselves, which we derived from practical reason. But Parfit infers from the claims that it is intrinsically better if lives are lived above the Wonderful Level and intrinsically worse if lives are lived below the Restricted Level the conclusion that "lives between these levels have no intrinsic value." (FG, P. 164) We could not concur in this since on our view all lives have value in terms of beneficence strictly according to the quantity of beneficence they contain.

What is the result of arguing that lives that are neither Wonderful nor Restricted have no intrinsic value? The immediate inference is that the mere addition of extra lives within this intermediate zone cannot create a state of affairs worse than that which previously existed. Parfit argues, however, that the new state of affairs need not be better than the old, on the grounds that "not worse than" is not a transitive relation. (FG, p. 166. I will not discuss Parfit's argument for the claim that "not worse than" is not transitive.) This argument creates, in
effect, an unnamed mere addition claim. I will call this Claim (B):

It is intrinsically better to add lives above a Wonderful Level to society, and it is intrinsically worse to add lives below a Restricted Level, but the mere addition of lives in the intermediate zone cannot create a state of affairs worse than that which previously existed, although the new state of affairs may not be better than the old.

Parfit also believes that, on both (total) utilitarian and egalitarian grounds,

a change in the distribution of welfare whereby a worse-off group gains more than a better-off group loses creates a better state of affairs than that which previously existed.

(Cf. FG, p. 159)

I will call this Claim (C). (The wording of both claims is my own.)

The result of endorsing these claims is that one ends up with no solution to a new paradox Parfit propounds, the Mere Addition Paradox.

The Mere Addition Paradox looks like this. (FG, p.158)
A is a world with a limited population and a high average quality of life. Everyone is equal. B (in my notation $B_1 + B_2$) is a world with twice as many people as A ($A_1$) and an average quality of life that is lower than A’s but still more than half as great. Again everyone is equal. $A^+ (A_1 + A_2)$ is a world identical to A except that it contains an additional population group ($A_2$) with a lower quality of life than that of A ($A_1$), yet a quality of life such that life for the extra people is still well worth living. (We shall say that $A^+$ has the same number of people as B, as it appears to have.) The two population groups are separated from each other by an ocean and have no contact with each other. Hence the extra group affects no one else. Divided B is exactly like B except that it is divided, like $A^+$, into two equal population groups separated from each other by an uncrossed ocean.

The paradox is generated by the following argument. The object is to defend the view that A is better than B. (Parfit gives no explicit grounds for this judgment, but sufficient grounds might be that A is better in average beneficence than B.) Divided B is identical to B in all morally relevant respects. Therefore, by the transitivity of identity, anything that is true of B is true of Divided B.

$A^+$ is generated from A by the mere addition of people with lives worth living. And, by Claim (B), the mere addition of lives worth living cannot make a state of affairs worse than that which previously existed. Therefore, $A^+$ is not worse than A. Moreover, $A^+$ is not worse for the $A_1$-people, who do not know that the $A_2$-people exist; and it is not worse for the $A_2$-people, who have
lives worth living which they would not otherwise have had. Therefore the change from A to A+ is worse for no one. Parfit does not make the point explicit, but the view that A+ is not worse than A is justified not only by Claim (B) but by Paretianism. Paretianism is the view that a state of affairs where no one loses any welfare and someone gains in welfare is better than (Parfit would say "not worse than") the previously existing state of affairs. No one in A1 loses anything by the transition to A+, and those in A2 gain lives worth living. Therefore the change is for the better (or at least not for the worse) on Paretian grounds. It has been argued that Paretianism is a feature of all utilitarian systems, and a vulnerable one. I will neither defend nor attack it here. I merely note that Paretianism supports Parfit's conclusion that A+ is not worse than A.

Parfit now argues that Divided B is better than A+. This is true on grounds of both total and average beneficence. Moreover, we can envisage generating Divided B out of A+. In this case the worse-off people would gain more than the better-off would lose, justifying the move to Divided B on grounds of both total utilitarianism and equality in accordance with Claim (C). But if Divided B is better than A+, and A+ is not worse than A, then Divided B must not be worse than A. And if Divided B is not worse than A, then B, which is identical to Divided B in the relevant moral respects, must not be worse than A. But the original object was to defend the view that B is worse than A. These beliefs are inconsistent. Hence the paradox.
In a true paradox apparently inconsistent premisses are shown to be consistent on a higher level. This is the case with the Paradox of Future Generations. I shall argue, however, that the Mere Addition Paradox is not a true paradox: its premisses really are inconsistent. Thus it will be resolved when we eliminate some of its premisses.

The argument for the paradox implies that we can make no valid moral comparison directly between A+ and B. But surely this conclusion is too strong, for we do have a basis of comparison between A+ and B without involving Divided B in the comparison. We can compare A+ and B on grounds of beneficence under the Impersonal Principles of Beneficence. In terms of average beneficence, A+ is worse than B and both are worse than A. In terms of total beneficence, B is better than A+, which is better than A. But the Mere Addition Paradox asks us to accept the conclusion that A+ is not worse than A, because A+ has more total beneficence than A, and at the same time to deny the premiss that B—which also has more total beneficence than A—is better than A, or at least not worse. But, if all we are looking at are principles of beneficence and not extra claims like (B) and (C), this denial can be made only on the grounds that A has greater average beneficence than B. And in that case A should be better than A+ as well, since A has higher average beneficence than A+. The Mere Addition Paradox is inconsistent.

In order to make the argument consistent we must deny that either total or average beneficence is grounds for preferring one outcome to another. We might then adduce Claim (B) and say that A+ is not worse than A solely on
the grounds that it adds extra people with lives worth living, and this cannot make one outcome worse than another.

Suppose we accept Claim (B), the claim that lives which are worth living but neither Wonderful nor Restricted have no intrinsic value. If we accept this claim, we deny that extra lives of less than Wonderful but better than Restricted quality like those in A+ either add anything morally relevant to beneficence or subtract anything from it. On this view such lives cannot make one outcome worse than another even though they cannot make it better. They allow for only partial comparability among outcomes. But this notion destroys the transitivity of beneficence which is necessary to any principle of rational beneficence. And, in my view, morality must be rational. Therefore, I shall argue, Claim (B) must be rejected.

Why does the claim that lives that are neither Wonderful nor Restricted have no intrinsic value destroy the transitivity of beneficence? Let us take the following cases. (In each case below we assume that the average and total beneficence in each group is roughly as drawn.)

\[ \begin{array}{cccccc}
A_1 & B_1 & B_2 & C_1 & C_2 & \\
\hline
B & B & C & & & \\
\end{array} \]

A, B, and C are identical except that B and C both contain an extra population group whose lives are neither Wonderful nor Restricted. All the lives in the identical groups
are wonderful. And as A, B, and C all have the same number of people in this group (group 1) they are morally identical thus far. But if we accept the idea that lives in the middle level have no negative or positive moral value, A, B, and C are indistinguishable in any morally significant way. Moreover, this progression can continue until we reach a case where the extra population is only slightly better than Z on the Repugnant Conclusion scenario. And still the moral equation will not change. Which is absurd.

Now consider a different situation where no one has lives of either wonderful or restricted quality.

(2)
No one's life is Wonderful or Restricted in any group. \( A_1 \) has 66% more total beneficence than either \( B_1 \) or \( C_1 \). But if we add \( B_2 \), which has the same number of people as \( B_1 \) at a much higher quality of life than that in the original groups, \( B \) comes out with more total beneficence and higher average beneficence than \( A \). And if we add \( C_2 \), at the same level of beneficence as \( C_1 \), \( C \) should have less total and average beneficence than either \( A \) or \( B \), whereas originally \( C_1 \) was identical to \( B_1 \) and worse in both total and average beneficence than \( A_1 \).

Are we to imagine that no one's life in this universe has intrinsic moral value? If not, \( A_1 \) must be in some relevant sense better than \( B_1 \) and \( C_1 \). And this sense must be that \( A_1 \) has more total and average beneficence. But if the addition of extra people to \( B \) and \( C \) has no moral import, even though \( B \) now has higher average and total beneficence than \( A \) and \( C \) has lower average and total beneficence than \( A \), \( B \) cannot be better than \( A \), which in turn cannot be better than \( C \). But either total or average beneficence must be our criterion of morality in deciding that \( A_1 \) is better than \( B_1 \) and \( C_1 \). So it should remain our criterion when extra people are added. Either this universe has no lives with intrinsic moral value, which is absurd, or the principles of beneficence cease to apply when extra people are added, so that they can allow us to distinguish between \( A_1 \), \( B_1 \), and \( C_1 \) but not between \( A \), \( B \), and \( C \), which is also absurd. To accept the claim that lives which are less than Wonderful and better than Restricted have no intrinsic moral value is to accept its counter-intuitive implications, including the implication that beneficence relations are intransitive. It is thus
to destroy the solution to the Paradox of Future Generations through a satisfactory principle of beneficence. The claim must, therefore, be rejected.

Parfit also arrives at the conclusion that the claim that extra lives in the intermediate zone are intrinsically valueless must be rejected, and on similar grounds to ours: he rejects the implications of the claim, specifically the New Repugnant Conclusion. This says that

For any possible and large population, say of eight billion, all with a very high quality of life, there must be some much larger imaginable population whose existence, other things equal, would not be worse, even though its members have lives that are not much above the Restricted Level.

(FG, p. 168)

But if we deny this claim we must affirm the negation of it. Lives that are neither Wonderful nor Restricted do have intrinsic value derived from the quantity of beneficence or utility they contain (or produce).

Suppose, however, that although we reject the claim that there is a zone within which lives have no intrinsic value we attempt to justify the claim that A+ is not worse than A on Paretian grounds: a state of affairs resulting from a change on which some are made better off and none made worse off cannot be worse than the previously existing state of affairs. The trouble is that although Paretianism and the Impersonal Principles of Beneficence coincide on Same Person Choices, i.e., choices on which the same people are affected whichever of our alternatives we choose, when different people are affected by our
alternatives the Impersonal Average Principle of Beneficence and Paretianism may diverge. For the Impersonal Average Principle denies that a change on which no one is made worse off and some are made better off is necessarily a change for the better (or at least not a change for the worse).

If newcomers are made better off by being caused to exist and have lives worth living, and those who previously existed are not affected by the addition of these extra lives, and the extra lives are of less than the previous average quality, the Impersonal Average Principle says that the change is for the worse, whereas Paretianism says the change is for the better. This result can be avoided—and Paretianism reinstated—only by assuming the existence of the newcomers on superior moral grounds to those of beneficence. This could be the case if the newcomers were our own future generations but not if they were merely added. But by itself the Impersonal Average Principle of Beneficence does not imply Paretianism. Therefore, we can accept the Impersonal Average Principle of Beneficence without accepting Paretianism. Indeed we cannot accept both where mere addition is concerned. Instead, we can argue that the mere addition of lives worth living should be morally judged by its effect on the average quality of life and that this is the only morally relevant effect where beneficence is concerned. In that case we can easily solve the Mere Addition Paradox by consulting the Impersonal Average Principle of Beneficence.

Parfit is also aware that the state of affairs that results from the mere addition of extra people in A+ is
worse on the Impersonal Average Principle than the state of affairs that previously existed. But he asks why the mere addition of extra people would make for a worse state of affairs than that which previously existed. We can reply that in order to solve the Paradox of Future Generations we needed a wide principle of beneficence. The principle that proved best was the Impersonal Average Principle of Beneficence. If that is our principle of beneficence of choice we should use it consistently.

It is true that it is not better either for the \( A_1 \) people or for the \( A_2 \) people that the \( A_2 \) people not exist. These people are nothing to the \( A_1 \) people, hence do not make things worse for them. And their own lives are worthwhile to themselves, hence do not make things worse for themselves. But it is worse in the morally relevant sense that extra people live at the lower quality of life through no fault of their own. That is, it would be better in the wide average sense if there were no extra people to drag down the average quality of life in the world, even though this would be better for no one who ever exists. We operate on this principle when we rely on the Impersonal Average Principle of Beneficence. (This principle cannot be applied to bring about a better situation in the Mere Addition Paradox, but it remains the same.) And the Impersonal Average Principle does not imply either Claim (B) or Paretianism. Therefore we can reject both these claims.

The solution to the Mere Addition Paradox just as to the Paradox of Future Generations, is to employ the Impersonal Average Principle of Beneficence. With this principle, \( A \) is better than \( A^+ \), because the average
beneficence in A is higher than that in A+. Also, A is better than B, because the average beneficence in A is higher than that in B. Divided B equals B. And A+ is worse than B and Divided B, because the average beneficence in A+ is lower than that in B. Thus we deny Claim (B), add the premiss that what matters morally is average beneficence, and come out with a consistent set of premisses and an intuitively satisfactory ranking of outcomes: A, B and Divided B, A+. And the paradox dissolves.

The solution is reached without recourse to other possible moral factors, such as Claim (C). These would not, as it happens, have changed the moral picture anyway, since A+ is worse than B with respect to Claim (C) and Claim (C) is irrelevant to A. Therefore, we can readily admit that B is better than A+ with respect to Claim (C) as well as with respect to both total and average beneficence. But, in the only morally relevant senses in which they are comparable, B is equal to A in equality and worse than A in average beneficence. Therefore, all things considered, B is worse than A, and A+ is worse than B.

The conception of morality we have developed thus solves the Paradox of Future Generations, avoids the problems that plagued past attempts at resolution, and dissolves the Mere Addition Paradox. And it can meet the obvious objections against it. Therefore, it appears that the general principle of beneficence we have been seeking is the Impersonal Average Principle of Beneficence as restrained by reason.
FOOTNOTES


3 "The Paradox of Future Individuals," Philosophy & Public Affairs, 11, no. 2, 93-112. Further references will be incorporated into the text under PFI.

4 "Future Generations: Further Problems," Philosophy & Public Affairs, 11, no. 2, pp. 113-72. Further references will be incorporated into the text under FG.

5 Cf. R.M. Hare, "Abortion and the Golden Rule," Philosophy & Public Affairs, 4, no. 3 (Spring 1975). Reprinted in Moral Problems, James Rachels (ed.), New York, Harper & Row, 1979, pp. 151-173. Hare notes that "we do commend actions which resulted in our own existence; every Sunday in thousands of churches we give thanks for our creation as well as for our preservation and all the blessings of this life; and Aristotle says that we ought to show the same gratitude to our earthly fathers as 'causes of our being.' So it is at least meaningful to say of God or of our fathers that if they had not caused us to exist, they would not have been doing as well for us as they could." (p. 171)
6 Claim (3) in Parfit. This becomes Claim (A) in Derek Parfit, "Future Generations," rough draft, [1983]. I shall call it Claim (A) since we have numbered no previous claims.

7 We assume for the moment that the better-off child will have a life of average quality in order to avoid the complication that on the Impersonal Average Principle of Beneficence it will be better for the potential mother to have no child than to have one who will lower the average quality of life.

8 The total principle says also that the mother should have a Happy Child. But in order for the conception of happy children to lead to a prisoners' dilemma on a total principle, their conception would have to reduce the quality of life so much that there would actually be less total beneficence, or a lower quantity of beneficence, in the world after their conception than before. This is very unlikely, as we saw.

9 There is a possible preference utilitarian answer to the Exploitation Policy. This is that discrimination which is not justified in utilitarian terms is an additional disutility. An opponent of utilitarianism can reply, however, that in that case all we need do is change the numbers to compensate for the extra burden. In this case utilitarianism might also be making an additional questionable assumption, namely that discrimination can be compensated by material benefits. I shall not discuss the issues of the additional disutility of discrimination or
of commensurability here. Rather, I will leave the
Exploitation Policy case as it stands. Neither issue
affects the argument.

10 Immanual Kant, *Foundations of the Metaphysics of
Morals*, Lewis White Beck, trsl., Indianapolis, Bobbs-
Merrill Educational Publishing, 1959. Further references
will be incorporated in the text under FMM. Instead of
referring to page numbers in Beck's translation, I shall
refer to sections in Kant in square brackets.

11 Cf. FG, p.114: "Our concern about future people
should not be less simply when, and simply because, they
will live later. If other things are equal, we ought to
have as much concern about the predictable effect of our
acts whether these will occur in 200 or in 400 years.
This has great importance." And elsewhere he has said
that we should not have less concern for a five-year-old
boy killed by a mantrap ten years after the trap is set
than we would have for one who was alive when the trap was
set. (Tsanoff Lectures, Rice University, March, 1982.)

12 The full claim is actually more general than this:
we cannot make up for harms we do people even by benefits
to those same people. But the general claim will seem
bizarre until we have explained the notion of harms below.

13 Because it can be completely impersonally formalized,
the moral system I am portraying is compatible with
Parfit's reductionist theory of personal identity with its
implication that morality should be impersonal (Cf.
"Personal Identity and Morality," rough draft.) and with the view that we can get a complete objective picture of the universe without introducing the notion of the individual point of view. (Cf. Christopher Peacock, Sense and Content, Oxford, Clarendon Press, 1983.) The system can be viewed as completely objective in its formal aspects (although its empirical applications will vary with the subjective assessment of facts).

14 "Ethical Theory and Utilitarianism," Utilitarianism and Beyond, Amartya Sen and Bernard Williams, eds., Cambridge: Cambridge University Press, 1982, pp. 23-38. Further references will be incorporated into the text under ETU.

15 "Morality and the Theory of Rational Behaviour," Utilitarianism and Beyond, pp. 39-62. Further references will be incorporated into the text under MTRB.


17 Because it judges each effect of a choice as the only outcome for those affected, hence judges the rationality of each final position independently, our system avoids the charge that dogs utilitarianism, namely that by lumping all the people affected together into one utility function and considering only utility information, utilitarians reduce respect for persons to a meaningless notion. Cf. John Rawls, A Theory of Justice, Cambridge, Mass., Harvard University Press, 1971, p. 27; and Amartya
Sen and Bernard Williams (eds.), "Introduction," Utilitarianism and Beyond, p. 4.

18 Our system avoids another problem that Hare acknowledges gives him particular trouble, namely that of the fanatic. (Cf. ETU, pp. 29-30). Since on Hare's system the moral agent is viewed as bearing all the effects of his own choice, a fanatic can justify, e.g., genocide since in his mind the good effects "he" receives outweigh the bad effects "he" receives. On our system the moral agent must be willing to prescribe for all and accept for himself any final position as his only final position. For example, he must be willing to prescribe and accept death for himself even though he has done nothing wrong except be born of the wrong race or religion. He cannot justify slaughtering some of "his" selves for the greater benefit of other of "his" selves. No rational person could think genocide moral if his only final position were to be that of victim. And so with other intuitively immoral choices that harm some people for the benefit of others. Our system does not justify fanaticism. Utilitarianism does. (Hare bases his defense of the utilitarian formal structure on the unlikelihood of the existence of real fanatics.)

19 Harsanyi also offers an axiomatic justification for average utilitarianism in morality. (MTRB, pp. 48-49). It is difficult to perceive all the philosophical ramifications of this theory since they are not spelled out. But of the three axioms of individual rationality, rationality of moral preferences, and Pareto optimality, the
first two are subject to the same objections I raise against the equiprobability model and the third is arguably not implied by a wide conception of average utilitarianism (Cf. my argument, pp. 141-143, against Parelianism as an implication of the Impersonal Average Principle of Beneficence); so it does no work for the system.

20 We shall see in the next chapter that, in fact, the three postulates of practical reason derivable from the categorical imperative are not sufficient to allow us to determine the intrinsic goodness of effects. We need a fourth postulate. (Cf. pp. 101-103.)

21 I owe the term "goal rights system," to Amartya Sen's paper, "Rights and Agency," Philosophy & Public Affairs, 11, no. 1 (Winter, 1982), 3-33), but the system itself is entirely my own.

22 Derek Parfit, "Future Generations," rough draft, [1983], pp. 22-23

23 Parfit acknowledges that this claim implies that there is no objection to our choice of the Risky Policy. (FG, p.118)

24 Parfit's argument is that for the man with the 14-year-old mother to claim his rights were violated when he was conceived is for him to make an "appeal to unfulfillable rights whose predicted violation no one would regret." ("Future Generations," rough draft, p.27) Our
counter-argument is that rights, like beneficence, must be understood to apply in the wide rather than the narrow sense. That is, they apply to whoever is born, not to specific individuals. And they require rational treatment of whoever is born.

25 Harsanyi also stresses that interpersonal utility comparisons must rest on a similarity postulate, which he defines as "the assumption that once proper allowances have been made for the empirically given differences in taste, education, etc., between me and another person, then it is reasonable for me to assume that our basic psychological reactions to any given alternative will be otherwise much the same." (MTRB, p. 50) I would stress that while I agree that the preferences of other people are to be determined on this basis, taking into account the information we have as to relevant differences of preference, only those preferences that can meet the other criteria of rationality are to count as rational. For example, if we know nothing about affected people's preference for sadism, we are to assume that they are like us - not sadists. If we happen to know that they are sadists we are still to call their sadistic preferences irrational, for if we ourselves (the sadists) were to be the victims we would dislike our position and could not recommend it for others. The stricture against sadism holds even if the moral agent is not only a sadist but also a masochist who would like to be tortured. For he can look around and see that in this morally relevant respect his potential victims differ from him, and he cannot morally act towards them in ways they consider to
be harmful. Therefore he cannot recommend that sadistic actions be a practical law. (The only possible exception would be a sado-masochistic relationship between consenting adults.)

Another point. Preferences, as I use the term, are fundamentally likes and dislikes, not interests or desires. While we can assume that, with certain allowances for the empirical givens of environment and perhaps personal taste, all people have roughly the same rational likings we cannot assume that all people affected by our choices have anything like the same interests, and there are many problems inherent in taking the maximization of desires as a moral end.

26 In this case we do not have to resort to calculations under the Principle of the Least Evil to determine which harm is worse. In other cases we would.

27 Impartiality does not require equal outcomes for everyone affected by a choice no matter what the choice. The moral agent is supposed to take circumstances into account. Therefore, if some people could not, because of unalterable circumstances, benefit as much as others from the same choice the moral agent will not necessarily find the position of the worse-off people morally unacceptable.

28 Notice that this argument says nothing whatsoever about abortion. It may seem to follow that if we should conceive only those children who will predictably make positive contributions to society we should give birth to
only such children. In fact, this need not follow. When we are debating whether or not to conceive someone, we are talking about an abstract future individual, one who cannot be harmed since he is non-existent. When we are debating whether or not to abort someone, we are talking about a present actual human being—at least from the moment that human being has brain waves. In this case the fetus is, in my opinion, a living person and therefore has full rights, specifically the right to equal respect. This right surely implies that we cannot merely eliminate innocent human beings. If we cannot morally kill people who actually exist on the grounds that they inconvenience us or are handicapped, we cannot morally kill fetuses with functioning brains. (That amniocentesis cannot tell whether or not a fetus is defective until sixteen weeks of gestation have passed is to me a medical problem but not a philosophical one: it is wrong to abort sixteen week fetuses. Fortunately, it seems that we are on the verge of pushing diagnosis back beyond the point at which brain waves begin. And when we have done that, it will cease to be immoral to abort defective fetuses, if we do so early enough—for the fetus aborted before twelve weeks is not a person.)


31 As a people. It is not the case that every individual values the perpetuation of himself more than increases in his own quality of life.

BIBLIOGRAPHY


Kavka, Gregory S. "The Paradox of Future Individuals." *Phil. & Public Affairs, 11,* No. 2 (Spr. 82), 93-112.


_______. "Rights and Agency." Phil. & Public Affairs, 11, No. 1 (Win. 82), 3-33.
_______. "Utilitarianism and Welfarism." J. Phil., 76, No. 9 (Sept. 79), 463-489.