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

MORAL CHOICE

by

EVELYN KEYES

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

DOCTOR OF PHILOSOPHY

Professor Baruch A. Brody
Chairman

Professor Richard E. Grandy

Professor Stephen L. Klineberg

HOUSTON, TEXAS
MAY, 1985
Copyright

Evelyn Vincent Keyes

1985
ABSTRACT

In this dissertation I attempt to devise a satisfactory general principle for moral decision-making. I assume such a principle would give complete and consistent directives; it would give intuitively sound directives; and its fundamental aim would be the maximization of the good, or utility. Future generations provide a test of principles of moral choice. For a sound principle should accommodate choices affecting all people at all times. But attempts to accommodate choices affecting future generations under a general principle of moral choice have led to two paradoxes: the paradox of future generations and the mere addition paradox. The principles of total and average utility both give complete and consistent directives in all cases, including those where future generations are concerned; but both have counter-intuitive implications. Total utilitarianism must be rejected, for it implies the "Repugnant Conclusion," i.e., that for every large population with an excellent quality of life there is an enormous population with a wretched quality of life but more utility which it is better to bring about. The principle of average utility, however, can reply to the major objection against it, that it forbids adding people who lower the quality of life. It seems, prima facie, to be the comprehensive principle we seek. But utilitarianism faces two great problems. First, it justifies any means to the end of utility maximization. Second, suppose we define utility as personal preference satisfaction. Then according to utilitarianism the good simply consists in preference satisfaction and the right in maximizing preference satisfaction. But the satisfaction of some preferences is intuitively bad.
Neither the problem of justifying bad means to a good end nor the problem of bad preference satisfaction can be solved within a utilitarian framework. But a principle of moral choice which gives priority to the satisfaction of certain preferences, designated under moral contracts as rights which we have a duty to respect, over the maximization of general utility will give satisfactory results in all cases where utilitarianism gives counter-intuitive ones. The principle of average utility can be retained as a moral remainder rule.
PREFACE

This dissertation was inspired by the Tsanoff Lectures in moral philosophy given by Derek Parfit at Rice University in March, 1982. In those lectures Parfit argued that no extant moral theory can satisfy all of our intuitions about moral choice where future generations are concerned. As I worked on the problems Parfit had raised, it occurred to me that the subject was much bigger than simply the accommodation of future generations in our moral calculations. A theory of moral choice that failed when our choices affected future generations would fail with respect to a great many of our choices, whereas a theory that successfully accommodated choices affecting future generations should be able to accommodate choices affecting any people at any time. So I turned to the question whether there might be a theory of moral choice that would always direct us to make the intuitively right choice.

My first thought was that a moral choice must be distinguishable by its conformity to Kant’s three formulations of the categorical imperative. That is, a moral choice should be one that conforms to a general principle which the moral agent could will to be a universal law; it should be one that respects all people, including the agent, rather than using some people only as means to another’s ends; and it should fall under a principle that a legislator would be willing to make into law if he himself were to be subject to that law.

The trouble with the Kantian approach to the theory of moral choice is that it is entirely formal. None of its provisions has any substantive content. I asked myself what it would mean to apply the categorical imperative; and I arrived at the conclusion -
surprising to me and antithetical to my expectations — that only utilitarianism seemed to offer a general plan for applying the categorical imperative to substantive choices.

Utilitarianism posits a substantive end of morality, as Kant does not — universal benevolence. It says that each person seeks to maximize his own good, i.e., his own personal utility or happiness, and that a moral choice maximizes the general good, or the utility or happiness of all those affected by it, relative to the alternatives. Moreover, given universal benevolence as the end of moral choice, utilitarianism tells us specifically how to achieve it. It directs us to place ourselves in the positions of every person affected by our choices, to count equally the equal units of utility and disutility that will be realized by those choices, no matter whose utility is in question, and to maximize utility over-all.

Universal benevolence is specifically disavowed by Kant as the end of morality, but it seems a peculiarly Kantian end; for the maximization of the good, as determined by the people affected, is surely an end that shows respect for persons as ends in themselves. And, given this end, utilitarianism seems to recognize all of the requirements of the categorical imperative as criteria of moral choice. Surely a moral agent could will that a principle which directs him to maximize the general happiness be a universal law. This is especially so if the principle treats people as equally worthy of respect; and the principle of utility arguably does this when it counts equal units of utility and disutility equally, no matter whose they are. Finally, the utilitarian principle requires a moral agent to put himself in the position of all the people
affected by his choice and choose that alternative which is best for all positions together - which seems to be exactly what a legislator subject to his own laws would do.

Yet clearly utilitarianism is flawed. It is subject to the charge that it commits a fallacy by assimilating the ends of each person to the end of all. If every person seeks his own happiness it does not follow that we respect the aims of each - or our own aims - by maximizing the happiness of all together. Moreover, utilitarianism gives intuitively wrong answers to moral questions too much of the time.

I decided that utilitarianism could not serve as a necessary and sufficient theory of moral choice; but it offered a good starting place in the quest for such a theory. If I could specify the flaws in utilitarianism and correct for them, a theory of moral choice should emerge that would always direct us to make the intuitively moral choice. This is the task I have attempted to carry out in this dissertation.

Several people have been of great help to me in this undertaking. I have benefited greatly from Larry Temkin's friendship, his willingness over several years to read repeated - and sometimes contradictory - drafts of the papers that finally went to make up the dissertation, his penetrating criticisms, and especially his interest in the same issues and willingness to discuss them with me. I have benefited from the erudition and keen critical mind of Baruch Brody, who directed this dissertation, and from his willingness too to read repeated drafts of the work and to play devil's advocate against the arguments. One of the greatest benefits from working with him was that I was often compelled to try
to find clear and persuasive arguments for my own point of view
against strong arguments for a different position. Some of my
arguments could not meet this test and had to be changed
accordingly. I am also grateful to Richard Grandy, who read two
vastly different drafts of the whole work and who gave me not only
couragement but the excellent criticism of a philosopher who had
the advantage of not being a professional moral philosopher but a
logician. Finally, I would like to thank Stephen Klineberg, who
read the whole thing through, not as philosopher but as a
sociologist. The final test of the work must be whether other
students of human behavior besides philosophers can read it, fit its
arguments against their own understanding of moral choice, and find
that it has something to say.

Evelyn Keyes
INTRODUCTION

This dissertation is about moral choice. It is an attempt to discover the actual way in which at our best we make moral decisions.

I make one fundamental substantive assumption in the argument. This is that the fundamental principle for moral decision-making is a principle of some sort for the maximization of beneficence, or utility, or the good. I further assume throughout that any comprehensive principle for moral decision-making will be rational in the sense that it will give an answer to every case of moral choice and it will give consistent answers. And when it is applied it will give intuitively sound results.

The dissertation is divided into five chapters.

In Chapter I, I examine the problems that future people present to the theory of moral choice. Choices affecting future people provide a good test of the adequacy of a theory of moral choice, since a theory which can accommodate these choices can accommodate choices affecting any people at any time. Yet there has been a great deal of controversy over the moral treatment of future people.

I argue that the rationale for encompassing choices affecting future people in our theory of moral choice is our belief that if we can make people better off than they would otherwise be we should, regardless of when they live. But this notion is more complicated than it sounds. For there is a real question whether we can be said to make future people better off than they would otherwise be - or benefit them - at all. And, assuming we can, we must still decide how a choice benefits future people, how much a
choice benefits them, and how great an obligation we have to them. I argue that our choices do benefit future people by causing them to exist and have lives worth living and that future people should receive the same consideration that we accord people presently alive in our moral decision-making.

But future people present special difficulties to the theory of moral choice. For different choices cause different future people to exist, so that different people benefit according to which of our alternatives we choose. Derek Parfit has called this "the non-identity problem." When we acknowledge the non-identity problem and also avow that the objective of morality is to make people better off than they would otherwise be two troubling paradoxes of moral theory arise.

The first of these, the paradox of future generations, says that if our moral objective is to make people as well off as possible any choice which causes future people to be born must fulfill the objective of morality - no matter what the other alternatives are. For if the people affected by our choice would not exist if any other alternative were chosen the choice which causes them to exist must be made as well off as possible; and since the people who would otherwise have existed never exist, they are not harmed. Any choice which influences the future course of events must, therefore, be as good as any alternative choice.

The second paradox is called the "mere addition paradox." It was developed by Parfit. This paradox says that if a population with a much lower quality of life than that prevailing in the world, but not such that life is not worth living, is merely added to the world without affecting those who already existed, the newcomers
will be better off and no one will be harmed. Hence there appears to be no objection to the mere addition of extra people to the world. But this claim conflicts with certain other moral notions we have, such as the notion that a smaller society with a high quality of life is better than a larger one with more total utility but a much lower quality of life.

Both these paradoxes appear to arise from utilitarianism - or the moral theory which directs us simply to maximize utility, or to make people as well off as possible - but I argue that in fact neither one does. The paradox of future generations arises from moral justification by Paretianism - the requirement that our moral choices be Pareto optimal, or make at least one of members of society better off without making anyone worse off. Utilitarianism, by contrast, is concerned only with the total or average utility of alternatives, regardless of what the position of those who are affected would otherwise be. Similarly, the mere addition paradox arises when Paretianism is combined with utilitarianism. Paretianism is inadequate in itself as a theory of moral decision-making where future people are concerned and pernicious in combination with utilitarianism. Utilitarianism, however, has no difficulty determining an ordering of all choices affecting present and future people on the basis of utility information. We should reject Paretianism and tentatively retain utilitarianism as a theory of moral choice capable of accommodating choices affecting any people at any time.

There are two basic forms of the principle of utility, the use of either one of which by itself solves both paradoxes: the principle of total utility, which directs the moral agent to choose
that one of his alternatives which will produce the greatest net amount of total utility, regardless of how many people share in the benefit; and the principle of average utility, which directs the moral agent to average the total utility of each alternative over the people affected by the alternative and to choose that alternative which maximizes net average utility, regardless of its total utility.

Parfit has argued forcefully, however, that neither form of utilitarianism is satisfactory as a moral theory, specifically because of the implications of each in the area of morality which deals with future people. The principle of total utility, according to Parfit, implies that for every large population with a very high quality of life there is a much larger population with a much lower quality of life, such that life is barely worth living, which contains more utility and is what we should seek to bring about. Parfit calls this the "Repugnant Conclusion." But the principle of average utility is no better, for it implies that we should not bring anyone into existence who will lower the average quality of life of society below what it would otherwise be; and this conclusion is draconian.

Because of the counter-intuitive implications of both total and average utilitarianism Parfit argues that we need to find a new principle of utility, "principle x." The hope is that a compromise principle of utility which gave weight to both quantity and quality of utility would avoid the counter-intuitive implications of both traditional principles without requiring the rejection of the principle of utility altogether.

I argue, however, that no rational or substantive criterion
can be devised for constructing a compromise principle of utility but that it doesn't matter in the slightest that it cannot. Although the Repugnant Conclusion does give us grounds to reject the principle of total utility as the utilitarian principle of choice, the principle of average utility is not as draconian as it appears. Indeed, a satisfactory reply can be made to every serious philosophical objection that has been brought against it. Insofar as any form of the principle of utility can satisfy our moral intuitions the average principle can.

The principle of average utility seems to be the principle of beneficence we seek; but I argue that there are certain cases in which no principle that determines the morality of choices exclusively by their over-all utility maximization will require the choice of the intuitively correct alternative. The problem is that any principle which judges the value of choices exclusively by their maximization of utility cannot preclude the use of intuitively immoral means to the moral end of maximal beneficence. I argue that this failure of utilitarian moral choice does not, however, give us reason to reject the notion that morality is substantively a matter of utility maximization. The question is whether it consists in mere utility maximization.

I conclude that the moral aim of maximizing average utility is necessary but not sufficient for moral choice. Moral choice requires more than a moral end; it requires moral means to that end. Given that the maximization of beneficence is our aim, we are still constrained to use moral means to achieve that aim. Moral choice must then consist in utility maximization within moral means.

Nevertheless, in Chapter III, I consider the suggestion that
by appropriately defining utility we may be able to find a utilitarian theory of moral choice without any counter-intuitive implications.

I begin by seeking a definition of utility; and I conclude that only one acceptable definition has ever been given: utility should be defined as the satisfaction of personal preferences. Thus preference utilitarianism is the only satisfactory form of utilitarianism so far developed. But I argue that although preference utilitarianism has great strengths and can meet certain objections against it, it cannot satisfy all our moral intuitions.

When a preference utilitarian defines utility as personal preference satisfaction he is defining the good as personal preference satisfaction. For in utilitarianism the good simply is utility. Moreover, in a utilitarian theory of moral choice a right choice is one which maximizes utility. Therefore, in preference utilitarianism a right choice is one which maximizes the satisfaction of personal preferences.

The trouble is that the satisfaction of some personal preferences — such as preferences for sadism, or envy, or spite, or gluttony — is intuitively bad. But, if this is the case, the preference utilitarian definition of the good is paradoxical: it calls all preference satisfaction good even though some preference satisfaction is intuitively bad. I call this problem "the paradox of bad preference satisfaction."

It has seemed to several leading utilitarians that the paradox of bad preference satisfaction can be solved — along with the problem of utilitarianism's approval of counter-intuitive means to the end of utility maximization — if we can find a criterion for
the exclusion of the satisfaction of certain preferences from utility.

I argue, however, that no criterion for the identification and exclusion of bad preference satisfaction will allow us to distinguish between good and bad preference satisfaction and still retain the utilitarian definition of the good. No formal criterion of bad preference satisfaction excludes all the preference satisfaction we want to exclude, because our objection to the satisfaction of these preferences is not that the preferences are formally incorrect; e.g., it is not that they are not personally rational or prudent. Rather, it is that their satisfaction is substantively repugnant. But substantive criteria cannot serve as criteria of good or bad preference satisfaction either, for they do not correct the utilitarian definition of the good: they replace it.

We must therefore either accept the preference utilitarian claim that the good is simply the satisfaction of personal preferences just as they are or try to come up with some other entirely different substantive definition of the good. But no other definition of the good has been even remotely successful. Moreover, substantive criteria of the good run the risk of dictating the good for people from a totalitarian point of view, which is entirely alien to the notion of the good as determined by individual preference. Therefore, for want of a viable definition of the good different from the utilitarian one, or even a suggestion as to how the intrinsic good could be discovered, I propose we keep the preference utilitarian definition, warts and all. But if we do this we will still not want to call it right to maximize utility no
matter what that utility consists in.

I argue that there is a way to solve both the problem of bad preference satisfaction and the problem of utilitarianism's endorsement of counter-intuitive means to the end of utility maximization. We can accept the utilitarian definition of the good as preference satisfaction; we can even accept the claim that there is only one end to morality, namely personal preference satisfaction; but we can reject the utilitarian claim that a right choice simply consists in the maximization of personal preference satisfaction, with all preferences accorded equal weight.

In Chapter IV, I argue that the objective of morality is the maximization of the general good — or preference satisfaction — within certain constraints. These constraints take the form of impersonal contracts understood to hold among all persons who occupy certain positions with respect to choices. Moral contracts designate certain interests — or the satisfaction of certain preferences — as rights and prescribe obligations, or duties, to protect those interests — or to satisfy those preferences. These moral contracts also prescribe conditions for the waiver of rights and appropriate compensations for rights violations. Appropriately compensated violations of rights do not count as contract violations.

Morality requires that respect for moral contracts be given priority over the mere maximization of personal preference satisfaction. If a moral contract will be broken, no matter what we do, morality as I envision it requires us to minimize contract violations. Only if there are equal contract violations under all our alternatives, or no contract violations, does morality call for
us to maximize utility.

This conception of morality requires a new model for moral decision-making in which the effects of contract violations are counted separately and before the other effects of choices in order to determine the morality of a choice.

Having proposed a new model for moral decision-making, I show that not only can the model meet objections against it, but it allows us to solve moral problems which eluded solution under the utilitarian model.

There is one particularly interesting objection which can be brought against the new model for moral decision-making, however, namely that it is nothing more than old rule utilitarian wine in new bottles. In my view rights and duties, like utility itself, are reductive: they are whatever people take them to be, although the justification for according them seems to be the belief that the maximization of the general good depends on their recognition by all. Rule utilitarianism also requires us to obey those rules which would give the best results if everyone followed them, even though in a particular case the results may be inoptimal.

The difference between my model for moral decision-making and rule utilitarianism is that rule utilitarians typically endorse exactly the same model for moral decision-making as act utilitarians, i.e., as those utilitarians who hold that the right action is that which will maximize utility in the specific case. Although rule utilitarians think of the right choice as that which falls under an optimal rule, they have no separate place in their model of moral decision-making for the recognition of the effects of a rule and even conflicts of rules, as distinct from the effects of
a choice in terms of general utility, as I do. Thus rule utilitarians cannot give priority to certain effects of a choice over other effects. They can only sum the utilities and disutilities that follow from everyone’s acting the same in similar circumstances. My theory of moral decision-making is therefore not rule utilitarianism. I call it "conditional utilitarianism." And I claim that not only does the conditional utilitarian model for moral decision-making give intuitively results but that it is the model which at our best we actually use.

The argument thus proceeds from the assumptions that moral choice is rational and intuitively sound and that it consists in utility aggregation of some form, through an examination of the strengths and weaknesses of various proposals for the best form of the comprehensive general principle for moral decision-making, to the assessment that preference satisfaction is all there is to utility but mere utility maximization is not all there is to moral choice. I argue that there are in our actual morality contractual obligations which require us to satisfy certain designated preferences, which we call rights, before all others. The best model of moral choice is one that gives respect for these contractual obligations priority over the general good. But rights extend only as far as implicit moral contracts between agents and the recipients of moral action provide. I propose a new model for moral decision-making which gives a place to respect for these moral contracts; and I argue that the new model both meets objections against it and directs intuitively sound moral choices where earlier models for moral decision-making do not. Moreover, it reflects the way in which at our best we really do make moral choices.
TABLE OF CONTENTS

Preface.................................................................iv
Introduction.......................................................viii

I. Moral Choice and Future
   Generations.....................................................1

II. Utilitarianism and Population Policy.........................31

III. Preference Utilitarianism.......................................74

IV. Utility and Constraint.........................................108
    Bibliography..................................................153
I. MORAL CHOICE AND FUTURE GENERATIONS

I. Thesis

Future generations provide a test of the adequacy of a theory of moral choice. For a great many of the choices we make affect those who live in the future. Yet these choices have proved very difficult to accommodate in a theory of moral choice. In this chapter I will address the special problems presented to the theory of moral choice by future people and I will take the first steps towards developing a theory that accords future people intuitively sound recognition.

II. Future Generations and Moral Choice: The Issues

The rationale for including choices affecting future people in a theory of moral choice is the belief that if we can make people better off than they would be otherwise we should, regardless of when they live. But there are a number of hidden problems lurking in this apparently simple concept.

For one thing, we need to determine whether or not our choices can actually be said to "benefit" future generations at all, properly speaking. If so, we need to determine the extent of the benefit of a choice to a future person. And we then need to ask whether future people have any claim on our beneficence and, if so, whether their claim is equal to the claim of people already alive or not. Once we have answered these questions we will see that we still are not done, for future people present peculiar problems for moral calculation which threaten to involve us in paradox; and we must meet that challenge as well. Derek Parfit in particular has raised all these issues. To a large extent, therefore, this chapter will pursue topics raised by Parfit. (1)
a. The causal effect of choices on existence and the non-identity problem

We wish to determine whether it makes sense to claim that our choices benefit future generations. If future people benefit from a choice we make they must benefit in the sense that the choice is an essential part of the cause of their existing and having lives worth living. So first we must establish that our choices are indeed essential parts of the cause of the existence of those future people who actually exist. The argument that our choices do cause the existence of future people is not new. Parfit and others have made it before. But it is useful to repeat it here so that we can see just how far-reaching the causal effect of our choices on the existence of future generations really is. (2)

Most people would agree, I think, that who we are depends upon which egg and sperm united to cause us to be born and that that union depended in turn upon the timing of our conception. If these claims are true, as they seem to be, my husband’s and my decision to have a child now will cause a different child to be born than our decision to conceive a child next year or my decision to have a child with another man; and, if I decide to have children, presumably a different number of people will exist than if I decide not to. Thus our reproductive decisions are an essential part of the cause of the existence of those future people who are actually born.

But personal reproductive decisions like these are not the only ones that have an essential causal effect on which people and how many people are actually born. Abortion policies and population policies, for example, pretty clearly influence the number and
identity of those future people who are actually born.

It is much less obvious that many of our most far-reaching and interesting personal and policy choices influence who is actually born; but they do, as the following example shows.

Suppose we have a choice between building a deep water port to receive super tankers and not building one. If we build the super port we will need to bring in extra workers and pay them. We will eliminate or alter other projects. These transfers of people and resources will change the situation now and in the future. For if we build the port previously unrepresented entities in the business community will open banks and branch offices and more countries will open consulates. A more cosmopolitan atmosphere will be created for the city, enriching the economy. And more money will attract investors to the city who will build highrises and open jewelry stores, boutiques, and restaurants. In many ways, anticipated and unanticipated, direct and indirect, the make-up of the city will change.

As the city changes because of our choice to build the port so will the population. For with the influx of new people, people who otherwise would never have met will meet, marry, and conceive children. It seems reasonable to conclude that after several generations have passed an entirely different population will exist and be affected if we choose to build the port than if we choose not to. Let us call the choice to build the port A and the choice not to build it B. Then we can call those future people who will exist if we choose A the A-people and those future people who will exist if we choose B the B-people. My argument implies that if we choose A the A-people will exist and the B-people will not, and if we
choose B the B-people will exist and the A-people will not.

Not only do our choices cause future people to exist but the choice of different alternatives causes different people to exist. Parfit has called this "the non-identity problem." It is simply the problem caused by the fact that different future people exist according to the different alternatives we choose. This problem will return to haunt us; but for now I will put it aside.

The important point at this juncture is that our choices — and not only our reproductive choices but many other choices, including perhaps all of our important choices — are essential parts of the cause of the existence of whatever future people actually are born.

b. Existence as a harm or benefit

The next question that must concern us is the question whether those future people our choices cause to exist benefit from those choices. I will assume that if future people never exist because we fail to make a certain choice they are not harmed. Non-existence is not itself a harm nor do non-existent beings suffer harms, for there is no one there to harm. Only actual people suffer harms and enjoy benefits. I will also assume that if a choice causes future people to exist and have lives not worth living that choice harms those people. It would seem, by the same token, that a choice that causes future people to exist and have worthwhile lives must benefit them.

Some philosophers have taken as self-evidently true both the claim that future people benefit from choices that are essential parts of the cause of their existing and having worthwhile lives and the claim that future people are harmed by being given lives that
are not worth living. (3) But other philosophers have endorsed what Parfit has called "the Asymmetry." This claim holds that while future people are harmed by being caused to exist and have lives that are not worth living they are not benefited by being caused to exist and have lives that are worth living. (4) It is hard to devise logical grounds for the distinction, and no one has succeeded, so far as I know. But, since this is an area of controversy, we cannot simply assume that future people do benefit from being caused to have worthwhile lives without arguing the point.

The argument that future people do not benefit from being caused to exist goes like this. People benefit from a choice only if they are made better off by that choice than they would otherwise be. We cannot compare the lot of future people whom our choices cause to exist with that they would otherwise have had, since otherwise they would not have existed, and it is simply incoherent to claim that existent beings are either better or worse off than nonexistent beings. Therefore we have no basis for claiming that future people are made better off by being caused to exist than by not being.

Parfit has argued that the claim that that future people benefit from our choices is not implausible. I will go further than this. I will argue that the claim that future people do not benefit from our choices is not only implausible but false.

We can reply to the argument that future people do not benefit from our choices in a number of ways.

First of all, there is no linguistic justification for confining the use of the term "benefit" to fully comparative
situations and denying that future people meet the conditions of full comparability because no one can compare his lot as an actual person with that he would have enjoyed as a non-existent being. In order to benefit someone I need only accord him something predetermined to be a good - not something that is necessarily better than something else. Certainly this is true if I do not at the same time take away some good he already has. And in giving life worth living, which is a good, I do not take away anything the person I bring into existence already has: there is no one there to deprive of anything until I give him life. So we can say that in causing someone to exist we benefit him. If we are preference utilitarians we must say that he prefers to exist. This phraseology might seem to prejudice our argument since preference sounds like a fully comparative notion even if benefit is not. But actually it makes no difference. In preference utilitarianism preferences are choices people make, or revealed preferences, and people manifest their preference for existence so long as they do not commit suicide. We can rationally expect that most people whom our choices cause to exist will not commit suicide. Therefore we can expect those people to reveal a preference for existence.

Moreover, the argument that future people do not benefit from choices that cause them to exist fails by reductio. Suppose we assume that people do not benefit from being caused to exist. And we agree that choices that influence the future course of events are essential parts of the causes of the existence of those future people who are actually born. So the decision to build the super port in our example is an essential part of the cause of the existence of those future people who enjoy the - what, benefits? -
of the port. But we just agreed that future people cannot benefit from being caused to exist. Therefore a fortiori they cannot benefit from any decision that is part of the cause of their existence. So, since only future people "benefit" from the construction of the super port, and future people cannot benefit from any decision that is part of the cause of their existence, no one benefits from the port. Which is absurd.

People make choices that influence the future course of events all the time, and, whether or not they realize the effect of their choices on the make-up on the populations affected, they evaluate their choices in terms of expected benefits for whoever benefits. They do not cease to calculate expected benefits just because some of the recipients do not exist at the time the choice is made. The alternative is to claim that no one who lives in the future benefits from any choice that influences the future course of events. And not only is there no reason to accept this claim but no moral agent acts as if it were true.

We can also make the same argument from the other end, not as moral agents but as recipients of benefits. Many past choices influenced the then future course of events and thus were essential parts of the cause of our own existence. These include the Protestant Reformation, the Copernican Revolution, the discoveries of the telescope and the New World, the American Civil War and the outlawing of slavery, even our own future parents' decision to build a house before having children or to accumulate a nest egg to finance the education of those future children. If only pre-existing people can benefit from a choice none of us benefits from any of these choices. This means that no one in college today
because his parents bought tax-free bonds before he was conceived benefits from those bonds. No one alive today benefits from the abolition of slavery. No one who builds or voyages in a spaceship benefits from the Copernican revolution or the discovery of the telescope. None of us benefits from the discovery of America. And no one benefits today because Luther raised an outcry against the selling of indulgences or because Jesus taught in Galilee. The only people who benefited from any of these choices were those who were presently alive when the choice was made. Which again is absurd.

Suppose people actually accepted this radical interpretation of "benefit." Then no one would accumulate anything for his unborn children since he himself would lose more than he would gain by accumulating for literally no one's sake and no presently actual person would gain anything, so no one would benefit from saving. Hardly any reform could be justified since most reformers look to the future for justification; but a non-existent future justifies nothing. And no one could rationally explore any new place or any new idea, since to do so would be to risk a safe existence for a risky one in the futile hope of vindication by the future.

The only justification for interpreting "benefits" as "benefits to pre-existing persons" is temporal prejudice; and prejudice is not justification. It makes sense only to conclude that future people do indeed benefit from choices that affect them and, moreover, that it is by the expected benefits for future people - whoever they may be - that we justify a great many of our choices, and rightly so. The claim that we cannot benefit future people by our choices fails. We can benefit future people.
c. The value of existence

Next we must ask how much future people benefit from a choice that is an essential part of their existing and having lives worth living. The answer is that each future person benefits to the full extent of the quality of life he enjoys and would not otherwise have had. Future existence is an all-or-nothing benefit of a choice. This claim is called the "full comparative view." I will not defend it. Parfit has ably done that. (5)

d. The claim of future people on our beneficence

Finally we must ask how strong the claim is that future people have on our beneficence. Some philosophers have argued that even though future people enjoy the full benefit of a choice that causes them to exist, we have no moral obligation to weigh their interests - or benefits to them - equally with the interests of people already alive when we make a choice. Instead, the benefits expected to be enjoyed by future people as a result of our choices should be ever more heavily discounted as those people are more distant from us in time.

I will first note that in determining the value of a choice in terms of the expected benefits we do habitually count the benefits we expect will be enjoyed by future people in exactly the same way as we count the benefits we expect to be enjoyed by presently existing people. That is, we count the full quality of life we can expect to result from the choice, whether those who enjoy it live in the present or in the future. Expected benefits are benefits, whoever enjoys them. (6) But this practice need not only be noted; it can be justified.

If we do not count all benefits expected to be generated by
our choices in the same way we greatly distort the effects of those choices. For example, the moral value of a choice I make to buy a bond today that will pay off for my expected child in twenty years should not be discounted because it will benefit someone in the future. If the bond pays ten percent, it pays ten percent, not ten percent less a discount for the non-existence of the beneficiary when the bond was bought. Nor should the benefits of fuel conservation be counted at less than the actual expected benefits simply because the beneficial effect of conservation falls in the future. If we save a billion barrels of oil in the ground we save a billion barrels of oil; and that is the benefit to future generations; not a billion barrels of oil less a discount for futurity. (7) The rationale for counting these benefits differently from those which accrue to people presently alive is nothing more than the same temporal prejudice that keeps us from counting benefits to future people as benefits at all.

In this section I have argued that our choices cause future people to exist, that people benefit from causes that are essential parts of the cause of their existence to the full extent of the quality of life they have (assuming they have lives worth living), and that we do and should treat these benefits of our choices like any other benefits. Future people should enter our moral calculations on the same footing as any other actual people. But this brings us to the special difficulties for moral choices affecting future generations attributable to the non-identity problem.
III. Special Problems of Non-identity: The Paradox of Future Generations

The moral problems posed by future generations stem from the fact that any choice which affects future people is an essential part of the cause of their existence, so that if that choice is not made those people will never exist. Since this is the case, and since we are assuming that choices which cause future people to exist and have lives worth living benefit those people, every choice we make that causes the existence of future people with lives worth living not only benefits those people but benefits them as much as possible. And, since those who would have existed if we had chosen a different alternative never exist (because the choice that would have been an essential part of the cause of their existence is not made and non-existent beings cannot be harmed), no one is harmed. Thus every choice that causes the existence of future people benefits whoever actually lives as much as possible and harms no one. In preference utilitarian terms, if every choice that affects only future generations whose existence it causes makes everyone it affects as well off as he can possibly be and harms no one, everyone affected will prefer the choice that is actually made and no one will prefer any other alternative. Therefore it appears that, if our criterion of moral choice is maximal beneficence, or making people better off than they would otherwise be, any choice which affects only future people must be as good as any alternative choice. This problem is called the "paradox of future generations."(8)

The paradox of future generations appears to point up a problem with utilitarian justification since a utilitarian moral
theory holds essentially that the supreme object of morality is to maximize benefits, or the good, or utility. But in the following sections I will show that this is not the case.

The paradox actually is generated by moral justification by Pareto optimality - or Paretianism - a concept which is irrational in itself where future people are concerned and incompatible with utilitarianism, despite the claims of many philosophers to the contrary. Moreover, when Paretianism is combined with utilitarianism, another paradox, Parfit’s mere addition paradox, is generated, as we will see below. Paretianism is the culprit where future generations are concerned. If we disavow it we can solve the special difficulties posed for moral choices affecting future generations by the non-identity problem.

IV. Paretianism and Utilitarianism

Pareto optimality is essentially an economic concept that has found its way into moral theory as the claim that a choice cannot be bad if there is no one for whom it is worse. I will call moral justification via Pareto optimality "the Pareto optimality requirement," or, for short, "Paretianism."

Paretianism derives from an efficiency principle propounded by the economist Vilfredo Pareto. (9) Pareto argued that if the distribution of utility in society is such that no one can be made better off without someone else’s being made worse off the distribution of utility in that society is efficient. And if there are two alternative distributions such that no one can be made better off without someone’s being made worse off both are efficient and the society is held to be indifferent between them. Some economists and moral philosophers describe any Pareto efficient
state of affairs as "Pareto optimal." (10) But efficiency is not
the only interpretation of Pareto optimality; nor is it the
interpretation I will use.

In moral philosophy the term "Pareto optimality" is
generally used to compare the status quo with states of affairs
containing extra utility. Suppose we can either maintain the status
quo or bring about another state of affairs containing more utility.
(True to our previous definition, we define utility as preference
satisfaction.) Economists and moral philosophers call the latter
state of affairs "weak Pareto optimal" if everyone will be better
off than before and "strong Pareto optimal" if someone will be
better off and no one will be worse off. Strong Pareto optimality
requires more than weak Pareto optimality, namely that no one be
made worse off in making someone else better off. In weak Pareto
optimality since everybody is made better off obviously nobody is
made worse off. In preference utilitarian terms, weak Pareto
optimality is defined as the claim that society will prefer a state
of affairs which every one of its members prefers; and strong Pareto
optimality is defined as the claim that society will prefer a state
of affairs which at least one of its members prefers over another,
so long as everyone else is indifferent between the two states of
affairs.

Pareitanism requires that we choose Pareto efficient and
Pareto optimal states of affairs over other alternatives. In this
it presents a moral "social choice rule," or rule for the moral
evaluation of choices affecting the welfare of society. (11)
Utilitarianism also presents a moral social choice rule, but
utilitarianism diverges from Pareitanism early on.
Utilitarianism calls for the maximization of utility but is notoriously indifferent to the distribution of utility, so long as each equal unit of utility, or util, is counted equally. Suppose Joe and Sam are both building bonfires on Guy Fawkes Day and we find some extra wood. We can simply give the wood to either Joe or Sam or divide it among them, or we can snatch some wood off Joe’s pile and add it to Sam’s along with our offering. The total and average utility will be the same whichever choice we make. It is a matter of indifference to a utilitarian which we choose. (12)

Paretianism, on the other hand, while endorsing the notion that utility should be maximized, contains a built-in preference for certain distributions of utility. It approves only those choices that make someone better off without making someone else worse off, or that will satisfy more preferences for someone without thwarting the preferences of anyone. It will not let us morally take some of Joe’s wood. It will only let us give our extra wood to either Joe or Sam or split it between them. Paretianism looks like utilitarianism, but it is more restrictive.

Moreover, as a social choice rule Paretianism is incomplete, for in many cases there is no Pareto optimal alternative. This lacuna in Paretianism occurs when maximal utility can be generated only by a choice that harms someone or thwarts someone’s preferences. These cases cannot be decided using a true Paretian criterion.

The following case illustrates the point. Either we will hang, draw, and quarter Guy Fawkes or we will let him go free. It is 100 of us, the members of parliament, against him, the traitor; and we want to make the choice as preference utilitarians and
Paretians. We all have a very strong desire to see Guy hung, drawn, and quartered. Not surprisingly, Guy prefers to go free. No one entertains any other preferences, such as sparing Guy’s life and sending him into exile, so no one ranks any other possible preferences. We can assume that the over-all utility of hanging, drawing, and quartering Guy Fawkes is much greater than the utility of letting him go free. A utilitarian will call for the former. But there is no simple Pareto optimal alternative. Whichever choice we make we must thwart someone’s preferences.

Paretianism can, however, be extended to cover just those cases where someone’s preferences will be thwarted whatever we do. This is done by arbitrarily making it a condition of social choice that society be indifferent to the state of affairs that results from a choice so long as someone’s gain compensates for someone else’s loss. If one person’s losses on A vis a vis B are exactly made up by someone else’s gains on A, the society will be indifferent between A and B. And if one person’s losses on A vis a vis B are more than compensated by someone else’s gains, A will be socially preferred to B. In our example, the thwarting of Guy’s preference for going free is more than compensated for by our satisfaction at seeing him drawn and quartered. This "Pareto extension rule" (13) incorporates the simple weak and strong Pareto optimality rules and extends them to cover alternatives that are incomparable under basic Paretianism. When Paretianism is extended in this way it produces a social choice rule which merely ignores distributional considerations.

I will call Paretianism which has been extended by the Pareto extension rule "extended Paretianism"; and Paretianism which
has not been extended "simple Paretianism."

Extended Paretianism looks very much like utilitarianism since both offer social choice rules which require the maximization of utility. When we come to future generations, however, we find that utilitarianism and simple and extended Paretianism all offer very different ways of evaluating choices. Because of their differences, utilitarianism has no trouble at all deciding between alternative choices where future people are concerned, while Paretianism founders completely.

V. The Solution to the Paradox of Future Generations

Every choice that affects only the future is Pareto optimal in the weak sense. That is, every choice that affects only future people makes those people as well off as they could possibly be, since had any other alternative choice been made besides the one that caused them to be born they would not have existed. Simple Paretianism must therefore be absolutely indifferent among all alternative choices that affect only the future, so long as those choices give people lives worth living. Every single one of these choices will meet the weak Paretian criterion that everyone—meaning all affected people—prefer the state of affairs that choice brings about.

Suppose we have two alternative choices, A and B, which affect only people who live in the future. If we choose A we can justify the choice as Pareto optimal by consulting the preferences of those affected, the A-people. If, on the other hand, we choose B we can still justify the choice as Pareto optimal by consulting the preferences of those affected, the B-people. Neither choice can be justified as better than the other on simple Paretian grounds since
each maximizes the preference satisfaction of whoever is affected. The paradox of future generations arises from the attempt to evaluate choices that affect the future using the simple Paretian method.

Suppose, then, we try to justify choosing A over B by the extended Paretian criterion instead of the simple Paretian criterion. In that case we will choose the maximally beneficent alternative by taking the point of view of society rather than that of any particular affected individuals. But there are two things wrong with this move.

In the first place, extended Paretianism is supposed to fill the gap where there is no Pareto optimal alternative. But here the problem is not the lack of a Pareto optimal alternative but an embarrassment of riches. It is not that neither A nor B is Pareto optimal but that both are. Moreover, there is no way to assign a point of view to a society that changes over time which does not involve an internal contradiction. We cannot expand our notion of society to cover both the A-people and the B-people in order to be able to say that A is better than B from the point of view of society, since the society whose existence is caused by A and the society whose existence is caused by B are mutually exclusive. If the A-people exist the B-people cannot exist and vice versa. Any notion of society that incorporated both the A-people and the B-people would therefore be an inherently contradictory notion.

We must conclude that simple Paretianism cannot distinguish between A and B, nor can it distinguish among any other alternative choices that affect only the future; and extended Paretianism cannot apply to these choices. Where our choices affect
only future people whose existence they cause both simple and extended Paretianism fail as useful principles of choice evaluation.

Utilitarianism, on the other hand, suffers no embarrassment when confronted with choices that influence the future course of events, because it does not pretend to judge the utility of choices from the point of view of any given affected society. In utilitarianism it does not matter whether any specific people constituting any given society are better off than they would otherwise be or not. What matters is whether those people who are affected by a choice are better off than those who would have been affected by an alternative choice would have been, even if this means comparing an actual future society with one which will never exist. Total utilitarianism merely directs us to choose whichever of our alternatives will maximize total utility, and average utilitarianism does nothing more than add a step to the total utilitarian calculus, averaging utility on each alternative over those who will enjoy it and directing the choice of that alternative which maximizes average utility.

Since utilitarianism judges alternative choices affecting the future by the utility the societies they cause to exist contain relative to alternative future societies rather than by the utility those societies contain relative to their own other possibilities, it can generate a complete ordering of alternatives affecting the future where Paretianism cannot. Utilitarianism solves the paradox of future generations, which Paretianism generates.

Anyone who espouses either simple or extended Paretianism must, I think, admit its collapse as a useful principle of choice evaluation in cases where only future people will be affected by our
choices. And since utilitarianism can evaluate the choices affecting future generations that Paretianism cannot evaluate we have reason to prefer utilitarianism to Paretianism.

But choices that affect only future generations are not the only ones where Paretianism and utilitarianism can be brought to bear on our alternatives. Many of our choices affect both people who are presently alive and people who live in the future; and these choices too are subject to both utilitarian and Paretian evaluation, with somewhat different results from choices that affect only future people.

VI. Special Problems of Non-identity: Conservation versus Depletion

When we try to justify choices affecting both present and future people from the standpoint of Pareto optimality alone we sometimes run into counter-intuitive results and sometimes into irrationality.

Suppose we can choose between two alternative choices that each affect both present and future people. Let us say that the choice is between a policy of conservation and one of depletion. Depletion will be better for those of us presently alive, but over the long run conservation will maximize utility both on the whole and on the average. (The example is Parfit's.)

Utilitarianism will require the choice of conservation.

Simple Paretianism, however, is confusing. It does not apply to the choice of conservation because someone - namely ourselves - will be worse off if we choose conservation than if we choose depletion. But if we choose depletion it does apply. We will not be worse off, and future people will benefit. Depletion is
Pareto optimal in the simple weak sense. Simple Paretianism approves of depletion. But as moral justification for an energy policy this result is - however comforting to ourselves - counter-intuitive.

Nor could we resort to extended Paretianism to decide between conservation and depletion, because there is no society that includes both the future people who live if we choose conservation and those who live if we choose depletion. That is, we cannot say that "society" prefers conservation to depletion because which people society consists in will differ according to which alternative we choose; and each alternative society will prefer the alternative on which it exists.

Enough has been said, I think, to show that extended Paretianism is not equivalent to utilitarianism and that it is incoherent where future people are concerned. I will have nothing more to say about it. But simple Paretianism is not so easily scotched.

VII. Special Problems of Non-Identity: The Mere Addition Paradox

In some cases affecting both present and future people, unlike cases affecting only future people, there is only one simple Pareto optimal alternative. This is true of the conservation case. In others there is more than one, and, as before, Paretianism cannot choose between them. But this is not all that is wrong with simple Paretianism where both present and future people will be affected by a choice. If we combine simple Paretian justification with utilitarian justification we can again run into paradox. This should finally convince us to reject Paretian justification for moral choices.
Derek Parfit's mere addition paradox is a case in point. This paradox arises because simple Paretianism and utilitarianism are both used to justify a choice to merely add new people to society. (14)

In the mere addition paradox we start with an initial society, A1. There are four alternatives. First, we may leave things as they are. In this case the original group, A1, will continue to exist alone. We can call this choice A. Second, we may bring into existence another population, A2, in addition to A1. A2 will have a much lower quality of life than A1, but life will still be worth living. The existence of A2 will have no effect on A1 since the A2 people will live in a distant land and will have no contact with the A1 people. They will be merely added to the world. Third, instead of bringing A2 into existence, we may bring into existence a population at a quality of life only slightly worse than that in A1; but it will cost A1 something to do this. The quality of life in A1 will sink to the same level as that in the new society. We can call the new society B. In this society A1 becomes B1 and the newly created group is B2. B has a slightly lower level of preference satisfaction, or average welfare, than A but more total preference satisfaction; and it has more total and average preference satisfaction than A+. Fourth, we can bring A+ into existence, and once it exists we can create Divided B out of A+. This will result in an improvement in both total and average utility over A+ and, as Parfit points out, may satisfy the requirements of justice by making the worse off better off at small cost to the best off. The alternatives look like this.
Parfit asks us to assume that A is better than B. A+, Parfit argues, cannot be worse than A because there is no one for whom it is worse – a Paretian claim. Therefore the A2-people prefer A+ to A and the A1-people are indifferent between the two alternatives, as required by strong Paretianism. Divided B, however, is superior to A+ in total and average utility, as well as on apparent grounds of justice, so that it seems Divided B must be better than A+. But Divided B is equal in every morally significant way to B. Therefore B must be better than A+, which is not worse than A, while B is worse than A. The paradox arises from these inconsistent claims.

From a rational standpoint if two types of moral justification are essentially incompatible we must reject one. We can solve the mere addition paradox by distinguishing simple Paretian from utilitarian judgments and rejecting one or the other. (15)

Among A, A+, B, and Divided B, A+ is the only simple Pareto optimal alternative. Neither B nor Divided B is Pareto optimal since the A1 people lose utility in bringing about either. If we are simple Paretians we will choose A+. But we will not be able to
rank the other non-Pareto optimal alternatives with respect to each other. Our social choice rule - or rule for the justification of choices affecting the welfare of society - will be incomplete. Moreover, if we had had yet another alternative, similar to B2 except that A1 would not have been affected by its creation, we would have had two Pareto optimal alternatives, A+ and the new alternative; and simple Paretianism would not have been able to distinguish between them. Simple Paretianism dictates the choice of A+, but it is evident we cannot depend on it as a social choice rule for cases involving both present and future people except in very special circumstances, namely where there is one and only one Pareto optimal alternative.

Utilitarianism offers the only necessary and sufficient means of evaluating choices affecting both present and future people with respect to utility information. Total utilitarianism approves of the move from A to A+, just like simple Paretianism, but on different grounds; it approves the move because A+ contains more total utility or preference satisfaction than A. And it approves the moves from A to B and from A+ to Divided B on the same grounds. Average utilitarianism, however, condemns the move from A to A+ because the average utility in A+ is less than that in A. And it similarly condemns the move from A to B, although it approves the move from A+ to Divided B. Both total and average utilitarianism evaluate B and Divided B equally. Total utilitarianism thus ranks the alternatives B and Divided B, A+, A; and average utilitarianism ranks them A, B and Divided B, A+. Either type of utilitarianism give us a complete ordering of alternatives. Either can therefore serve as justification for choices affecting both present and future
people. (16)

Since the use of utilitarianism and Paretianism together generates both the paradox of future generations and the mere addition paradox, and simple Paretianism by itself offers insufficient justification for choices, while utilitarianism suffers from no intrinsic inadequacies, we should reject Paretianism and use utilitarianism in either the total or the average form as a social choice rule capable of justifying choices affecting future generations.

VIII. Conclusion

In this chapter I have argued that our rationale for considering future generations when we make moral choices is our belief that we have a moral obligation to make people better off than they would otherwise be, no matter whether they live now or in the future. I have addressed the issues that set future generations apart with respect to moral choices; and I have concluded that future people deserve to be treated just like any other people when we make a moral choice. But because of the non-identity problem special difficulties attend the attempt to accommodate future people in moral decision-making on an equal footing with everyone else. These difficulties result in two important paradoxes, the paradox of future generations and the mere addition paradox. I have argued that both these paradoxes arise when we attempt to combine Paretianism - or the requirement that our choices be Pareto optimal - with utilitarianism. Paretianism is fatally flawed on its own account as a principle for justifying choices affecting future generations, and utilitarianism is not. We should reject Paretianism in favor of utilitarianism. We will then be able to
solve both paradoxes. (17)

I do not claim that if we reject Paretianism in favor of utilitarianism all our problems in arriving at a satisfactory general theory of moral choice will be over. We will not even be home free on the issue of accommodating future people in moral decision-making in a way compatible with our intuitions. For Parfit has argued forcefully that both total and average utilitarianism have counter-intuitive implications where choices affecting future generations are concerned.

The important point for our purposes here is that utilitarianism does solve the special problems that arise when we try to accommodate choices affecting future generations in our theory of moral choice. Moreover, a utilitarian type principle — i.e., one that compares the lot of those who do live with that of those who would have lived had a different choice been made — is apparently the only kind of principle that does. If we are to arrive at a satisfactory general theory of moral choice further work lies ahead. My argument suggests we should proceed by looking more closely at utilitarianism.
FOOTNOTES


2. I do not claim that they are the entire cause of the existence of those future people who actually live. Other factors, such as chance, are also essential parts of the cause of the existence of future generations. All that is necessary to my argument is the claim that our choices are essential parts of the cause of the existence of future people.


6. I do not mean to imply that all expected effects should count the same, no matter how uncertain or unlikely. A sound utilitarian theory should discount for the unlikelihood of certain effects whether they fall in the present or in the future. Only more or less direct and anticipated or anticipatable effects should enter the utilitarian calculus. But remoteness in time is not itself a guarantee of the insignificance or unlikelihood of effects.

7. The claim that the utility expected to be enjoyed by
future people should count equally with the utility expected to be enjoyed by present people has also been defended by Parfit. Parfit 1984, p. 357.


11. The concept of social choice is ambiguous between the notion of a choice made by all the members of society (for example, a vote or the body of free market selections) and the notion of a choice which respects the preferences of the members of society, or the choices they would make for themselves. I am using the concept in the second sense.

12. A utilitarian would be expected to object that taking from Joe to give to Sam actually generates extra disutility. We can meet this objection by having Sam’s extra-large bonfire generate more utility for the greater number of neighbors who will see it than if Sam’s bonfire were smaller or Joe’s were larger.


15. The conflict between Paretianism and utilitarianism is noted by Thomas Schwartz (Schwartz 1979). "Welfare Judgments and Future Generations," Theory and Decision 11 (1979), 181-194. Schwartz argues, however, that the mere fact that Paretianism and utilitarianism conflict in the mere addition paradox is sufficient reason to reject "all versions of utilitarianism broadly conceived, including all admixtures of utilitarianism with criteria of distributive justice." (p. 190) R. I. Sikora also notes that average utilitarianism can conflict with strong Pareto optimality. He takes this conflict to be a counter-intuitive consequence of using the average utilitarian principle and concludes that "it is enough by itself to demolish average utilitarianism." R. I. Sikora, "Is It Wrong to Prevent the Existence of Future Generations?," Sikora and Barry, p. 116. I think the conclusions of both men outrun their premises.

16. On my view the distributional issue is something of a red herring. We could not rationally endorse both a principle of distribution and a conflicting principle of maximal utility any more than we could endorse both Paretianism and utilitarianism. When rules conflict we must choose between them or bring them into accord with each other.

17. This argument has ramifications beyond the scope of this paper.

few weak claims about social rationality — including the Pareto optimality claim — rational social choice is impossible. This is called Arrow’s “impossibility theorem.” But Arrow does not make clear exactly what is impossible — society’s having a rational preference, or a choice of its own determined by the choices of its members, or there being a way to aggregate the preferences of the members of society so that a person could make a rational choice that respected those preferences. That is, he does not sort out the ambiguities in the concept of social choice.

My argument shows that the notion that “society” can have a Pareto optimal preference of its own is incoherent if we consider the wishes of future people in determining society’s preference. Moreover, it shows that no principle of social choice which aspires to aggregate the preferences of the members of society in a rational manner can be derived from the principle of Pareto optimality. John Rawls’ difference principle is such a principle. The difference principle is a principle of just distribution which picks out a point on the curve of Pareto efficient social situations — that where the worst off are not made worse off by a choice — as the morally significant point. (Cf. Rawls 1971, pp. 66ff. for the derivation of Rawls’ difference principle from the Paretian efficiency principle.) It takes any distribution of society’s goods to be just so long as the worst-off are not made worse off. The argument against Pareitanism shows that a Pareitan-based principle of just distribution which tries to take future people into account will collapse. For no distributional decision whatsoever that affects future people will make those people worse off than they would otherwise be, so long as they have live worth living. The
elimination of aid to dependent children, for example, would be fine if we were to use the difference principle as a principle of just distribution. Rawls handles this problem by having the difference principle be chosen as a principle of just distribution by people in an original position behind a veil of ignorance as to their future position in society, but people whose own future existence is not in question. But my quarrel is not with the initial choice of the difference principle by people in the original position. It is with its actual practical use as a principle of just distribution in a society whose membership changes over time.
II. UTILITARIANISM AND POPULATION POLICY

I. The Argument

The goal of a moral population policy is to give our children and our children's children a better life than they would otherwise have had. If our population is growing too fast to maintain the quality of life we try to limit its growth; if it is declining too precipitously we encourage people to have more children. (1)

There are two major philosophical questions about population policy, however. The first concerns the question whether we can make sense of an obligation to people who do not yet exist. The second—assuming we can make sense of such an obligation—concerns the question what type of principle could be used morally to justify a population policy.

Both issues have recently been raised by Derek Parfit.

Parfit argues that the special moral problems presented by future generations can be overcome by taking a utilitarian approach to population policy but that both total and average utilitarianism have unacceptable implications for population policies. He calls for the development of a compromise principle of utility, "principle x," that will avoid these counter-intuitive implications.

I will argue that indeed there is no theoretical obstacle to developing a moral population policy along utilitarian lines. I also agree with Parfit that the total utilitarian principle should be rejected as the utilitarian principle of choice in population policy. But I will argue that the compromise principle Parfit seeks—principle x—is a chimera. There is no satisfactory way to devise such a principle. We need not despair, however: the
principle of average utility not only satisfies our intuition that a population policy should make life better for future people on the average than it would otherwise be, but it can satisfactorily meet all charges brought against it except one, and its one shortcoming is shared by every version of the principle of utility. This is that utilitarianism justifies any means whatsoever to the end of utility maximization, no matter how intuitively immoral, so long as those means are the most expedient.

I will argue that the strengths of the average utilitarian principle give us reason to retain it as the welfare principle of choice in population policy, but that its shortcomings give us reason to condition its application on the prior satisfaction of a principle of means to the utilitarian end, one that requires respect for people's rights. I conclude that the principle of average utility is a necessary but insufficient principle for the justification of a moral population policy.

II. The Non-Identity Problem and the Paradox of Future Generations

The whole concept of a moral population policy risks foundering on what Derek Parfit has called "the non-identity problem." (2) The problem is this.

Many factors act together to determine which people are actually born. Some of these are beyond our control or are not direct results of our reproductive choices. But reproductive choices are an essential part of the cause of the existence of those future people who are actually born. If Mary Merry enters a convent, for example, she will probably not have any children. If, on the other hand, she marries Ralph Handsome she probably will have children. And since a different sperm and egg will be produced
today than five years from now, if Mary and Ralph decide to start their family right away they will have different children than if they decide to wait five years to begin their family. Also, if they both decide to have a career they may have a different number of children than if Mary primarily thinks of herself as a wife and mother. And what is true of Ralph and Mary is true of all of us: our reproductive choices are essential parts of the cause of the number and identity of those future people who are actually born. The purpose of a population policy is to influence those choices. A population policy is therefore itself a choice that helps determine a set of reproductive choices. If we choose between different population policies different people will exist.

The non-identity problem generates a serious moral problem, the "paradox of future generations." (3) If a choice is an essential part of the cause of someone’s existence he will never exist if an alternative choice is made. If he enjoys a life worth living as a result of a choice it is plausible to say he benefits from that choice, and if he must endure a life not worth living we can say he is harmed by the choice. (4) But if he never exists he is not harmed. Every choice is therefore as good as it could possibly be for whoever actually lives as a result of it, assuming that he has a life worth living. Indeed it is better for him than any alternative choice would have been, since on any other alternative he would not have existed. And if it affects only future people whom it gives lives worth living it harms no one. There seems to be no moral reason to choose one alternative course over another in the interests of a better life for future generations, so long as on either choice people will have lives worth living, since whatever
alternative we choose everyone who ever lives in the future will be as well off as he could possibly be and no one will be harmed. Unless we can solve this paradox it will be hard to justify any kind of population policy, or any other policy or choice that affects future generations, as better for our descendants than any other.

The paradox of future generations can be solved, however, if we take a utilitarian approach to choices that affect the future, for in utilitarianism it does not matter who benefits from a choice. If someone will benefit from choice A and someone else will benefit more from alternative choice B it matters not at all that if we choose A the people who would otherwise have lived will never exist, and likewise if we choose B. A utilitarian judges a policy or a single choice solely by its effect on human welfare, impersonally considered. The non-identity problem and the paradox of future generations to which it lends give us reason to think that we should take a utilitarian approach to decision-making when our choices will affect the welfare of future people.

There are two traditional forms of the utilitarian principle, the principle of total utility and the principle of average utility. Either solves the paradox of future generations; but each does it differently. Total utilitarianism calls for the maximization of total utility, or welfare, no matter how many people share in the total, and average utilitarianism calls for the maximization of the level of utility when total utility is averaged over the number of people sharing in it. Thus where different numbers of people will be affected according to which alternative we choose – as is quintessentially the case with population policies – total and average utilitarianism can give different directions.
But Parfit argues that the use of either principle has counter-intuitive implications for moral population policy. Both principles do indeed solve the paradox of future generations, but both do so unsatisfactorily.

III. Total Utilitarianism and the Repugnant Conclusion

Suppose we believe that life worth living is a good thing in itself and that the object of morality is to maximize the total amount of life worth living. We are total utilitarians. It follows that we should want to create as much happy life as possible, taking into account the state of all those who live. But, Parfit argues, the principle of total utility implies the Repugnant Conclusion. This says that for every large population (A) with a high quality of life there is a much larger population (Z) with a much lower quality of life that has more total utility than the former population and is what we should seek to bring about, even though life in Z is barely worth living. (Cf. Parfit, 1982, p. 140ff. and Parfit, 1984, p.381ff.) The Repugnant Conclusion is morally unsatisfactory. We should reject the principle of total utility which implies it.

Parfit offers the following situation in illustration of the claim that morally reaching the Repugnant Conclusion is a real possibility under the principle of total utility. I will call it the Repugnant Conclusion scenario.

We, the presently existing population A, decide to have so many children that the total size of the population doubles. So long as the average quality of life in the new population B, consisting of ourselves and our children, is more than half the average quality of life that previously existed in A, the principle of total utility justifies our decision to bring about B. And if
the population can again be doubled in the next generation at a cost of less than half the average quality of life in B, bringing about population C, the principle of total utility will endorse the move to C. And so on to Z, where the population is enormous and life barely worth living.

It is important to be clear about the Repugnant Conclusion. A total utilitarian will not insist on increasing the population if more total utility will be generated by limiting population growth. Nor does the total utilitarian approve of increasing population to the point of maximal sustainable population. If he did, he would have to approve increasing the population even though the total utility of the society was falling, up to the point where there were so many people that some actually began to die off from starvation or disease. A total utilitarian approves of increasing the size of the population only so long as the total utility of the society can be increased by the addition, i.e., only so long as the population increase adds marginal utility to the society.

Whether or not we actually reach the Repugnant Conclusion depends upon empirical factors. If the addition of extra people means that each succeeding generation will have less than one half the average welfare of the previous generation but twice the number of people or more, the total utilitarian principle will tell us to restrict population growth, and we need never reach the Repugnant Conclusion. But we do not have to accept the strong claim that we must reach the Repugnant Conclusion: the mere fact that total utilitarianism would call moral a population policy under which we would go on adding ever more people at a cost in the average quality of life less than the gain in the total quality of life lived while
the quality of life sank to the level of Z, together with a scenario
for attaining this result, is enough to make us take the Repugnant
Conclusion seriously.

I think its direct implication of the Repugnant Conclusion
gives us sufficient reason to reject total utilitarianism as
justification for a population policy.

IV. Average Utilitarianism, Impartiality, and Temporal Neutrality

In contrast to total utilitarianism, average utilitarianism
justifies the type of population policy it seems we intend to
justify when we undertake a moral population policy, namely one
where everyone is as well off as possible on the average; and it
cannot directly imply the Repugnant Conclusion since it forbids
bringing people into the world who will lower the average quality of
life below what it would otherwise be. But to many moral
philosophers average utilitarianism provides no better or even worse
justification for a population policy than total utilitarianism.

Parfit and others have brought many arguments against
average utilitarianism. In this section I will address two of them.

The first argument against average utilitarianism has been
put most forcefully by L. W. Sumner. Sumner calls it the argument
from utilitarian impartiality. He argues that the dictum "each to
count for one, no one to count for more than one" — meaning "each
unit of utility to count for one unit and none for more than one
unit" — is an essential feature of utilitarianism going back at
least to Mill and that average utilitarianism subverts this dictum.
(5)

We can reply, however, that all that utilitarian
impartiality requires is that each unit of utility be counted in the
same way as every other; and average utilitarianism does this. The only difference from total utilitarianism is that instead of stopping with the total utility expected to be generated by a choice average utilitarianism goes one step further and averages utility over the persons affected. In adding this step to the total utilitarian formula average utilitarianism can hardly be said to introduce favoritism or partiality into utilitarian calculation. Rather, it can be argued that it takes a step in the right direction, namely towards a greater recognition of the results of choices for individual persons than total utilitarianism allows. The real argument boils down to this: average utilitarianism is not total utilitarianism; therefore we can reject it. This argument from authority is too weak to have any real force against average utilitarianism. We can disregard it.

Let us assume, then, that average utilitarianism really is impartial, not only with respect to persons but also with respect the time at which they live, since it judges choices in the same way whether they affect people presently alive or people who will live in the future. The average utilitarian principle is not only impartial but atemporal. This brings us to another problem.

Derek Parfit has charged that temporally neutral average utilitarianism is absurd. He argues that temporal neutrality on a utilitarian principle requires that we add new people only if doing so will most contribute to the welfare of all the people who ever live considered together. But the requirements of temporal neutrality are different on the principles of total and average utility. A temporally neutral total principle of utility will approve adding any new people with lives worth living. But a
temporally neutral average principle of utility runs into trouble, for it requires that we add new people only if doing so will most contribute to average welfare, atemporally considered. The only way to make the necessary calculations is to figure in the average welfare of all the people who have ever lived or ever will live, given our choice, to see whether or not that choice will improve the average. If the ancient Egyptians had the highest quality of life of any people who have ever lived, we have lives of lower quality, and our children would have lives of the same average quality of life as ourselves, hence lower than that of the ancient Egyptians, our children’s existence would lower the average quality of life of all people atemporally considered. The atemporal average principle of utility would say we should not have children. But it is absurd to think our moral calculations require us to know the average quality of life of all the peoples who have ever lived. Therefore we should reject temporally neutral average utilitarianism. (Parfit 1984, p. 420.)

But in fact the temporally neutral form of the principle of average utility would not require us to figure the quality of life of everyone who ever lived into our population policy calculations. Or, rather, it would have this effect only if we could make decisions from a timeless point of view. But we cannot. We make our decisions at a specific time; and, because of time’s forward arrow, those decisions can affect only present and future people. Since a utilitarian of any kind judges choices exclusively by their effects on people, he must judge his own choices by their effects on present and future people – the only people they can affect. These are the people the atemporal utilitarian principle requires must be
treated alike. Since in practice no decision of ours can affect the ancient Egyptians, we need not consider their welfare in determining the value of any choice of ours. So in practice the atemporal average utilitarian principle cannot have the absurd implications with which Parfit charges it.

V. Average Utilitarianism and the Mere Addition Problem

Parfit has, however, advanced yet another, very important, objection to average utilitarianism. This is the argument from "mere addition." (Cf. Parfit 1982, p. 159ff., and Parfit 1984, p. 420ff.) Average utilitarianism tells us it is wrong to bring extra people into existence who will lower the average quality of life, even though their existence benefits them and hurts no one else. But, Parfit asks, "What is wrong with a lower average, if it involves no loss of any kind, but only mere addition?" (Parfit 1982, p. 159) He argues that the mere addition of people with lives worth living who affect no one else - no matter how much worse off they are on the average than those who already existed - cannot make a society worse, although it need not make it better. Parfit argues that unless we can show why we should consider the mere addition of people with lives worth living to the world worse than not adding them we must reject average utilitarianism.

I argued against the mere addition claim in Chapter I. (6) I argued that the mere addition claim is a form of the Pareto claim that a choice which is worse for no one cannot be bad. Contrary to the common philosophical perception, Paretoism is incompatible with any form of utilitarianism, including total utilitarianism. We cannot keep both Paretoism and utilitarianism. Utilitarianism is a much richer and more valuable source of moral
justification than Paretianism. We should keep utilitarianism and reject Paretianism and with it the mere addition claim. That argument need not be repeated. But there are other arguments against the mere addition claim.

A. Mere addition and the Repugnant Conclusion

At first blush the mere addition claim merely seems to repeat in a weak form the total utilitarian claim that so long as the addition of people to society adds marginal utility their addition cannot be bad. If this were all it said we could bring arguments against it similar to those against total utilitarianism.

Suppose a large population with an extremely high quality of life already exists and it is possible either to bring no one else into existence or merely to add a second population with many times as many people as the first population, all of whom will live at slightly better than subsistence level. We can suppose the first population is equivalent to A and the second is equivalent to Z on the Repugnant Conclusion scenario. The addition of the second population will not affect those who already exist.

The mere addition claim says that if we merely add Z to A the resulting society will not be worse than that which previously existed since those who will be added will be given lives worth living — hence will be benefited — and those who already existed will not be affected in any way — hence will not be harmed. This seems just as repugnant as the original Repugnant Conclusion, however. Adding Z to A seems just as bad as bringing Z about instead of A. If a universe contains only very happy people we do not intuitively make it better by adding a lot more people with mean and cramped lives that are barely worth living.
The only difference between the A plus Z scenario and the Repugnant Conclusion scenario is that in the former Z is merely added to A and in the latter Z is reached serially from A. But position in time is not a morally relevant factor.

Nor need there even be a distinction between the Repugnant Conclusion scenario and the mere addition scenario. The mere addition claim itself directly implies the Repugnant Conclusion.

Suppose we in A can merely add a second population group twice as big as our group but with an average quality of life lower than our own by almost 50%. The second group will not affect us in any way. The average quality of life in the whole society consisting of ourselves and the newcomers, i.e., in A+, will, however, equal the average quality of life in B on the Repugnant Conclusion scenario. Mere addition approves the transformation of A to A+ and so does total utilitarianism. A generation later repeat the mere addition, adding a population at a lower quality of life than that in A+ - low enough to cause the average quality of life in the world to drop to the level of C in the Repugnant Conclusion scenario. And so on until the quality of life has reached the level of Y. Now let the people in Y merely add a group equivalent to Z and let the population in Y die out of natural causes. The only remaining population will be Z.

The outcome on this mere addition scenario is identical to the outcome on the RC scenario. This is possible because the question of the average and total utility in society taken as a whole is independent of the question of the distribution of utility. That is, it doesn’t matter whether the newcomers affect those who already existed or not. All that matters is the effect of the
addition of new people in terms of the total and average utility in the world. The mere addition claim implies the Repugnant Conclusion.

Our objection to mere addition is the same as our objection to total utilitarianism: both principles imply that there is nothing wrong with bringing about Z under conditions which do not seem to justify bringing it about. Therefore, since the mere addition claim permits A plus Z, or simple Z itself, we should reject it as a morally satisfactory claim if we reject the Repugnant Conclusion.

B. Mere addition and irrationality

But this is not all that is wrong with the mere addition claim. We have rational grounds to reject it also. For it says that the mere addition of people with lives worth living cannot make society worse although it need not make it better. Now in total utilitarianism the addition of people with worthwhile lives - no matter how restricted - does make society better, so long as their addition does not reduce the utility of others. So A+Z is better than A from a total utilitarian standpoint. And in average utilitarianism the mere addition of people with lives of worse than average quality actually makes society worse. So A+Z is worse than A from an average utilitarian standpoint. With either the total or the average utilitarian theory we can rank societies according to their welfare, but with the mere addition claim we cannot: A+Z need not be better than A, but it cannot be worse.

If we accept the mere addition claim in addition to either the total or the average utilitarian principle we destroy the completeness of utilitarianism. Yet completeness is an essential
feature of rational choice. It is the requirement that when two outcomes of a choice are compared either one outcome should be preferred to the other or the two outcomes should have equal utility. But with the mere addition claim an original society and the society that results from merely adding people to that society can contain different total utility and different average utility, and yet neither need be preferred to the other. Our only ground for comparing two societies on a utilitarian principle is the utility contained in the society, and in the special case of mere addition that ground is declared not to count. So if we accept the mere addition claim we destroy an essential feature of rational utilitarianism. Yet most utilitarians have thought of utilitarianism as a rational theory.

Its endorsement of Z and A+Z and its interference with rationality seem to me reason enough to reject the mere addition claim and answer enough to Parfit's question, What is wrong with mere addition?

VI. The Compromise View

A variation on the mere addition claim advanced by Parfit and other philosophers seems, however, to present a separate objection to average utilitarianism. Suppose, these philosophers argue, that the universe contained only a small population, but everyone in that population was ecstatically happy. We can call that society heaven. Now suppose God could also choose to bring into existence a second population which would not interact with the first in any way. The people in the second population would be only slightly less happy than the people of heaven. Indeed, everyone would be happier than anyone who has ever lived in the world. We
can call their home Eden. Average utilitarianism says God should not create Eden in addition to heaven. This is counter-intuitive.

Total utilitarianism, however, fares no better. Since the same amount of total utility could be contained in societies with vastly different numbers of people and levels of welfare, heaven plus Eden could contain exactly the same amount of utility as A plus Z. Hence, heaven plus Eden could be morally equivalent to A plus Z. So neither total nor average utilitarianism satisfies all our intuitions about the addition of Eden to heaven.

It seems to some philosophers, including Parfit, that we need a compromise principle of utility to designate a level of welfare between heaven and Z below which we would morally refuse to add people and above which we would be morally required to add people. (7) Since the search for a compromise view is itself complicated, and this is a natural place to introduce it, I will consider the compromise principle of utility here, before giving the average utilitarian response to this challenge.

The main problem with the development of a compromise principle of utility - principle x - is that there is apparently no rational, or well-defined, place to set the moral limits if the ideal lies between the point of maximal average utility and the point of maximal total utility. And if there is no rational place to balance the quality of life lived and the number of people who live it between the points endorsed by total and average utilitarianism, the stopping place must be - at least from a mathematical standpoint - arbitrary, or irrational. It seems wonderful to add Eden to heaven; and if Eden why not America; and if America why not Bangladesh? It seems there is no point short of the
point of 0 marginal utility at which we can say that people whose
lives are worth living should not be brought into existence from a
rational point of view unless we say that people whose existence
will lower the average quality of life of society should not be
brought into existence.

One way to avoid the conclusion that there is no rational
stopping place between total and average utilitarianism is to try to
find a formula which will permit the addition of people to society
so long as their addition does not cause a great drop in the average
quality of life.

Notice that a principle that seeks the compromise point
cannot, however, take the immediately preceding average quality of
life as the level it will not permit to be drastically lowered if
new generations are added. The total principle of utility already
does this: it does not permit those in an immediately preceding
generation to lose more than those in the succeeding generation gain
at any point on the downward slope from A to Z and the Repugnant
Conclusion. It may seem that what is needed, therefore, is a
principle that will not allow a drastic lowering of the quality of
life of the original group - A - i.e., the people who already
exist when the population policy is decided upon. Parfit calls a
principle that gives special consideration to people who already
exist when a choice is made "a person-affecting principle."

Parfit has offered two person-affecting principles of
utility - one derived from the total principle and one from the
average principle - designed to endorse the addition of new people
to society between the point where average utility starts to drop
and the point of 0 marginal utility, but not to endorse the addition
of so many people that the average quality of life of those who originally existed is drastically lowered.

A person-affecting utilitarian principle does not measure the total utility of a choice simply by calculating the overall outcome in terms of utility. Rather, it takes into account separately how the present and future population groups affected by a choice will fare. Like the principles of total and average utility, the person-affecting total principle counts the total effect on any future generation of the one alternative where it exists, since those people are not affected by the other alternative. But, unlike the principle of total utility, it counts the cost or gain for presently existing people as the difference between their alternatives, i.e., the difference between the state of affairs where new people are added and the state of affairs where they are not. For example, if people enjoy a present quality of life of 100 utils and those same people would lose 20% of their quality of life if new people were added to the population, the person-affecting total principle counts the effect of the state of affairs where no one is added as +20 utils per person for those presently alive and the state of affairs where newcomers are added as -20 utils per person for those presently alive. It does not, like the principle of total utility, look at the final result of each decision and count the utility of the one alternative as 100 utils per person and the other as 80 for those presently alive. Thus it does not treat presently existing and future people alike. (The average person-affecting principle merely adds the extra step of averaging results.)

The result, however, on both person-affecting principles is
ultimately the same with the Person-affecting Restriction as without it, as Parfit shows. So long as the newcomers will benefit more than those who previously existed will lose by a population increase we should increase population. Both person-affecting principles imply the Repugnant Conclusion. (8)

It does not seem possible to find a satisfactory principle that determines the morality of a population policy by its effect on the original generation.

Parfit, therefore, suggests that a successful replacement for the traditional principles of utility in population policy might be one that would call the mere addition of lives that are neither "wonderful" (equivalent to the lives in heaven, or even Eden) nor "restricted" (equivalent to the lives in Z, or slightly better) "valueless." The addition of lives within the valueless zone would not change the moral equation. (Parfit 1982, pp. 163-169) As Parfit acknowledges, however, all this move does is to draw the boundaries of good and bad societies a little narrower. Instead of ending up approving Z, anyone who made this move would end up approving Higher Z; and instead of endorsing the Repugnant Conclusion, he would end up endorsing the New Repugnant Conclusion. (Parfit 1982, p. 168)

But there is more to the problem than even this. The addition of a valueless zone reintroduces incompleteness into utilitarian calculations. The addition of extra people just below the level of heaven would be morally equivalent to the addition of extra people just above the misery level. It would again be impossible to rank societies by their utility levels. This would destroy the rationale for the use of the principle of utility.

Finally, even though a rational stopping place between
maximal average utility and maximal total utility is elusive and, if I am correct, non-existent, we must consider the possibility of someone's coming up with a substantive criterion that will allow us to endorse creating Eden in addition to heaven without also endorsing creating Z in addition to A. This substantive criterion will predictably be just as elusive, however, as the rational criterion. Eden looks good from our standpoint. It might not look good from the standpoint of the angels. From our standpoint Egypt has immense social problems with poverty, overcrowding, and pollution. Yet from the standpoint of Ethiopia, where people face starvation, Egypt may seem a society to strive for.

Which should we take as the substantive criterion of the appropriate stopping place between total and average utilitarianism - Eden or Egypt? If Eden, we, whose children would have no hope of aspiring to Edenic existence, would have a moral obligation not to have children. If Egypt, we would have a moral obligation to continue having children until the average quality of life in the world dropped to the level of life in Egypt. And if we take the current population of the world and the average quality of life in the world today, or some such figures, as the ideal to strive for, we may set up an unattainable ideal for future generations, given the facts of their existence, or we may decide that a much lower quality of life than that within their reach is the ideal. We would not want to measure the ideal compromise between numbers of people and average quality of life for ourselves today by a substantive standard determined by monks in the Dark Ages with reference to then contemporary Europe. Nor would we want to say that the ideal compromise between numbers of people and average welfare was
determined during the reign of Julius Caesar and remained the same after the fall of Rome.

A substantive criterion for compromising between total and average utilitarianism seems, therefore, no more attainable than a rational one.

I myself think there is a straightforward objection to the compromise view in that it implies that a society with more people and a lower quality of life is sometimes actually better than a society with fewer people and a higher quality of life. So if we are at A we have a moral obligation to strive for B, which is a better society. Why? Simply because it has more people, hence more total utility. And if we are at B we have a moral obligation to strive for C, which is a better society. And so on to some arbitrary point. Someone who espoused the compromise view would have to agree that he didn’t really want whoever lived to be as well off as they could possibly be; he wanted them to be less well off — how much less being a matter of spreading the wealth to other people who would exist only because of an (irrational, to my mind) insistence on the moral value of numbers. If these extra people were never created they would not suffer, so the concern with numbers cannot be a concern for people but only for quantity itself, i.e., for an abstraction.

The notion of sharing the wealth in the abstract seems an inadequate basis for rejecting the principle of average utility in favor of a compromise between total and average utilitarianism, especially when to entertain the notion of a compromise would entail denying that our real moral objective was to make people as well off as possible and would also either entail irrationality or require
the formulation of a substantive criterion of well-being which seems chimerical.

But we need not despair. Average utilitarianism is not the ogre it seems when it precludes the creation of Eden in addition to heaven.

VII. Average Utilitarianism and Elitism

The real concern of philosophers who adduce the heaven plus Eden example and who espouse the compromise view is that the adoption of the average principle of utility would result in a very small elitist society — one so elitist that even Eden might not qualify to be brought into existence. This vision offends people's democratic sensibilities. (If Adam and Eve in Eden cannot morally qualify for existence, who could?) But in fact the adoption of the average principle of utility would not result in very small or elitist societies. The true goals of the compromise view are met in actuality by the average view, without any violations of rationality and without invoking any chimerical substantive standards of well-being, as I will now show.

In support of the claim that average utilitarianism is elitist, those who favor the compromise viewpoint especially to two counter-intuitive implications of average utilitarianism. These are the implication that we should have no children whose existence will decrease the average quality of life and the even stronger implication that we should have just those children whose existence will most raise the average quality of life. Many philosophers have seen in these requirements the most serious objection to average utilitarianism. (9)

Average utilitarianism does imply that we should have just
those children whose existence will most raise the average quality of life; but, as with other claims, it is important to understand exactly what is meant by this requirement. This does not mean that we should have no children whose lives will be of less than average quality. At times philosophers seem to assume that people are nothing more than vessels of utility themselves and consumers of the utility of others. This assumption is false, even from a utilitarian point of view. People are producers of utility as well as consumers. When we consider the utility of adding an extra person to the world we must take into account what quality of life we can reasonably expect him to have himself and what extra utility and disutility we can expect him to produce. We cannot assume that only people who have lives of better than average quality themselves will contribute most to the average utility of society. Moreover, it would be a mistake to believe that average utilitarianism requires us to have only highly productive or inventive children. A society composed only of theoreticians would not be the society with the highest possible average quality of life. Average utilitarianism thus does not require us to have only those children whose lives will be of better than average quality themselves or those who will be the most clever or the most productive. Average utilitarianism requires us ideally to bring into existence just those future garbagemen, electricians, and clerks, as well as those philosophers and biochemists, whose interaction will yield the highest possible average quality of life.

From an average utilitarian point of view the decisive moral consideration in adding a new person to the world is whether the society which includes him will have a better or worse average
quality of life than the society which does not. Not only does this seem a defensible moral position to take; it seems to me to be the position we actually do take when we ask ourselves whether it is moral from a social point of view to bring a child into the world. (10) One real difficulty with taking an average utilitarian approach to population policy, however, is that the calculations necessary to determine the morality of reproduction from an average utilitarian point of view can only be very rough and ready. Nevertheless, this is not to say that they should not be or are not made. On the contrary, I think they are made and are made in this fashion and are no less valid for all that.

VIII. Average Utilitarianism and Narcissism

There is still a problem with average utilitarianism as an ideal, however. We can call this the challenge from narcissism. Let us imagine a two-member society that is presented with 50 extra units of utility of which to dispose any way it sees fit. Suppose the average utility in that society is 60 utils. The two members of the society can each take 25 utils, raising their own average utility to 85, or they can bring an extra person into the world and give him all 50 extra units. The existence of the newcomer will lower the average level of utility below what it already is and far below what it would be if the presently existing members of society consumed the extra utility themselves. Average utilitarianism counsels against it. But if average utilitarianism has this effect it seems to support rampant self-indulgence over parenthood - no doubt a moral comfort to the narcissists among us. (11)

Average utilitarianism would indeed have the implication that society should have no children if things were as they are in
the scenario just given. But in fact this is not the case. Only under extraordinary circumstances will the average quality of life of society be higher if no one has children.

A remark.

My argument that average utilitarianism does not imply narcissism depends upon empirical fact — the facts of society's actually being a certain way and of ordinary circumstances' prevailing. This suggests an objection. I argued that total utilitarianism should be rejected because it directly implied the Repugnant Conclusion in principle, regardless of whether or not the Repugnant Conclusion was in fact reached as a matter of total utilitarian policy. I did, however, give a scenario for reaching the Repugnant Conclusion, one which is hard to deny as a plausible description of how society’s quality of life might decline through over-population. It may, nevertheless, seem inconsistent to turn around and argue for the principle of average utility on empirical grounds when it too has counter-intuitive implications in principle.

My reply to this objection is that when we see what the average principle implies for all types of empirical cases we will see that, in direct contrast to total utilitarianism, it tells us not to have children in just those cases where we think we should not have them and to have children in just those cases where we think we should have them. I will illustrate this claim with what I believe to be a plausible scenario of my own. Thus although the argument makes use of empirical facts about society it is an argument for the justification of population policy by average utility in principle, given the facts as they actually are.

If we justify a population policy on average utilitarian
grounds we will add people to society so long as their addition will maintain or increase the average quality of life and we will cease to add people to society when the addition of extra people will cause the quality of life of all to drop below what it would otherwise be. This does not mean that we would not add people if the quality of life of society was dropping. We would weigh how low the quality of life of society would be over our lifetimes if no new people were added against how low it would be if new people were added.

If the quality of life would be expected to be negative whatever we did — i.e., if life would not be worth living — an average utilitarian would hold that no life at all was better than any life. His population policy would dictate not having children. This makes sense, however, because intuitively we do not think it is moral to bring children into a society where the quality of life of those children will be such that life is not worth living. (12)

The only cases left are those where life can be expected to be worth living whether we have children or whether we do not. Let us assume, then, such a society — a society like our own. The question is whether the quality of life of those of us already alive in such a society will be better if we decide to reproduce ourselves than if we do not.

The flaw in narcissism is its assumption that in ordinary circumstances we can maintain our own average quality of life over our lifetimes at a higher level by not having children than by having them. This is almost certainly false. The amount of utility in society does not necessarily either remain static or expand, as in our example. If it did, we would indeed increase average utility
by producing extra utility for ourselves instead of by having children. But in reality new generations are needed to maintain the highest possible average quality of life in the society as previous generations age. If we have no children the average quality of life of society will fall over time below what it would have been if we had reproduced ourselves. A society that fails to reproduce itself is one which chooses a lower future quality of life for its presently existing members than they and their children together would have had. In order to maximize average utility for those who would exist on either alternative society must reproduce itself.

Suppose we Americans and indeed people all over the world suddenly discovered an amazing consensus among ourselves: children are more bother than they are worth; they use up resources we adults could otherwise appropriate to our own use; they demand our care and attention and reward us with ingratitude; and even the contribution they can be expected to make to society when grown does not compensate for the burdens their existence places on other people. All things considered society would be better off if no one had any more children. As we are all utilitarians, we agree that all people should submit to sterilization in the moral interest of promoting the average utility of those who actually live - namely ourselves. No one in any future generation will be harmed since no future generations will ever be born; and we will be better off. Everyone is voluntarily sterilized. The question now is whether in fact we can expect our own average utility to be greater than that we and our children together would have enjoyed. The answer to that question will determine whether we have made the right decision.

Let us project a bit ahead. At first everyone is delighted
to be freed of the burden of providing for generations yet unborn. Those who are stuck with small children who beat the conception deadline are perhaps envious of their footloose neighbors, but they console themselves that child rearing is not so bad as all that; and anyway the sterilization program, whenever it took place, was for the good of all. The society throws itself into a fandango of liberation. No more sacrifices need be made for future generations. No more time or resources need be devoted to the care of the young. All can concentrate on making life better in the here and now.

A few occupations immediately lose their utilitarian justification, it is true, namely all those that are justified by their effects on the quality of life of future generations. Cancer researchers, for example, can proceed with their projects as good utilitarians so long as they can realistically expect results to benefit the present generation, but not otherwise, and so on with pure research scientists of any kind. In politics peace in our time is an adequate goal. And in culture there is no point in building better than those presently alive know. Better a Michael Jackson who bursts across the firmament today than a Bach who will shine tomorrow.

But perhaps we should say so what to these distortions of previous values. If a shift to immediate gratification rather than future gratification is the only ramification of our decision, the argument against it doesn’t seem very strong. It might be no more than an argument from prejudice and — again — elitism. Besides, against these losses we can set the gains in the quality of life we will now be able to provide for people already alive who would otherwise have led lives of substandard quality. And we need not
tell ourselves we are robbing the future to do so. We need a stronger argument than has been given if we are to be persuaded that it would have been in our own best interest as a people to have had children. No argument is forthcoming; and so we go on for ten years of bliss.

At the end of ten years we discover that no – or few – advances have been made in science and technology – less, in fact, than would have been made if there were a future beyond our own to plan for. So people live lives of somewhat less comfort than they might have had, and more people die of disease at an earlier age than would have been the case had research projects been continued. Also, those who were in the oldest generation when we undertook the sterilization program have become truly old. They require great amounts of care at great expense. There are no younger generations coming along to share the burden of their care, either physically or monetarily; it falls hard on the rest of us. Meanwhile, jobs that require great physical strength or endurance which would in other times have been handled by the young must now be handled by the aging. Need I remark also that the pleasures of hedonism begin to pall in our middle age – as notoriously they do? And we find we have denied ourselves a principal pleasure of old age, the presence of grandchildren growing to manhood and womanhood? We dare not think about this. Indeed, we dare not think ahead at all, because the dictates of real life are that we will spend our own old age in misery and despair – with no one to care for us or about us, dying one by one of unrelieved disease and want. Alternatively we could commit suicide before we become old. But this would be an admission that we could not sustain worthwhile life as long as we could have
had we had children, assuming that up to some point the dependent old still have worthwhile lives.

Only if this scenario is implausible can we plausibly claim that the average quality of life of all those who will ever live will be better if we have no children than if we do have children and share their lot. I claim the scenario is very plausible indeed. More than this, I claim that it is the only plausible scenario when a society simply stops having children despite life's being worth living. Isolation in time is not obviously advantageous to human society.

An interesting thing happens when we realize that the goal of maximizing average utility requires society to reproduce itself, for, once we realize that, we realize that if we in generation A take as a moral goal the maximal average utility of the society of which we are a part we must assume the existence of our children in B. From this point on we are no longer talking about maximizing our own average utility but about maximizing the average utility of ourselves and our children together, i.e., the average utility of some B1,...,Bn. Moreover, since average utilitarianism requires us to conceive just those children who in our best estimation will contribute most to the average welfare, it requires us to bring into existence future generations which will contain not only those individuals who will contribute most to society, whatever their own personal attributes or quality of life, but also the optimal number of contributing individuals. It will require us to bring about that Bi with the highest average quality of life possible. And this is precisely the goal of a moral population policy.
IX. Average Utilitarianism and the Repugnant Conclusion

But, wait a minute, we might want to say, doesn’t this argument simply exchange the Scylla of our ordinary understanding of average utilitarianism for the Charybdis of the Repugnant Conclusion? For I have just endorsed B over A; and if that is so why should I not endorse C over B and so on to Z?

The key to the difference between total and average utilitarianism lies in recognizing that B is a variable and that different theories of utility maximization will require us to bring about different B’s.

The point of divergence between total and average utilitarianism is that every time total utilitarianism asks whether we couldn’t squeeze out just a bit more marginal utility by bringing into existence some extra person with a life barely worth living average utilitarianism asks whether we couldn’t make life better on the average by restricting population growth. When our choice looks like this, total utilitarianism will tell us to have the extra child and average utilitarianism will tell us to limit population growth.

It is possible to reach the Repugnant Conclusion via average utilitarianism, but only indirectly and in very particular circumstances. Specifically, average utilitarianism endorses bringing into existence a population with a lower quality of life than a present population enjoys only when there is no alternative population with a quality of life equal to or higher than that presently enjoyed which could be brought about instead. And even then it tells us to minimize the loss in the average quality of life. So, although average utilitarianism can lead to the Repugnant Conclusion, that conclusion loses some of its intrinsic repugrance
under force of circumstance: average utilitarianism will endorse bringing about Z only when the alternative is a society which chooses to remain childless even though its own members would have a higher quality of life over-all if they had children and its children would have lives worth living and no other society with a higher quality of life than Z is possible. The conclusion that we should bring about Z is still repugnant, but it is less repugnant than the alternative.

By contrast, total utilitarianism endorses bringing about a population with a lower quality of life whenever there would be more total utility in that society than in a smaller society with a higher quality of life but less total utility. It directly and unconditionally implies the Repugnant Conclusion.

Average utilitarianism can meet the challenge from narcissism. Moreover, from a practical point of view, average utilitarianism answers one of the major concerns of those philosophers who have espoused its rivals, total utilitarianism and the compromise view - and answers it satisfactorily, as they cannot. Most philosophers who believe in total utilitarianism or a compromise between total and average utilitarianism do so, at least in part, because they believe that numbers of people count for something morally. But, as we saw, the total utilitarian insistence on adding people up to the point of 0 marginal utility leads directly to the Repugnant Conclusion, and the compromise view involves irrationality or a chimerical substantive standard of ideal well-being. Yet philosophers complain that numbers of people do not matter at all to an average utilitarian: if average utility will be maximized by eliminating all but one happy person average
utilitarianism says only that one person should exist; and this is morally repugnant in itself. An average utilitarian can reply, however, that in actual practice numbers of people count do count morally. They count just up to the point where average utility begins to fall because of their addition below what it would otherwise be; and it will begin to fall when we surpass that point where the greatest number of people possible live at the level farthest removed from subsistence. This will not in practice constitute a very small society.

X. Elimination and Replacement: the Insufficiency of Utilitarianism

There is still one great problem with average utilitarianism, however, as Parfit and others have argued.

Suppose that society is rapidly approaching the point at which the maximum number of people can live at the level of welfare the farthest removed from subsistence, or is even beyond that point. Here average utilitarianism reaches what is perhaps indeed its most critical problem, for a population policy based on average utilitarianism permits society to do whatever is necessary to maximize average utility. Theoretically it even allows one person or group of people to exterminate everybody else if that will raise average utility most, considering both the direct and indirect consequences of the action, as Sumners and Nozick, for example, have noted. And, if this is the case, it certainly allows people to harm others, perhaps by sterilizing them against their wishes, or condoning infanticide, or imposing severe penalties on people who have more than a prescribed number of children, if this will most increase the average quality of life of society, or prevent society
from traveling down the road of overpopulation. (13) More theoretically, but no less seriously, average utilitarianism as I have described it can justify A’s bringing into existence a slave population at the level of Z so long as it could argue that its own quality of life would be less without the newcomers than with it.

This grave type of challenge to average utilitarianism argues, however, against the means of raising average utility; and the problem of means to the utilitarian end is a problem of the sufficiency of utilitarianism itself as a moral theory, not one of average as opposed to total utilitarianism. We can see this when we consider that total utilitarianism also allows the murder of people already alive if they can be replaced by people whose lives will be of better quality. (Nozick alludes to this possibility.) And it readily allows a society to bring a slave population into existence, as Gregory Kavka has argued. (14) First, I will show that the problem of means to the utilitarian end is a problem for any kind of utilitarianism – not just average utilitarianism. I will then consider the implications of this finding.

Parfit has an ingenious argument of this type against the principle of average utility. I will argue that it actually constitutes an argument against utilitarianism itself. I will call this the argument from mere elimination.

Suppose, Parfit argues, that at some point in the future there exist only two different population groups – the French and everybody else. Everyone in both populations is better off than any of us are today, but the French are far better off than anyone else. Now suppose a disease comes along that sterilizes all the rest of the world’s population except that of France. Everyone except the
French ceases to exist. Because the high level of welfare of the French depended in part on interaction with the rest of the world, the quality of life of the French falls slightly below that they previously enjoyed. Nevertheless, because it was so extraordinarily high in the first place, it remains above the level previously enjoyed by the rest of the world. How can this state of affairs, which is worse for everyone involved, be better than that which previously existed?, Parfit asks. Yet from an average utilitarian point of view it is better: the average quality of life of those who survive is actually higher than the average quality of life of those who previously existed. The average utilitarian conclusion is counter-intuitive. Moreover, this problem does not plague total utilitarianism, which gives the intuitively correct answer: it is better that both the French and everyone else exist.

Parfit's argument against average utilitarianism cannot be defeated if we think mere elimination is bad or if we think it counter-intuitive to approve of a move where everyone loses utility on the grounds that average utility is greater. It is a genuine counter-argument to average utilitarianism in population policy.

But we can make a parallel argument in which total utilitarianism gives the intuitively incorrect answer and average utilitarianism the correct one, which will show that arguments of this type tell equally against both forms of utilitarianism, and are not merely problems of average utilitarianism.

Suppose the initial situation is exactly as Parfit describes it. There are two groups of people, the French and everyone else; and everybody who lives in either group is very much better off than anyone alive today, although the French are much better off than
anyone else. Instead of a sterilizing disease striking everyone else, however, the same disease strikes the French. The elimination of the French causes everyone else's quality of life to drop slightly. In order to maintain their quality of life at as high a level as possible they begin to reproduce at a faster rate. As the French die out they are replaced by even greater numbers of everyone else at a lower level of quality of life than before, but not so low that the loss in quality of life is not compensated for by the gain in quantity of worthwhile life lived. The mere elimination of the French plus the mere addition of children to the remaining society counts as a gain from a total utilitarian point of view since it is possible to replace the French with more people than before at a smaller loss in quality of life than gain in quantity of life, even though every individual is worse off than in the first state of affairs. This is counter-intuitive for the same reasons as Parfit's example: a society where everyone is worse off is not intuitively better than one where everyone is better off; and mere elimination plus mere addition, or replacement, seems just as bad as mere elimination alone.

Since my example is no less plausible than Parfit's - being a variation on it - the two examples taken together - mere elimination and mere elimination plus replacement - must either tell against both utilitarian principles or against neither. My claim is that they tell against both.

It is cold comfort, however, to be able to argue that both forms of utilitarianism are tarred with the same brush. If average utilitarianism is insufficient to allow us to justify all and only those population policies which are intuitively sound we have to
decide what to do.

We have three alternatives.

First, we might want to argue that the problem is not the insufficiency of utilitarianism - that there is something wrong with the average principle that might be corrected by devising a new principle of utility. We have already shown, however, that it is unlikely we can find a principle of utility which maximizes utility in a more intuitively acceptable way than average utilitarianism. But indeed if the real problem here is the insufficiency of utilitarianism to satisfy our complete notion of morality in the area of human welfare, we would not expect revision of the principle to work. The failure of revision would be confirmation of the conclusion that the problems of utilitarianism are not all problems of the measurement of utility.

Second, we could reject the average utilitarian principle. But this move seems rash, considering that average utilitarianism both solves the paradox of future generations and dictates exactly what we ask of a population policy - a better life for future generations. Moreover, no moral principle that is not end-directed can justify utility maximization as a moral goal. Since it seems clear that when we set moral population policies we do have the moral goal of making people better off than otherwise and that the principle of average utility best expresses this goal, I will assume that these strengths of average utilitarianism give us reason to retain it.

Finally, we can place our average utilitarian concerns within a larger moral context. I will argue that this strategy is the best. But if we make this claim how can we possibly respond to
Parfit? Isn't it perfectly obvious that this principle directs us to prefer a counter-intuitive alternative to an intuitively sound one? Well, yes and no.

The principle of average utility says that the state of affairs where no one exists except the French is a better state of affairs than that where the French exist at a very high quality of life and the rest of the world at a somewhat lower quality of life and that the average quality of life matters morally; but unless we are simple average utilitarians - for whom utility maximization is a necessary and sufficient aim of morality - we can accept the claim that the latter state of affairs is preferable to the former in terms of utility without agreeing either that average utility is all that matters morally or that a mere increase in average utility is all that is going on morally in the move from the former state of affairs to the latter.

Yes, mere elimination is repugnant to our moral sensibilities; and no, we could not morally undertake it if we were in a position to do so. But our objection is not to the maximization of utility by the change in states of affairs. Our objection is to mere elimination itself - even when supplemented by replacement. This is the case even though in Parfit's case no one chooses mere elimination. We are repelled by it because in the real world we associate the elimination of persons with suffering and loss. We cannot believe that utility maximization - although good in itself - is all there is to morality.

I claim that what is required to make a utilitarian-based population policy morally acceptable is its subordination to a comprehensive principle that precludes achieving the end of utility
maximization by counter-intuitive means, including mere elimination and mere elimination plus replacement. Such a principle would not figure in Parfit’s mere elimination case because there would be no opportunity to employ it. But it would meet our objection to mere elimination where we do have a choice — i.e., where our morality makes a difference. More than that, it would meet Sumner’s and Nozick’s objections to the maximization of average utility through the murder of worse-off people and Kavka’s objection to the maximization of utility through the addition of a slave population.

The required comprehensive principle of means must, it seems, be one which requires that we respect people’s rights. If it exists and takes priority over utility maximization, it must be a principle which determines right action from an essentially non-utilitarian standpoint without denying that utility maximization is a moral goal. (15)

XI. Conclusion

My argument in this chapter has been in the first place an attempt to establish the claim that utilitarianism — and not just any form of utilitarianism but average utilitarianism — is essential to the moral justification of population policy. It has been in the second place an attempt to establish the insufficiency of average utilitarianism to determine a moral population policy. If these arguments are correct, average utilitarianism offers necessary but not sufficient justification for a moral population policy. We need the principle of average utility in our morality to justify choices like population policies. But we also need a principle which prevents the implementation of that policy by just any means so long as they are the most expedient.
In the next chapter I will introduce another problem of utilitarianism and will show that it also cannot be solved within the utilitarian framework. I will then argue for a comprehensive principle for moral decision-making which solves both problems while incorporating the insights of utilitarianism.
FOOTNOTES

1. Other types of population policies are possible besides moral ones. For example, we could design a population policy to create the greatest number of Aryans possible and eliminate everyone else in order to populate the world with a master race. I will be concerned in this paper only with population policies that are moral on the face of it, i.e. with policies that have as their goal the improvement of living conditions for future generations over what they would otherwise have been.


4. There are a number of issues involved here. A particularly important issue arises over the question whether the notion of benefit itself requires that someone be made better off than he would otherwise be. If so, future people cannot benefit from choices that affect them, since they cannot be made better off than they would otherwise be. For the purposes of this chapter I will assume that future people do benefit from choices that affect
them. For the argument for this view cf. Chapter I, "Moral Choice and Future Generations."


8. Since in a case where future people gain less than present people lose by a choice, the person-affecting principles will counsel against the addition of new people, it is possible to use either principle in selected cases and reach the impression that the principle will not lead to the Repugnant Conclusion or that it has contradictory implications. These problems are only apparent, however. They can, as Parfit claims, be avoided by giving both person-affecting principles a more complicated form on which they lead to the Repugnant Conclusion. Cf. Parfit 1982, p. 146. I discuss these principles at length in my M. A. thesis, "The Paradox of Future Generations," Rice University, 1984.

especially L. W. Sumner, Sikora and Barry 1978, p. 103ff.

10. We make other moral determinations as well — for example, whether this person will predictably have a miserable life himself although he makes others very much better off. This introduces a moral complication that I think need not concern us here. From a social point of view, which is the point of view of utilitarianism, we seem to justify bringing children into the world by their usefulness to society.

11. L. W. Sumner alludes to this problem. Cf. Sikora and Barry 1978, pp. 102-103. He does not specifically tax average utilitarianism with favoring narcissism, however, but rather with favoring the existence of present people over future people. Jefferson McMahan has responded to Sumner’s argument by arguing — correctly — that future people do not exist to be discriminated against and therefore that average utilitarianism cannot favor present people over them. Cf. Jefferson McMahan, "Problems of Population Theory," Ethics no. 92, October, 1981, pp. 96-127. I think the thrust of the challenge is somewhat different from what both Sumner and McMahan believe and that it goes beyond the question of discrimination against future people. Thus it requires a further response in addition to McMahan’s.

12. It is easy, but mistaken, to assume that average positive utilitarianism implies average negative utilitarianism. That is, it is easy to assume that if we should maximize the average quality of life when our choices will give people lives worth living, we should minimize the negative average quality of life when under all our available options people will exist and have lives not worth living.
If average utilitarianism did indeed have this effect it would have counter-intuitive implications for population policy. For negative utilitarianism implies that, if we have alternative choices, one of which will bring a billion people into existence with an average quality of life of 999,999 negative utils while the other will bring one person into existence with a quality of life of one million negative utils, we should choose the former alternative even though the total amount of suffering would be immensely greater. This is very counter-intuitive, as Parfit has argued. Cf. Parfit 1984, p. 406ff.

The answer to this problem is simply to realize that positive average utilitarianism does not imply negative average utilitarianism. The view that it does neglects the fact that no life at all is better on the average utilitarian principle than life not worth living. What the example really shows is that negative average utilitarianism does not offer acceptable moral justification in cases involving human suffering. It does not show that positive average utilitarianism offers unacceptable moral justification in cases involving human well-being.


14. For the argument that total utilitarianism, at least, allows bringing someone into existence to be a slave, cf. Kavka 1982, p. 100ff.

15. Protected rights could obviously not include a right to reproduction.
III. PREFERENCE UTILITARIANISM

I. The Argument

Sound definitions of the good and the right are essential to any moral theory; yet no moral theory so far advanced has managed to be persuasive both as to what is good and as to what is right.

Preference utilitarianism offers a particularly strong and clear definition of the good as the satisfaction of personal preferences and of the right as the maximization of personal preference satisfaction. These argue in its favor as a theory of moral choice.

But preference utilitarianism has paradoxical implications. Specifically, the satisfaction of some preferences—such as a preference for sadism—is intuitively bad. Yet on the preference utilitarian definition of the good this intuitively bad preference satisfaction is just as good as intuitively good preference satisfaction. Moreover, since utilitarians call a choice right if it maximizes good consequences, a preference utilitarian must call any choice right which maximizes personal preference satisfaction, no matter how intuitively bad the satisfaction of those preferences is. I will call this crucial problem for preference utilitarianism "the paradox of bad preference satisfaction."

The natural assumption is that we can solve the paradox of bad preference satisfaction by finding a means for identifying intuitively bad preference satisfaction and excluding it from counting as good. This approach has indeed been tried by several of the most eminent utilitarian philosophers. I will argue, however, that no criterion of bad preference satisfaction can be formulated that will keep intuitively bad preference satisfaction from counting
as good and at the same time allow us to retain the definition of
the good as the satisfaction of personal preferences. Moreover,
even if a criterion of bad preference satisfaction could be devised,
the utilitarian theory of right action could not satisfactorily
accommodate it.

I will argue that we can solve the paradox of bad preference
satisfaction, not by rejecting the notion that the good consists in
personal preference satisfaction, but only by replacing the
utilitarian definition of the right in favor of a definition which
gives the satisfaction of certain preferences priority over the
satisfaction of others.

If this strategy succeeds, we will have made progress
towards a sound moral theory beyond preference utilitarianism - a
theory which, if it cannot allow us to call the satisfaction of some
preferences intrinsically good and the satisfaction of others bad,
can allow us to avoid calling the satisfaction of all preferences of
equal intensity equally right.

II. Why Preference Utilitarianism

A. The strengths and weaknesses of previous utilitarian theories

Utilitarianism is essentially that moral theory which holds
that the object of morality is the maximization of the good, or
utility. Although it has been disputed by important moral theories,
this core notion of utilitarianism seems sound. At least it should
not be dismissed out of hand. But utilitarianism has had a great
deal of trouble coming up with a satisfactory definition of utility.

The earliest utilitarian theory, that of Jeremy Bentham,
defined the good as pleasure and the right as the maximization of
pleasure. Utilitarianism of this type is called "hedonistic
utilitarianism." It is a version of eudaemonism. Eudaemonistic moral theories all interpret the good as happiness and the right as the maximization of happiness. Hedonistic utilitarianism further interprets happiness as pleasure. It calls for the maximization of the pleasure experienced by those affected by a choice.

Many utilitarians have not rested content with hedonistic utilitarianism, however, since it seems too narrow and in cases plainly false. Bentham's theory implies that 'all things being equal, pushpin is as good as poetry.' Yet many things seem good to us out of proportion to the pleasure they afford; and it does not seem to be the case that any pleasure is as good as any other of equal intensity.

One major alternative to hedonism - ideal utilitarianism - was proposed by G. E. Moore in response to these problems. Moore argued that the good consists in states of mind of intrinsic worth and that the object of morality is to maximize these states of mind. While it was plausible to Moore that every intrinsically good mental state must contain at least some pleasure, it was evident that the concept of pleasure did not fully capture the notion of the intrinsically good. Moore did not, however, specify any general way to identify all intrinsically good mental states. Quite the contrary. He explicitly denied that there was any general characteristic of intrinsically good mental states other than their being good - which is question-begging as a definition of the good, or utility. (1) Nor did Moore give a means of measuring the intrinsic goodness of states of mind vis a vis each other and vis a vis different occurrences of the same type of state. His ideal utilitarian theory is therefore seriously incomplete.
A second alternative to hedonism was offered by John Stuart Mill. Although developed earlier than Moore's theory, Mill's is less problematic. Mill believed like Bentham that morality consisted in the maximization of pleasure; but unlike Bentham he believed that there were higher and lower pleasures. His theory called for the maximization of pleasure in those people affected by choices, but he assumed that people who had been exposed to both the higher pleasures and the lower ones would all choose more or less the same thing and that they would choose the higher pleasures over the lower ones.

The difficulty with Mill's view is that there is no real reason to think people, if exposed to both pushpin and poetry, would settle on poetry as the higher pleasure, as Mill apparently thought they would. Also, failing consensus as to the hierarchy of pleasures, Mill, like Moore, would be forced to devise a criterion for determining which pleasures were to count as higher than others and by how much. And there is no telling what such a criterion would look like.

Nevertheless, in principle Mill's theory allows people to decide for themselves what gives them greater pleasure by mere acquaintance with the various pleasures. And in thus locating the good in people's personal estimations of their own happiness without restricting happiness to simple intensity of pleasure, Mill's version of utilitarianism lacks the principal drawbacks of both hedonistic and ideal utilitarianism. It is, instead, a precursor of preference utilitarianism. (2)

Bentham's, Moore's, and Mill's versions of utilitarianism all, however, suffer from a fatal flaw in that all give mental-state
definitions of the good. That is, all locate the good in the production of certain mental states, whether those mental states are states of maximal pleasure or mental states of maximal intrinsic worth. And, as Robert Nozick has pointed out in *Anarchy, State, and Utopia*, no mental state theory of the good can ever suffice.

Suppose, Nozick argues, we could hook ourselves up to an experience machine that would give us whatever experiences we chose. If we wished to write the world’s greatest novel we would have the experience of having done exactly that. If we wished to live lives of eternal bliss, the experience machine would give us eternal bliss. But nothing in the world would be changed by our experiences; they would be entirely manufactured by the machine. It seems clear that we would not call the results of our hooking up to the experience machine good. We think of the good as consisting in more than mere experiences of certain types. Nozick concludes that certain specific things beyond our experiences matter to us, namely actual achievement, our being a certain way and not merely thinking we are, and our having actual contact with a "deeper reality." No form of mental state utilitarianism can provide an adequate moral theory because all ignore the connection between mental states and reality that we actually require of the good. (3) I am not sure we have to go so far as Nozick in defining utility, namely so far as to require contact with a deeper reality or to require our being a certain way. But it does seem that the achievement of the good requires not merely that we have the experience of things being a certain way but that they actually be that way.

B. The strengths of preference utilitarianism

Of all forms of utilitarianism so far devised only one meets
Nozick's objection - preference utilitarianism. In preference utilitarianism the satisfaction of the preferences people have - their likes or desires - is considered intrinsically good and the satisfaction of their aversions intrinsically bad. And preference satisfaction is all that is irreducibly good or bad. Personal preferences exist in the mind as likings or desires, just as pleasure and "mental states of intrinsic worth" exist in the mind. But, unlike pleasure and mental states of intrinsic worth they themselves are considered neither good or bad. There are no restrictions of any kind on what constitutes "good" preference satisfaction. The good simply is personal preference satisfaction, by definition. Nor do preference utilitarians call the production of a preference right - as hedonistic utilitarians call the production of pleasure right, for example. Preferences look out beyond themselves to the world. They are choices that a certain state of affairs exist. Preference utilitarians call the satisfaction of people's preferences by states of affairs good. And they call the maximal production of such states of affairs right.

Not only does preference utilitarianism alone meet Nozick's objection to mental state utilitarianism, but it also lacks the other defects of its precursors. Preference utilitarianism retains the notion that the good has a mental component, namely happiness. But that component is neither overly restrictive, as in hedonistic utilitarianism, nor mysterious, as in ideal utilitarianism. Nor are the higher pleasures assumed to be the same from person to person, as in Mill's theory. Rather, preference utilitarianism gives us a way of defining happiness which allows for the intuitively sound
notion that what happiness is differs from person to person and is best determined by whoever who experiences it. Moreover, we can measure the happiness generated by a choice in terms of preference satisfaction.

These strengths of preference utilitarianism are unmatched by any other utilitarian theory. They seem to give us sufficient reason to stay with preference utilitarianism to the extent we are utilitarians at all.

And there are strong reasons to prefer a utilitarian, or utilitarian-type, moral theory. First, the notion that the right consists in the maximization of the good seems sound, as does the utilitarian notion that the good consists in the actual satisfaction of people’s conceptions of their own happiness. Moreover, preference utilitarianism, uniquely among contemporary moral theories, offers conceptions of the good and the right that are, at least prima facie, clear, complete, and measurable – which is to say it appears to be the only extant well-defined moral theory.

But at this point we must ask whether there are no weaknesses of preference utilitarianism which would preclude its use as a moral theory. (4)

III. The Paradox of Bad Preference Satisfaction

Unhappily, preference utilitarianism suffers from a grave weakness, namely its paradoxical implications. For if the good is defined as personal preference satisfaction, it must include the satisfaction of people’s actual preferences no matter how intuitively bad the satisfaction of those preferences is. The right then consists in the maximization of preference satisfaction, no matter which preferences are satisfied. (Analogues to the same
problem, incidentally, plague other forms of utilitarianism: if the good is the production of pleasure, then pleasure derived from a "bad" source is just as good as pleasure derived from a "good" source.)

Amartya Sen has given a striking illustration of the paradox of bad preference satisfaction. In Sen's case there are a miserable policeman (P) and a romantic dreamer (R): "The dreamer has a happy disposition ("the future is ours") and also happens to be rich, in good health, and resilient, while the policeman is morose, poor, ill, and frustrated, getting his simple pleasures out of torturing." As Sen sets up the case, the total net utility - or preference satisfaction - in the state of affairs where P tortures R is greater than the net utility in the state of affairs where he does not. Preference utilitarianism must call the satisfaction of the miserable policeman's preferences just as good as the satisfaction of those of the romantic dreamer. Moreover, as Sen points out, since more personal preferences are realized in the state of affairs with torture than are realized in the state of affairs without torture, the utilitarian theory of the right calls the satisfaction of the policeman's preferences right. (5) But the satisfaction of the miserable's policeman's preferences is intuitively bad while the satisfaction of those of the romantic dreamer is intuitively good. It seems false to say that the satisfaction of a preference for torture is good and that in choosing the state of affairs with torture we are doing the right thing.

The alternatives are to reject preference utilitarianism or to attempt to solve the paradox. The argument I have given for preference utilitarianism suggests that we should not lightly
discard it. It seems better to seek a solution to the paradox of bad preference satisfaction.

Several eminent utilitarians have implied that we can solve the paradox of bad preference satisfaction by developing a criterion of good and bad preference satisfaction and excluding bad preference satisfaction from counting as utility. In the following sections I will show that attempts to solve the paradox of bad preferences by developing a criterion of good and bad preference satisfaction must fail. If we are unwilling to give up preference utilitarianism we must find another way to save it.

IV. Rational and Irrational Personal Preferences

The most popular approach to the paradox of bad preference satisfaction has been the attempt to construct a criterion of good and bad preference satisfaction on the basis of individual rationality. Philosophers who take this approach to the paradox argue that not all personal preference satisfaction is good but rather only the satisfaction of people’s rational preferences; hence morality requires us to satisfy only rational personal preferences. They assume that if we satisfy only rational personal preferences we will not produce any intuitively bad preference satisfaction. I will show that this is not the case.

There are certain senses in which everyone agrees that some personal preferences are irrational. Thus their satisfaction should not count as good; nor do they place on us any obligation to satisfy them by our choices. Preferences are irrational if, as Hume noted in a famous passage, they are based on false beliefs or propose ends insufficient to achieve their own ends. They are irrational also, as Derek Parfit has noted, if they draw arbitrary distinctions. (6)
For example, it is irrational to prefer jam tomorrow and jam yesterday but never jam today.

These criteria of irrational preference, however, offer no more than a partial solution to the paradox of bad preference satisfaction. The miserable policeman’s preference for torture, for example, would be irrational if it derived from the false belief that torture actually satisfied the romantic dreamer’s preferences. But the policeman may be well aware that the romantic dreamer dislikes torture, and he may still prefer to torture him or anyone else, and this preference will count as rational under both Hume’s and Parfit’s criteria.

But two strategies for defining irrational preferences and excluding their satisfaction from counting as utility do seem to offer a solution to the paradox. Richard Hare and John Harsanyi have argued, independently of each other, that the satisfaction of those preferences which are not members of the set of preferences a person would have if he were well-informed and perfectly prudent are irrational and should not be counted as utility. (7) Richard Brandt agrees with Hare and Harsanyi, but he also argues that the satisfaction of intrinsically irrational preferences as defined by cognitive psychotherapy should not count as prudent.

I will argue that Hare and Harsanyi are right about one thing: the satisfaction of imprudent preferences is irrational and should be excluded from counting as utility. But imprudent preference satisfaction does not correspond to intuitively bad preference satisfaction. And Brandt’s theory of intrinsically irrational preferences is unsound. So neither strategy offers a genuine solution to the paradox of bad preference satisfaction.
A. Imprudent Preferences

As utilitarians, Hare and Harsanyi argue that utility consists in the satisfaction of people's prudential preferences. But what is prudence?

We can arrive at a nice definition of prudence if we enlarge upon utility theory as developed by John von Neumann and Oskar Morgenstern. Utility theory takes its inspiration from the plausible notion that rational people try to maximize their personal preference satisfaction. On this theory personal preferences are simply the choices people make when confronted with alternatives. They are neither rational nor irrational in themselves. But if a person has a continuous set of preferences, a unique numerical value can be determined for each preference.

Von Neumann and Morgenstern argue in their seminal work on utility theory, *The Theory of Games and Economic Behavior*, that we can determine the value of any preference held by a single individual by asking him to rank his preferences *vis a vis* each other. They prove that through a process of conducting wagers with himself about his various preferences a person can arrive at a cardinal value for each of his preferences, assuming only that his set of preferences is continuous and complete, that is, assuming that he prefers something in every situation and that his preferences are consistent with each other. Each person should then have a utility function determined by his own preferences. Indeed, individual rationality is defined by von Neumann and Morgenstern as the individual's maximization of his own utility function when he is confronted with a choice. (8) (This does not mean that a person must prefer only states of affairs that are better for *himself* than
alternative states of affairs; it means that, whatever value he assigns to alternative states of affairs, on whatever grounds, he must prefer that alternative which he himself ranks highest. This is a very important feature of the theory.)

Practical rationality with respect to any choice then consists in the maximization of the numerical value of one's own personal preference satisfaction. Utility theory thus formally defines individual rationality with respect to a choice as a function of personal preferences which are themselves neither rational nor irrational.

Suppose utility theory is right and we can define individual rationality in any given situation as the maximization of one's personal preference satisfaction. (9) Now let a well-informed person estimate the preference satisfaction of all alternative courses of action open to him, counting equal preference satisfaction equally regardless of when the preferences will actually be satisfied, but discounting for the probability of their satisfaction. And let him do this every time he is confronted with a choice. We can now define a perfectly prudent person as one who always maximizes his own personal preference satisfaction indifferently over time. (10) A utilitarian who maximizes only rational personal preference satisfaction will maximize the utility curve of all the prudential preferences of the individuals affected by his choice.

The problem is that since the satisfaction of any preferences that contribute to the maximization of personal utility counts as the satisfaction of rational preferences and the satisfaction of rational preferences counts as good, any personal
preference satisfaction that contributes to personal utility maximization cannot be excluded as bad, no matter how intuitively bad it is.

It might be good, for example, for the Mongol hordes to ravage central Asia. Indeed it would be, if the maximal satisfaction of the Mongols' personal preferences depended on their devastating the villages of the steppes and the satisfaction of their personal preferences outweighed the thwarting of the personal preferences of the villagers. But the satisfaction of preferences for torture, rape, and pillage is intuitively bad. So the attempt to restrict good preference satisfaction to prudential preference satisfaction does not exclude intuitively bad preference satisfaction from counting as good in this case. Nor need it exclude even the intuitively bad satisfaction of the preferences of Sen's miserable policeman with which we started. His preferences too may be perfectly prudent and their satisfaction therefore good.

We can and should adopt Hare's and Harsanyi's conclusion that utility consists only in the satisfaction of rational personal preferences - i.e., prudential preferences - but this will not solve the paradox of bad preference satisfaction.

B. Intrinsically Irrational Preferences

In *A Theory of the Good and the Right*, Richard Brandt accepts the claim that the satisfaction of preferences, or desires, is irrational if it interferes with the achievement of the entire set of outcomes a person most wants - that is, if it is imprudent in approximately the sense just given - but he argues that preferences can also be intrinsically irrational. Brandt defines a fully rational person as one who maximizes the satisfaction of his
intrinsically rational preferences, i.e., those preferences which would survive cognitive psychotherapy. (11) A preference fails to withstand cognitive psychotherapy, hence is irrational, to the extent it extinguishes or diminishes under repeated confrontations with the facts in "value-free reflection." (p. 113)

It may seem at first glance that Brandt is simply agreeing with Hume that a preference based on a false belief is intrinsically irrational. But Brandt goes beyond Hume to characterize as irrational not only preferences based on false beliefs but, for example, certain preferences produced by "artificial desire-arousal in culture-transmission" (p. 116); those produced by "generalization from untypical examples" (p. 120); and those characterized by exaggerated responses due to early deprivation. (p. 122)

Preferences produced by generalization from untypical examples can be interpreted as expressing a type of false belief because they are produced by faulty logic. Like Hume's and Parfit's criteria of irrational preference, this criterion of irrational preference is not controversial, but it will not solve the paradox of bad preference satisfaction. The miserable policeman, for example, need not generalize defectively from a past experience in which he enjoyed torturing someone in order to prefer torturing people. Any solution to the paradox of bad preference satisfaction through the intrinsic irrationality of preferences must therefore stem from the other types of alleged intrinsically irrational preference.

Let us look at the notion that some desires aroused by artificial culture-transmission are intrinsically irrational. It should become apparent that the same arguments which can be brought
against this notion can also be brought against the notion that preferences which stem from early deprivation can be intrinsically irrational.

Take, for example, Brandt's account of the desire for achievement. Brandt believes that there is a native pleasure in doing difficult things well. This pleasure is not irrational. But the preference for achievement, or superiority in relation to some standard of excellence, stems from middle-class parents' urging their children to greater future success through reward and punishment. It is artificial. And at the level "where it can be reached only by a mixture of threats, misinformation, and good company" it is irrational and a candidate for extinction or diminution by cognitive psychotherapy. (pp. 118-120) Brandt argues that a person can determine the level at which his own preference for achievement becomes irrational by asking "where the total activity is leading." (p. 120)

Suppose I am a philosophy graduate student with an innate desire to do moral philosophy well. By Brandt's definition this is an intrinsically rational preference, or desire. But I also have a desire to excel in some more practical field - namely law. My desire to excel in law is, however, artificial. It was inculcated in me by bourgeois parents who rewarded interests likely to make me self-sufficient in the long run and discouraged interests that did not have this tendency. And it is sustained by the threat of poverty and the pressure of society that I support myself: I still have no natural preference for law. But if I excel in law I will be able to satisfy the greatest number of my intrinsically rational preferences. I will be able to support my family, for example. And
if I pursue philosophy I will not be able to satisfy the greatest number of my rational preferences: the task, although intellectually rewarding, interferes with my ability to provide for my family and to pursue a number of my other interests.

Since my desire to excel in law is aroused by artificial culture-transmission, it is, according to Brandt’s theory, a candidate for extinction by cognitive psychotherapy, at least at a certain level. I want to determine that level. I confront my desire to excel in law with the facts. But what are the facts? Social pressure and the threat of poverty are facts and so is my lack of a natural inclination towards the law; and my mere recognition of the fact that my desire to excel in law is artificially induced does not alter the other facts. But the confrontation of my desire to excel in law with the facts of life should tend to strengthen it, while my confrontation of that same desire with the facts of its origin and its lack of comportment with my most intense innate desire – the desire to do philosophy – should tend to weaken it.

On reflection, I give up the desire to excel in law to the extent that it is supported by misinformation or baseless threats or extrinsic approval or disapproval without consequence. But this is only to say that I give it up to the extent it is sustained by falsehood or by irrelevant considerations (assuming I consider approval and disapproval intrinsically irrelevant; I may not, and I may have grounds for considering these facts not to be irrelevant). So far I am merely agreeing with Hume. I ask myself to what extent beyond this my desire is irrational.

Now I reflect that achievement in law is essential to my
optimal preference satisfaction. This fact alone gives me reason to retain my preference for excelling in law in sufficient measure to maximize my personal preference satisfaction. And it also gives me reason not to retain it beyond that point. Otherwise I will satisfy this desire to the detriment of my over-all preference satisfaction, which is irrational in Brandt’s view and in utility theory. But this makes otiose the notion that my preference for excelling in law is irrational in any sense other than Hume’s or Hare’s and Harsanyi’s. I can in fact determine the extent of rationality of my preference for excelling in law simply by seeing what degree of the preference for achievement is required to enable me to maximize my over-all preference satisfaction.

Next I look at my desire to do philosophy. According to Brandt this desire is intrinsically rational. But its satisfaction interferes with my over-all preference satisfaction. This leads to a dilemma. Since the desire is intrinsically rational, its satisfaction should be intrinsically rational. But it can be satisfied only by my failing to satisfy the desire to excel in law, a desire essential to the achievement of my rational goal of maximal preference satisfaction and to that extent a rational desire. Since I can satisfy my desire to do philosophy only by not satisfying a desire whose satisfaction is essential to my rational preference satisfaction, the satisfaction of my desire to do philosophy is irrational; so the desire itself seems irrational. But this contradicts Brandt’s claim that it is rational.

I have two options. I can give up the notion that my desire to do philosophy is intrinsically rational. Otherwise I would have to say that my desire to do philosophy should be satisfied even
though it will lead to less satisfaction of my rational preferences over-all - which is counter-intuitive. Or I can give up the notion that the satisfaction of intrinsically rational desires must itself be rational. But if I give up this notion and concede that it can be irrational to satisfy a rational desire the notion of rational desire becomes otiose. What matters is what it is rational to do, not what it is intrinsically rational to desire; and it is not rational to satisfy my desire to do philosophy. Thus the claim that my desire to do philosophy is intrinsically rational is either false or otiose. Either way it can be disregarded.

The same argument that can be brought in this case against the concept of intrinsically rational and irrational desires as Brandt defines them can be brought in others. To the extent preferences are based on false beliefs they are irrational on grounds already advanced by Hume. The exclusion of their satisfaction from counting as utility will not solve the paradox of bad preference satisfaction. But to the extent intrinsically irrational preferences as Brandt envisions them are not based on false beliefs their satisfaction can be determined to be rational exactly insofar as required by prudence. Otherwise the concept of intrinsically rational preference satisfaction will interfere with the concept of prudential preference satisfaction - with counter-intuitive results. But this makes the claim that there are intrinsically irrational preferences based on true beliefs whose satisfaction is bad either false or otiose.

There is still another objection that can be brought against Brandt’s theory of irrational preferences as those which would not survive cognitive psychotherapy. It would clearly be impossible in
many cases to determine which preferences were intrinsically irrational. People might be able to determine the intrinsic rationality of their own preferences. They would often have no way at all of determining the rationality of other people’s preferences. So a practical utilitarianism that called for maximizing the satisfaction of people’s intrinsically rational preferences as determined by cognitive psychotherapy could not be constructed. And if such a utilitarianism could not be constructed, the attempt to solve the paradox of bad preference satisfaction by eliminating intrinsically irrational preferences from moral consideration must fail, a fortiori.

We should reject the claim that the satisfaction of only intrinsically rational personal preferences as Brandt defines them offers a solution to the paradox of bad preference satisfaction, just as we rejected the claim that the restriction of good preference satisfaction to the satisfaction of prudential preferences offered a solution. If we want to retain preference utilitarianism we must look elsewhere for a solution to the paradox of bad preference satisfaction.

V. Anti-Social Preferences

Harsanyi ultimately concludes that restricting moral preference satisfaction to the satisfaction of "true" or rational personal preferences will not solve the problem of bad preference satisfaction. Therefore he proposes that the satisfaction of anti-social preferences should be excluded from the utilitarian calculus. He calls for the development of a formal criterion for the determination and exclusion of "all clearly antisocial preferences, such as sadism, envy, resentment, and malice."
(Harsanyi 1977, p. 56.)

Now an anti-social preference must be defined on the basis of what is acceptable and unacceptable to society. And utilitarianism has its own criterion of socially good and anti-social preference satisfaction, if not of socially good and bad preferences. The satisfaction of a preference is socially good if it is the satisfaction of a preference actually held by someone in society; and it is anti-social if it thwarts the satisfaction of a preference of a member of society. (The satisfaction of some preferences is therefore good to a certain extent and bad to a certain extent.) This criterion of socially good and bad preference satisfaction derives from the notion that society is nothing more than the aggregate of its members. So what is socially good and bad must be determined by what the members of society take to be good and bad.

Any formal criterion of anti-social preferences other than the utilitarian one would have to deny the utilitarian criterion for determining what is and is not acceptable to society. Therefore it would have to deny utilitarianism itself.

But even if we could in principle devise some formal criterion of anti-social preferences, there is no reason to think that criterion would eliminate intuitively anti-social preference satisfaction from counting as good, any more than the rational criterion of good and bad personal preference satisfaction as prudential preference satisfaction eliminated all intuitively bad personal preference satisfaction from counting as good. The reason is the same in both cases: our objection to intuitively anti-social preferences is not formal but qualitative. The trouble with any
formal definition of bad preference satisfaction - personal or social - is that the results it gives will be intuitively acceptable or not according to the qualities of the preferences which get plugged into it.

Suppose, then, that we abandon the attempt to devise a formal definition of anti-social preferences and admit that what we really seek is a qualitative criterion of socially good and bad preferences. This may not be impossible to find, although the many previous unsuccessful attempts to give qualitative definitions of the good do not bode well for it. But this much is clear: a successful qualitative criterion of good and bad preferences could not be merely grafted onto preference utilitarianism to enable us to distinguish bad preferences from good ones.

The new qualitative definition of the good would have to replace the personal preference definition of the good, not modify it. Preference satisfaction would not be good simply because it was the satisfaction of preferences, or of rational preferences. Rather, the intrinsic goodness or badness of preferences would derive from their qualities. For example, a preference for art would be good because art is good; and a preference for sadism would be bad because sadism is bad. Therefore, a qualitative criterion of good and bad preference satisfaction - could one be devised - would not offer a solution to the paradox of bad preferences. It would destroy the essential feature of preference utilitarianism - its definition of the good.

Neither a formal nor a qualitative criterion of good and bad preference satisfaction can be devised that will allow us to solve the paradox of bad preference satisfaction and retain preference
utilitarianism. Rather, the notion that a utilitarian can devise a satisfactory criterion of bad preferences is chimerical. There is no such criterion. We must conclude that his only choice is to bite the bullet and accept the conclusion that all actual personal preference satisfaction is good by definition and should be included in the calculation of utility. The utilitarian definition of the good is ultimately, then, reductive. It reduces to no more than personal preference satisfaction.

But if the paradox of bad preference satisfaction cannot be solved by finding a criterion of good and bad preference satisfaction we must either totally reject preference utilitarianism or find some other way to modify it. I suggest that if we cannot declare the satisfaction of certain preferences bad we may nevertheless be able to call it wrong to satisfy those preferences. This strategy would require the modification of the utilitarian definition of the right rather than the modification of the utilitarian conception of the good. I will advocate this strategy. But first I will deal with another tempting - but I think mistaken - strategy for solving the problems of preference utilitarianism.

VI. The Exclusion of Bad Preference Satisfaction

It is a great temptation to believe that if only bad preference satisfaction could be identified and excluded from counting as good, or better yet, counted negatively, utilitarianism would always dictate intuitively moral choices. That is, it is a temptation to believe that if the good could be adequately redefined within a preference utilitarian context the right would fall into place. This belief is false. Even if a criterion of bad preference satisfaction was possible, it would not solve the problems of
preference utilitarianism.

Harsanyi has taken this strategy. He suggests that once we have devised a criterion of anti-social preferences we can simply exclude the satisfaction of those preferences from counting when we maximize utility.

This strategy will not work because the exclusion of bad preferences destroys the completeness and continuity of preferences. And the completeness and continuity of preferences are essential features of rational choice. Indeed, they constitute two of the conditions of rational choice recognized by Harsanyi. So this does not represent a viable option.

Let me elaborate. We saw that an individual can be said to make a rational choice if he has a complete and consistent set of preferences and maximizes his own preference satisfaction. And if he does this every time we said that he could be called prudent, and his preferences could be called prudential preferences. We accepted the arguments of Hare, Harsanyi, and Brandt that utilitarianism requires the maximal satisfaction of the prudential preferences of the people affected by a choice. Now we are supposing we have found a means of identifying anti-social preferences. And we are claiming that we want to keep those preferences from figuring in our calculations, even if they are prudential preferences. We want to know how a utilitarian can make a rational choice that will not satisfy any anti-social preferences.

Suppose that almost all of P's (the miserable policeman's) preferences meet our definition of anti-social preferences. We disregard these preferences. But now certain preferences look stronger than they really are. Suppose we disregard P's strong
anti-social preferences for torture and the like but we see that he
takes delight in smiling at the people he tortures as if to say he
didn’t really mean it. This preference is left as P’s only socially
acceptable preference.

We look, then, at P’s good preferences and we see only one:
the preference for smiling. Since we disregard his other
preferences we must count his one preference as one very worthy of
satisfaction. We have nothing to weigh it against so we assign it a
perfect value of 1 on a scale of 1 to 0. We of course weigh the
negative effects of the satisfaction of P’s taboo preferences as we
did before, subtracting them from the preference satisfaction R
enjoys in the state of affairs with torture. But we have placed
such a high value on the satisfaction of P’s preference for smiling
that the state of affairs with torture emerges as better than that
without.

Our social welfare function, or utility function, is
distorted. It is not accurately reflecting people’s prudential
preference satisfaction any longer, because it is not accurately
reflecting P’s prudential preference satisfaction. We have not
solved the problem of bad preference satisfaction.

A utilitarian who finally succeeded in identifying bad
preferences might, however, try counting only the negative effects
of those bad preferences. There is a tension implicit in this move
since bad preferences would represent people’s likes but they would
be counted as dislikes. This tension would again be reflected in
the individual’s utility function. Suppose P tortured R as before.
We would count the effect of the torture on R as a loss of utility,
just as before. But this time we would count it as a loss for P as
well. But for P it would not weigh against any particular
counter-vailing preference. That is, the satisfaction of P’s
preference for torturing someone would not thwart any measurable
preference of P’s. We would have no way of accurately determining
the intensity of the badness of the satisfaction of this preference
for P. P’s utility function would again be distorted.

Even if we could devise a criterion of anti-social or bad
preference satisfaction we could not solve the paradox of bad
preference satisfaction within the utilitarian framework.

VII. The Hierarchy of Preferences

If we cannot devise a criterion of intrinsically good and
bad preferences, however, we need not despair of retaining the
essential feature of preference utilitarianism, namely the notion
that the good consists in preference satisfaction. It may be that
we can solve the problems of preference utilitarianism by simply
giving up the notion that the right consists in the mere
maximization of preference satisfaction.

T. M. Scanlon has taken exactly this type of approach to the
problems of preference utilitarianism. Scanlon argues that although
morality does call for preference satisfaction, not all preferences
are equally worthy of being satisfied. He calls for a hierarchy of
socially good, better, and best preferences, or a "hierarchy of
social urgency." (13)

What Scanlon is arguing is that simple utility maximization
is not enough to determine the rightness of a choice. We need a way
to pick out which preferences should receive priority of
satisfaction. If we use Scanlon’s strategy, we will not be able to
call the satisfaction of any actual personal preferences bad, but we
may be able to place the satisfaction of some preferences very low in our scale of priorities. This notion of giving priority to the satisfaction of certain preferences over others while not denying that utility consists only in preference satisfaction is alien to utilitarianism. It, in fact, amounts to a redefinition of the right.

The trouble with Scanlon’s suggestion that morality requires a hierarchy of good, better, and best preferences is that not only does Scanlon not give a social hierarchy of preferences, but he does not even suggest what any criterion of social urgency would look like. And, lacking a standard, it is hard to see why we should depart from utilitarianism. For utilitarianism itself uses the plausible criterion of maximal preference satisfaction to determine the social urgency of preference satisfaction. And we may well want to ask, If I like capital punishment for murder and you like life imprisonment, who is to decide whose preference should prevail from a social point of view, and on what grounds? Moreover, one cannot avoid the suspicion that Scanlon’s hierarchy of social urgency would have to be a qualitative hierarchy. And, if that is so, the notion of qualitatively good, better, and best preferences would compete with the notion that the good simply consists in preference satisfaction, the better in more preference satisfaction, and the best in optimal preference satisfaction. Scanlon’s theory is incomplete and potentially unstable.

VIII. Conclusion

We are left then with a dilemma. We have a strong and serviceable definition of the good as the satisfaction of personal preferences. This definition, uniquely among extant definitions of
the good, is clear, well-defined, and measurable. But not only have we not solved the paradox of bad preference satisfaction; we have argued that there is no satisfactory formal or qualitative criterion which will allow us to call the satisfaction of certain personal preferences bad and the satisfaction of others good and still retain the preference utilitarian definition of the good. Yet clearly morality cannot call for us to give equal weight to the satisfaction of a personal preference for torture or murder and preference for smelling the flowers, so long as they are held with equal intensity.

In this chapter I can only indicate where I think the solution to the dilemma must lie.

I believe we should devise a more complex model of moral decision-making than utilitarianism offers. The new model would allow for recognition of a commitment to people that supercedes the commitment to the impartial satisfaction of the maximal number of personal preferences. It would allow us to give priority to the satisfaction of certain designated preferences, specifically those that accord rights to people. In effect, any preference satisfaction designated as a right would serve as a trump in the moral game.

If we can designate preferences like the preference to be free from torture as rights worthy of being given priority of satisfaction over preferences like the preference for torture, we should be able to avoid having to call the decision to satisfy a preference for torture right, even if we could not call the satisfaction of that preference bad. And other types of intuitively bad preference satisfaction could be dealt with in the same way.

But just as utilitarian decision-making gives intuitively
wrong results when the preference satisfaction it approves is intuitively bad, so a type of decision-making which designated certain preferences as rights and gave their satisfaction precedence over the satisfaction of other preferences would give intuitively wrong answers if it failed to give priority to the satisfaction of preferences that actually meet our conception of preferences worthy of priority.

We can predict, however, that the new model would at least preclude the approval of intuitively bad preference satisfaction in many cases where utilitarianism approves it. And it would allow us to keep the essential insight of preference utilitarianism — that the good consists in personal preference satisfaction. If so, it would mark an important stride forward in the construction of a well-defined and serviceable moral theory.

The next task is to see what such a model for moral decision-making would look like and what effect it would have on reducing the number of times we must call intuitively bad preference satisfaction moral.
FOOTNOTES


4. There is one serious charge against preference utilitarianism which I think can be satisfactorily answered. The charge has frequently been made that preference utilitarianism requires the interpersonal comparison of utilities — or the comparison of preference satisfaction from person to person — which is impossible. If this is so, the preference utilitarian definition of the good is incomplete and the preference utilitarian definition of the right as maximal preference satisfaction is incoherent. That is, a state of affairs may observably contain some preference satisfaction, hence count as good, but if it is impossible to determine how much good it contains it will be impossible to construct a theory of right action on the basis of the production of the good.

People who bring this charge against preference utilitarianism do not dispute the claim that a person can determine
the value of his own preferences. He does this by asking himself which outcome of a gamble he would take—the certainty of having a specific preference satisfied or an even chance between having either of two other preferences satisfied. In *The Theory of Games and Economic Behavior*, John von Neumann and Oskar Morgenstern prove that an individual can establish a unique value for each of his preferences by persistently repeating these personal gambles, or lotteries, thus ever narrowing down the intervals between one preference and another. (Cf. John von Neumann and Oskar Morgenstern, *The Theory of Games and Economic Behavior*, 3rd edition, Princeton, New Jersey, Princeton University Press, 1953.)

There are two objections against extending this method of preference measurement to other individuals besides oneself, however.

First, a common point is required to bring the different utility functions of different individuals into line with each other. I propose that we meet this objection by arbitrarily picking a preference and assigning it the same value for all persons. Perhaps it could be the preference for a drink of water.

The second objection is that it is impossible to tell how much more or less an individual would value the satisfaction of any particular preference vis à vis the satisfaction of his preference for a drink of water or the satisfaction of another of his preferences. This objection has, I think, been entirely overrated.

First, in order to determine how much the outcomes of choices mean to different people we make use of whatever information we have about people’s preferences. Often there is a great deal of external evidence as to what people actually prefer and as to their
relative preference for a given outcome. We may know, for example, that Mary has no winter coat while John does have one. We rank Mary’s preference for a coat higher than John’s — unless John lives in the Yukon and has only a lightweight coat and Mary lives in Miami and has a heavy sweater; and Mary tells us she is warm and John tells us he is cold; and so on.

A moral agent should inform himself as fully as possible about how each person who will be affected by his choice would himself rate a certain effect. That is, he should determine as fully as possible the value of the individual preferences of the people affected by a choice in their own estimation. Perhaps he can merely ask them how important a certain effect is to them. Perhaps he is in a position to observe their behavior — to see what they choose for themselves and what they reject and when. These observations will give him a more or less rough idea of the value of preferences to the people themselves. But this idea will never — or surely seldom — be complete. An agent will seldom or never be in a position to observe what all people affected by his choice actually will choose when presented with all the possible different alternatives.

To fill in the gap between observation and reality the moral agent must use what John Harsanyi has called a "similarity postulate." (Cf. John Harsanyi, "Morality and the Theory of Rational Behavior," Social Research 44, 1977, reprinted in Utilitarianism and Beyond, Amartya Sen and Bernard Williams, eds., Cambridge, Cambridge University Press, 1982, 50.) In my formulation (not Harsanyi’s), this postulate says, "To the extent that the preferences of others cannot be determined by observation of the
facts assume the similarity of their preferences to your own." The use of this postulate is quite innocuous because whenever it is contradicted by evidence to the effect that the preferences of the people affected by a person's choice are different from his own he will judge the preference satisfaction of the choice according to the demonstrable preferences of those affected, not his own.

Using this method of preference evaluation, a person can discover the value of interpersonal comparisons of utility by the same process as that by which he determines the value of his own preferences. He can conduct a mental wager between the preference satisfaction (according to the evidence, plus similarity if appropriate) if x receives a certain benefit versus the preference satisfaction if y receives the benefit. He can do the same thing in a negative sense with costs. And by conducting a series of such wagers he will fix the value of his choice as nearly as it can be fixed.

I think it is worth remarking here that the alleged difficulties afflicting the interpersonal comparison of utilities are not peculiar to interpersonal comparisons. They also afflict intertemporal comparisons of utility for a single person. That is, they afflict the comparison of the value of the satisfaction of an individual's own preferences over time.

The only virtue that a person's assessment of his own utility has over his assessment of the utility of others is the virtue of privileged access to his own preferences, so that his information as to their value is more nearly perfect. But even a person's measurement of the value of his own preferences remains to some extent imperfect, since we are all influenced by subconscious
preferences. And, in fact, whenever someone projects the effects of his own personal choices into the future - even the near future - he has to fill in the gaps in his knowledge of his own preferences in the same way as he fills in his knowledge of the preferences of others. In order for a person to know what he himself will prefer at any time other than the immediate present he has to use whatever evidence he has of what his preferences will be at that time. If he has no evidence they will be different he assumes they will be the same. Yet no one seriously challenges the argument that a person can assess the value of the same one of his own preferences at different times. Therefore we should not find it impossible to compare the value of the satisfaction of the same preference for different people.

Interpersonal comparisons of utility are in fact made every time we debate whether to do something for one person or for another, as we frequently do. So they are possible. And I believe I have shown how they are made and how they are essentially the same as the comparison of our own preferences over time, which no one challenges. So the problem the interpersonal comparison of utilities presents to preference utilitarianism is not insurmountable.


8. Cf. fn. 4 for the use of von Neumann and Morgenstern's technique for establishing the value of personal preferences. The establishment of the value of a personal preference is an essential step in the determination of a rational personal choice, i.e., a choice that maximizes one's own preference satisfaction.

9. The notion that individual rationality can be defined as a person's maximization of his own utility over time is, although traditional, very problematic for a utilitarian, even though Harsanyi himself uses this definition of individual rationality. Cf. Harsanyi 1977. The reason is that utilitarianism requires the impersonal satisfaction of the preferences of all the people affected by a choice, whereas, if a person is affected by his own choice, rationality may seem to require his putting his own preference satisfaction first. I will not attempt to resolve this very difficult problem, which was first expounded by Sidgwick in The Methods of Ethics, Book IV.

10. Derek Parfit has argued against prudence as descriptive of rational personal behavior on the grounds that a person does not retain the same identity over his lifetime; therefore, it is not
irrational for a person to fail to treat all parts of "his" life equally. (Parfit 1984, p. 317) The definition I have given of prudence can be adapted to meet Parfit's objection. We can impersonally count the preferences of all the people we can expect to live serially in our bodies and minds and for whose behavior we are peculiarly responsible without essentially changing our definition of individual rationality as maximal personal preference satisfaction.

11. Richard B. Brandt, A Theory of the Good and the Right, Oxford, Clarendon Press, 1979. Brandt does not define prudence exactly as we have defined it. He argues that a rational person maximizes "the whole set of outcomes he most wants." (p. 11; p. 111) "The whole set of outcomes a person most wants" is a highly ambiguous concept which Brandt explores, but it is close to prudence as given above except that Brandt himself equates prudence not only with the maximization of one's over-all preference satisfaction but also with the intrinsic rationality of the desires, or preferences, to be satisfied. Cf. p. 81.


IV. UTILITY AND CONSTRAINT

I. The Argument

Utilitarianism offers a clear, complete, and well-defined model for moral decision-making. These traits are essential to an adequate theory of moral choice. Yet utilitarianism fails to capture the way in which at our best we really do make moral choices. (1) Its model for moral choices is defective. In this chapter I will argue that a model for moral decision-making can be devised which corrects for the deficiencies of utilitarianism. We should accept that model as a picture of the real model for ideal moral choice.

Utilitarianism defines the good as utility, which it further interprets as happiness or as preference satisfaction; and it defines the right as the maximization of the good. The morality of a choice is determined by its consequences. Each unit of utility and disutility that will be caused by the choice is counted, and the units are summed to determine the net utility of the choice. A moral choice is right if and only if it maximizes the utility of all the people affected by it relative to the alternatives.

These basic structural characteristics of utilitarian moral theory lead to the following problems. (2)

First, utilitarianism justifies the use of intuitively bad means to the good end of utility maximization. If it is always right to maximize utility and wrong not to, it must be right to use the most expedient means to maximize utility, no matter how evil those means appear to be. Yet, in fact, we would call the use of evil means to the good end of utility maximization immoral.

A corollary of this drawback is that utilitarianism cannot
avoid giving counter-intuitive answers where bad consequences are
entrained along with the good so long as the good consequences
outweigh the bad; for it tells us to choose whichever of our
alternatives will produce the greatest net utility - regardless of
how much suffering we have to figure in to reach that net sum; but
often a choice that generates less utility but harms no one is
intuitively better than a choice that generates more net utility but
does harm someone.

Second, utilitarianism has not been able to solve the
problem of bad preference satisfaction. Most utilitarians today are
preference utilitarians. They define utility in terms of the
satisfaction of personal preferences. If we interpret utility as
personal preference satisfaction, however, utilitarianism must call
for the maximization of personal preference satisfaction, no matter
what the preferences that are satisfied actually are; and the
satisfaction of some personal preferences - such as a preference for
sadism - is intuitively bad. (3)

I will argue that there is a single answer to both types of
intuitive objections to utilitarianism as a theory of moral choice.
We do have the moral goal of utility maximization, or the
maximization of personal preferences, just as utilitarianism claims.
But we also believe that there are moral constraints on the means to
the achievement of this end. An adequate theory of moral choice
must allow for the maximization of utility only insofar as it can be
maximized within these moral constraints. I will show what these
constraints are and how they dictate a more complex model for moral
decision-making than that recognized by utilitarianism.
II. Utility and Constraint

The first questions we might ask are why utilitarianism cannot solve its own problems by incorporating constraints into utilitarian decision-making or else why extant constraint theories cannot provide a satisfactory model for moral decision-making by themselves without utilitarianism. In this section I will address these questions.

In all moral theories the notion of moral constraint is expressed by the theory's concept of duties and rights. Duties are moral obligations which constrain moral agents to respect certain corresponding interests called rights.

Utilitarianism recognizes only one fundamental duty - the duty of maximizing beneficence. Its recognition of this one supreme duty implies that moral constraints can always be overridden, directly or indirectly, in the interest of utility maximization. In utilitarianism all subsidiary rules - those from which constraints on the maximization of utility are to come - are essentially rules of thumb for moral action; and rights are the interests those rules protect. (4) These subsidiary rules are ultimately justified by appeal to their potential for maximal preference satisfaction in one of two ways.

An act utilitarian justifies an act under a rule only if the consequences of obeying the rule are better in a particular case, all things considered, than the consequences of breaking it. In act utilitarianism, therefore, rules have no distinct moral role and cannot act as a final constraint on utility maximization. If the direct and indirect consequences of breaking a moral rule are better than the consequences of keeping it, act utilitarianism will
always say break it.

A rule utilitarian, however, posits a different relationship between rules and the consequences of single acts. Rule utilitarianism contains a generalization provision. It justifies acting under a rule, regardless of whether the outcome in a particular case is maximally beneficent, just so long as more personal preferences would be satisfied if everyone followed the rule than if everyone followed some conflicting rule. Rule utilitarianism thus offers the possibility of acting inoptimally in a particular case so long as the rule that dictates the act is itself optimal.

The generalization provision is the source of rule utilitarianism's strength over act utilitarianism; but it is also the source of its weakness. For rule utilitarianism does not recognize the possibility that different rules may conflict in a particular case, with one having better consequences than another. And it does not recognize the possibility that some of the consequences of a choice may be attributable to an optimal rule and others simply to general utility or disutility. Rather, it requires us to seek out the optimal rule, no matter how complex that rule may be, and act on that rule. But the genuinely optimal rule will always be the utilitarian rule. So rules ultimately collapse into the act utilitarian rule; and when they collapse they take rights with them. Even a rule utilitarian must admit that no right which impedes the maximization of utility holds.

A pure constraint theory, however, cannot serve as an adequate replacement for utilitarianism. At least, no pure constraint theory developed so far comes close to offering an
adequate model of moral choice. We can see why this is so if we examine the structure of constraint theories.

As opposed to utilitarianism, which is a reductionist theory - i.e., one that proceeds by reducing each moral decision to units of utility to be summed - the extant constraint theories are all global theories. That is, these constraint theories hold that the substantive laws of morality are objective and complete and only await discovery. If the basic objective substantive duties and corresponding rights were revealed, a moral agent could assess the value of a choice simply by subsuming the facts under the applicable moral principles. It is the task of the moral philosopher to reveal the underlying objective system of moral laws.

The task of providing a complete delineation of all the objective rights and duties there are by derivation from some general principle or some small number of intuitively sound insights is Herculean. This is especially true given the fact that the presumed perfect moral theory has never been reflected in real moral theories. Nevertheless, many constraint theorists have undertaken to give a global picture of objective morality.

Some constraint theorists argue that duties and the rights which correspond to them cannot conflict: all would be well if only the exact extent of each duty and right were known, since every case would then be covered by an appropriate duty. Unhappily, these theorists never manage to give the exact extent of any duty or right; nor is it easy to know how one could do this, since every conceivable situation would have to be covered by a single, appropriate duty. (5) Indeed, it is more plausible to argue that the search for a complete map of duties and rights is hopeless.
because these are fuzzy concepts, i.e., concepts that have no
delineable boundaries. In this is true, such a theory of moral
decision-making is fundamentally misguided. We do not specify in
our own minds every situation where a certain duty or right should
apply and every situation where some other duty or right should
apply. Rather, we not only leave the scope undefined but we often
think of duties and rights as actually conflicting in practice
rather than as nicely pulling back to allow each other room to
operate. A realistic moral theory should make provision for
conflicts of duties and rights rather than attempting to define the
scope of each specific duty and right. (6)

Other constraint theorists hold that duties and rights can
conflict and that in these cases the more basic principle should
apply. These theories typically divide duties into perfect duties,
i.e., duties which are never to be violated, and imperfect duties,
i.e., duties which can be violated to protect a perfect duty.

The problem is that the resolution of moral conflicts by
recourse to perfect and imperfect duties tends to have
counter-intuitive results. For example, such a theory may typically
recognize a perfect duty to keep a promise and an imperfect duty to
protect someone from harm by one’s beneficent action. But in that
case a moral agent is only doing his duty if he keeps a promise to
tell a potential murderer where his intended victim is hiding.

Constraint theorists who recognize perfect and imperfect
rights typically attempt to escape from the dilemma in either of two
ways.

Their first recourse is to invoke the notion of intention
and the theory of the double effect: only the intended effect
(keeping the promise) should count and not the unintended one (permitting the murder). Unfortunately this argument has proved impossible to sustain, since it cannot be shown that an action that predicably will result in a certain consequence known to the agent is unintentional.

Their second recourse is to consequences: a constraint theorist may propose that a moral agent can determine the weight of two perfect duties by the consequences of violating them. The trouble with this is that in all these theories the exact formula for telling when the evaluation of a choice by its fulfilment of the provisions of a duty should give way to its evaluation from consequences is missing, and so is the argument that would reconcile these two essentially different modes of choice evaluation. (7) Yet such a formula is necessary to a well-defined model for moral decision-making.

The most sophisticated type of constraint theory is one which makes provision for the conflict of prima facie rights, determining the actual right action by the most important right in the circumstances. W. D. Ross offers a constraint theory of this type. Ross distinguishes a number of prima facie rights, with the right action to be determined by the weightiest right in the particular circumstances. Ross does not, however, have a means of determining which right should prevail, although he does claim that this judgment cannot be made on logical grounds. And, although he makes provision for taking account of the consequences of an action, since he recognizes the validity of the principle of beneficence or utility as one genuine moral principle among several, he does not show how the utilitarian method of choice evaluation can be
reconciled with the evaluation of a choice by reference to a right. Thus Ross's theory is not finally persuasive. (8)

Pure, or almost pure, constraint theories tend to be cumbersome and unconvincing, but the main objection to these theories is that none of them has ever succeeded in producing a well-defined model for moral decision-making, much less one with predictive power, and it is hard to see how one could. An elegant and persuasive pure moral theory that derives a model for moral decision-making exclusively, or primarily, from conformity to rights and duties may not be impossible; it may only be that it has not yet been developed. But the lack of success of pure constraint theories in showing what characteristics a moral choice must have does, I think, give us reason to doubt that such a moral theory can provide an adequate model for moral choice.

Rather than accepting the pure constraint account of moral choice, it is simpler and more plausible to believe that we can determine the value of a moral choice by the consequences of that choice, just as utilitarianism claims. But it is clear that we cannot judge the value of a moral choice just by summing utilities. We need a model for moral choice that not only evaluates utilities but also can accommodate a distinct role for moral constraints, or rights and duties, in moral decision-making. That is, we need a hybrid moral theory whose method of moral decision-making takes into account both the over-all consequences of actions in terms of utility, or preference satisfaction, and our duties towards individuals.

This conclusion is supported by a very strong counter-example Amartya Sen has proposed to both utilitarianism and
III. Sen’s Counter-argument to Pure Utilitarianism and Pure Constraint

Amartya Sen gives a hypothetical case in which a gang of bashers are planning to bash Ali, an East African storekeeper living in London, in some secluded spot. Donna, a friend of Ali’s, learns of the plan, but Ali is out for the day and Donna does not know where he is. She does know, however, that Ali has left a message about his plans on the desk of his business contact Charles, who is, unfortunately, also out for the day. Donna can arrive and save Ali only if she breaks into Charles’ room and reads Ali’s message. But Donna knows Charles will be embarrassed if someone sees his private papers; and she respects his right to privacy.

Sen shows that utility will be maximized if the bashers bash Ali: Ali will suffer more than any one of them will gain, but their aggregate gain will be greater than his loss. Thus utilitarianism does not give the intuitively right answer in this case, namely that Donna should break into Charles’ room, find out where Ali is, and go warn him. Rather, it gives exactly the opposite answer: that the bashers should bash Ali because the state of affairs resulting from their action will contain more utility than the alternative state of affairs. (9)

But rights theories, Sen argues, do no better. They are exclusively concerned with the preservation of rights, such as Charles’ right to privacy. On a rights theory Donna should not violate Charles’ right to privacy regardless of the consequences to Ali in terms of negative utility. (10)

Sen is surely right that any constraint theory which simply
holds that the right action is determined by reference to perfect and imperfect duties, respect for privacy being a perfect duty and the right to beneficent action an imperfect duty, will give the wrong answer in this case. A constraint theorist might object, however, that the scope of Charles’ right to privacy extends only to the boundary of Ali’s right to Donna’s beneficent action. The trouble is that the theory would have to show where to draw the line; and theories of this type are notoriously weak on that score. And if the constraint theorist argued that the rights to privacy and to beneficent action do conflict in this case and the more important right should prevail — here the right to beneficent action — we would still want to know how he determined which was the more important right. Lacking these answers, we must assume that Sen’s claim is right: a pure constraint theory cannot give the intuitively correct answer in this case.

But there is a possible answer to Sen’s case. Someone might propose a hybrid theory on which rights do conflict and the relative importance of different rights is determined from case to case by the consequences of violating each right. The consequences to Ali from the violation of his right to protection from bashing are worse than the consequences to Charles from the violation of his right to privacy. Donna should break into Charles’ room, find out where Ali is, and warn him.

Such a strategy would involve taking a consequentialist approach to rights. A philosopher who used it would have to claim that rights violations can be judged by their consequences, just as utility maximization can be determined by consequences.

The development of such a theory would not be unproblematic,
however. A theory of moral choice that judges rights violations by consequences runs the risk of collapsing into a type of negative utilitarianism — i.e., a utilitarianism that judges choices not by the maximization of utility but by the minimization of disutility. And, even if the theory did not collapse into negative utilitarianism, its reply to Sen’s case would be only a partial answer to the general problem Sen has raised, which is the insufficiency of either a pure utilitarian theory or a pure rights theory as a complete theory of moral choice. For in some cases of moral choice no one’s rights are violated. In those cases there would have to be a positive determination of rights fulfilment, rather than a negative determination of rights violations. If this positive determination were made consequentially the positive theory would closely resemble the positive utilitarian theory — if it did not actually collapse into it. And if it were not made consequentially it is not clear how it could be made.

Sen himself argues for a moral theory that acknowledges both the concept of utility maximization and the concept of rights-based constraints on moral action. He argues that what is required is a "goal rights theory," i.e., one which permits consideration of rights information in addition to utility information in the evaluation of those states of affairs which are the goals of our moral action. (Sen 1981, p. 15)

The shortcoming of Sen’s argument is that rights and utility are both simply left there, floating around together in the final state of affairs that is the goal of a choice. Sen does not define rights in such a way that they must emerge as a moral force independent of other interests or units of utility, nor does he
provide a formula for subordinating either respect for rights to utility maximization or utility maximization to respect for rights or for balancing both moral concerns.

Sen's counter-example to both pure consequentialist and pure constraint theories of moral choice is persuasive; but his own positive theory of moral choice is incomplete. The task of developing a satisfactory theory of moral choice only begins when we acknowledge that both rights and utility are essential determinants of the morality of a choice. We must still determine how they fit together in that ideal morality which common-sense morality only imperfectly shadows forth. That determination is the task of this paper.

In the next section I will give my own concept of moral constraints; and in the following section I will show how we can devise a model for moral decision-making that accommodates both utility and constraint.

IV. Moral Constraints

I will not try to argue for my view of moral constraints in this paper. I will merely present it. If it leads to a better model of moral choice than that we currently possess - one which consistently gives intuitively sound directives - the view will be justified.

I think we can best view moral constraints on the means to the end of utility maximization as contracts between the agents and the recipients of social choice. Insofar as these contracts are only moral, and not legal as well, they are non-enforceable and very often implicit. They impose obligations on the agent of moral choice to respect certain specific interests - or preferences, if
you will – of persons in specific positions in specific types of circumstances. These obligations are called "duties" and the preferences moral agents are bound to respect are called "rights."

Moral contracts are contracts – or quasi-contracts – of a peculiar kind, however.

They are not made between particular individuals. Rather, anyone who accepts morality extends its contractual provisions to all people in appropriate positions. (11) Thus moral constraints are impersonal and atemporal. That is, they do not hold between named individuals or only at specified times. Rather, they apply to whatever persons happen to find themselves in the relation of moral agent and the recipient of a moral action in given types of circumstances. (12) Moral contracts hold between types, not tokens – to use the jargon of linguistics. (13)

But, while they are impersonal and atemporal, moral contracts are not universal in every sense of the word. For different contracts have different scope – some being extended to all people, others only to the members of a given society or to some smaller group. And they certainly do not exist in some Platonic universe where they await discovery. Moral contracts are made by people. Only then do the rights for which they provide come to be extended impersonally to those who fit the descriptions of recipients of moral largesse. And only then do the duties they impose come to be regarded as binding. (14)

If moral rights and duties are contracts, we can also view them as containing provisions for acceptable violations or suspensions of the protections they grant and the obligations they prescribe. We can conceive of appropriate compensation for the
violation of a right as being agreed under the contract which grants the right. Prima facie violations of rights for which appropriate compensation was paid do not count as rights violations: they are not violations of the actual contract. (15)

But moral constraints must come from somewhere, just as utility comes from somewhere. We might, therefore, want to ask whether there are no formal or substantive restrictions on what can constitute a right or a duty.

From a purely rational standpoint we can build in protection for whichever personal preferences we wish, subject to only two restrictions.

Rights and duties must be consistent with the concept of constraint and with each other. No right can contradict the concept of a vested interest; no duty can contradict the concept of an obligation; and the set of rights can ideally contain no inconsistent rights. I say "ideally" because it seems obvious that in reality we do accord people inconsistent rights, both individually and as a society; and we then have either to reject certain rights or to restrict their scope.

I do think it is important, however, to distinguish between a conflict between rights in principle and a conflict between rights in application. Reason forbids us to have in our set of rights any two rights which conflict in principle; but it does not forbid our having in our set of rights rights which can conflict in a situation. As Sen’s case shows, the right of privacy and the right of protection from harm can conflict in application; but there is no conflict between them in principle, as there would be, for example, between a fetus’ absolute right to life and a mother’s absolute
right to an abortion.

If there are no rational restrictions on what can count as a right or duty except those posed by consistency, we may still ask whether there are any substantive restrictions on moral contracts.

I think we can best view rights and duties as having no necessary substantive content and no specific division into greater and lesser, or perfect and imperfect rights and duties. That is, rights and duties are reductive, just as utility is. The system of moral constraints consists in the moral contracts to which people adhere, just as utility consists in the preferences people have. People can contract with others to protect whatever interests they like.

Nevertheless, I believe there is a certain consistency in the interests people do choose to protect: people usually choose to protect interests that promise to promote the general good, extending rights like the right of innocent people to life. This leads me to believe that the ultimate substantive goal in assigning rights and duties is indeed the promotion and betterment of society – just the goal that utilitarians claim for morality.

The question may, therefore, justly be asked whether my theory does not collapse into rule utilitarianism, under which rules and the rights they protect are validated by their promotion of maximal utility – which I have called an unstable theory. I will address this question after I have presented my own conception of the model for moral decision-making dictated by our concern with utility and with constraint.

The important points here are the claims that the principle of utility dictates the ultimate goal, or end, of morality, the
maximization of the good; moral contracts dictate the permissable or
required means to this goal by placing moral constraints on action;
and the whole system of moral choice is a reductive system. All
that is objectively given is the form of the system itself. A
choice is moral if and only if it conforms to a fundamental law
which requires the maximization of utility subject to the
constraints posed by moral contracts, or at least, if it has the
same results as would have been achieved had the law for moral
decision-making been followed. In the next section I will show what
I think this fundamental law of moral choice actually looks like.

V. A New Model for Moral Decision-making

What is required is a new model for moral choice evaluation
incorporating the beliefs that there are moral constraints which are
worthy of our respect and that respect for these moral constraints
takes moral priority over the mere maximization of benefits. If I
am right, moral constraints take the form of contracts which
prescribe duties and rights adhering to the various positions in
which moral agents and the recipients of the results of choices can
find themselves. Therefore, what is required is a moral rule which
enjoins us to minimize the violation of moral contracts in the
application of the principle of utility. This principle must
perforce be a principle of means. When it is imposed as a condition
on the application of the principle of utility it will be a
principle governing moral means to the end of utility maximization.

The obvious candidate for the needed principle for the
minimization of violations of moral contracts is the traditional
principle of the least evil. This principle simply says, 'Do as
little evil as possible.' When the application of the principle of
utility is made contingent on the prior fulfilment of the principle of the least evil the moral directive says,

Given two alternative choices, neither of which violates any moral contracts, choose that alternative which maximizes utility. Given two alternatives, only one of which violates any moral contracts, choose that which violates no moral contract. Given two alternatives, both which violate moral contracts, choose that which violates moral contracts to the least degree. And given two alternatives which violate moral contracts equally, choose that which has greater utility.

On this conditional principle of utility maximization a right action is that which maximizes the good subject to the least uncompensated violation of moral contracts. The moral agent considers certain ones of the consequences of his alternatives as being provided for under contract. These consequences take precedence over all others. If any of the final positions of any of his alternatives contains a contract violation and there is an alternative choice which violates no contract, he will reject the alternative that violates the contract. If all his alternatives violate moral contracts, he will choose that alternative which violates them least, regardless of the utility of the choice. Only if none of his alternatives violates moral contracts or if all violate them equally will he choose that which generates the greatest amount of general utility.

It is important to remember, however, that moral contracts contain provisions for their own abrogation and for appropriately compensating infractions of their conditions. When these provisions are complied with the contract is not considered to be violated. That is, rights are not considered to be violated if they are suspended or compensated in accordance with the terms of a
pre-existing moral contract.

But in some cases moral rights are violated. We therefore require a means of judging to what degree these contracts have been immorally violated. I propose that we judge the violation of moral contracts just as we judge the achievement of the good — by consequences in terms of utility. I do not think this is any more difficult than judging preference satisfaction by consequences. We can look at choices that incorporate moral contracts and predict what the consequences of violating those contracts will be. Indeed, it seems to me that we do exactly this when judging the gravity of violating someone’s rights.

The important thing to remember is that only certain consequences will count as consequences attributable to the fulfilment or violation of the contract. These will be those consequences specifically (if often implicitly) provided for in the contract. They will consist in a subset of all the consequences of the action. It is crucial that we not get the impression that whenever moral contracts are involved in a moral choice all consequences of the choice are alike to be attributed to the fulfilment or abrogation of the contract.

For example, let us grant that I have a duty to help others and that I also have a duty not to harm myself and that these two duties conflict in a certain case. Then it may appear that the acknowledgement of both duties requires that I assess right action by the consequences tout court: if someone would die unless I help him I am obligated to help him up to the cost of my own life. On my theory this conclusion is false. My obligation to save life is a limited obligation expressed as an obligation to save only
certain lives and then only if the cost to myself is not great. Beyond this any act of saving life is an act of supererogation, not duty.

The crucial point, again, is that we remember to keep the consequences of violations of moral contracts separate in our calculations from our reckoning of the general good. Violations of moral contracts—not those waivers and abrogations of rights provided for in the contract, but genuine violations of moral constraints—should be taken into account before we calculate the costs and benefits of choices which are not related to rights.

If this view of morality is correct, the principle of utility acts as a remainder rule when appeals to moral contracts have been exhausted. But the maximization of utility within the rules is the only end of morality. Moral contracts are nothing more than moral constraints on means to that end. At the same time, however, that moral contracts act as constraints on the maximization of utility, the end of morality—utility maximization—acts as a constraint on what we are willing to accept as rights and duties prescribed by contract, for every right and duty we accord takes precedence over utility maximization. It therefore behooves us not to accord so many rights to people and prescribe so many duties that we thwart our ultimate moral aim of making life better for all.

I believe the model of moral decision-making I have given represents the true way in which at our best we do make moral choices, but at this point we are hardly home free. We must next address certain fundamental objections to the model and show how they can be met. Then we must show that the model does indeed do what I have claimed it does, namely allow us to justify the
intuitively correct answer to moral problems, which utilitarianism by itself cannot do.

VI. The New Model and Act and Rule Utilitarianism

It can readily be shown that the new model does not collapse into either act or rule utilitarianism.

In act utilitarianism all consequences of a choice are weighed together, whether those consequences follow from a rule or not. My model for moral decision-making is different. It gives priority to that preference satisfaction which is achieved under a rule or contract over general preference satisfaction. It counts the consequences of breaking a contract separately and before other consequences. It then sums the consequences which belong properly to general utility. This model of moral decision-making is very different from act utilitarian decision-making.

A rule utilitarian determines the morality of a choice by what happens if everyone follows the optimal rule. Therefore he refines the scope of his rules according to the circumstances of a case so that no more than one rule ever applies and that rule is the optimal rule. So all of the consequences of a choice figure in rule utilitarianism as consequences that follow on the use of the optimal rule. They are help define the scope of the rule.

My model is still different from the rule utilitarian model for moral decision-making. On my model rules are assumed to have fuzzy edges and to be able to conflict in an actual choice. And not all consequences of a moral choice will be attributable to respect for, or violation of, a single comprehensive rule applicable to the situation. Moreover, consequences proper to violations of moral contracts will take precedence over other consequences. This
contrasts with the model for moral decision-making which rule utilitarians themselves give, in which all consequences of a choice are assimilated to a single rule and no consequences take precedence over any other of equal degree. So the new model for moral decision-making is not a rule utilitarian model. It might more properly be called conditional utilitarianism.

I have claimed that the new model for moral decision-making rectifies the defects of utilitarianism. But it too is subject to objections. In the next section I will address the first of these objections to the new model.

VII. Objections to My Model for Moral Decision-making: Arbitrariness

It may be objected that moral contracts should not be given priority over utility maximization in every case. A philosopher who made this objection would argue that my model for moral decision-making is too rigid and arbitrary. He would argue that we can and should weigh the general good against respect for moral contracts whenever we make a moral choice. If the importance of maximizing the general good outweighed the value of keeping the contract we should break the contract.

I, of course, agree that we should weigh the effects of according a right - or entering into a moral contract - against the effect on the general good. But I think weighing the virtues of breaking a contract against those of keeping it in a particular case would be pernicious, especially once two things about my model for moral decision-making are understood. First, my model assures that certain prima facie violations of the conditions of the contract will be provided for within the contract itself. These do not count
as contract violations. So the theory is not as restrictive with respect to the protection of rights as it might seem at first. Second, my theory decrees that if some moral contract must be broken no matter what we do we should break the contract whose violation involves the lesser loss of utility. So it does provide for breaking moral contracts in well-defined circumstances. It is not a rigid rights theory.

My theory of moral choice stands on its own merits, but there are two reasons, unrelated to that theory, to resist a process of weighing and balancing moral contract violations against the general good.

First, a person who took this approach to moral choice would need a well-defined basis for weighing rights against the general good; and it is hard to imagine what this would look like. Suppose he rejected my model of moral choice and decided to weigh the virtue of keeping a contract against that of breaking it in a particular case solely on the grounds of whether the greater good would be served by breaking it than by keeping it. This is simply act utilitarianism and suffers from all its embarrassments. Therefore we must suppose that he would not want to go that far, but would nevertheless want to claim that we should break moral contracts in the interests of the greater good. He would then need some formula for determining when the greater good should take precedence over moral contracts - a formula different from either the act or the rule utilitarian formula. Would he argue that we should respect a contract so long as the consequences of respecting it were, say, two-thirds better than the consequences of promoting the general good? It would be very difficult to make such a determination. And
how would this be any less arbitrary than simply according contract
violations absolute moral priority over the maximization of the
general good?

Second, in the absence of a well-defined basis for weighing
respect for rights against the general good, the attempt to weigh
and balance respect for moral contracts against utility maximization
from case to case would reintroduce all the confusion and
uncertainty of pure constraint theories into morality, wreaking
havoc with people’s expectations of their treatment by others. For
people would have no way of knowing from case to case whether the
rights accorded them by moral contract would be respected by others
or not.

We must conclude that finding a well-defined formula for
weighing and balancing the general good against respect for moral
contracts would be involved and difficult, if not impossible, and
that the lack of such a formula would be intolerable.

I think, therefore, that the objections to weighing and
balancing respect for contracts and utility maximization from case
to case outweigh any gain in moral flexibility. By contrast, giving
respect for moral contracts absolute priority over general utility
maximization allows us to generate a well-defined model for moral
decision-making. And I will show below that this model gives
intuitively correct answers to moral questions. The objection from
arbitrariness can thus be met.

VIII. Scylla and Charybdis: Positive and Negative Utilitarianism

In Section III, I claimed that, while a consequentialist
theory of rights would seem to offer the most promising solution to
those cases of moral choice which cannot be decided to our intuitive
satisfaction on utilitarian tenets alone, such a theory would not be unproblematic, for it would risk collapsing into either positive or negative utilitarianism. My own theory of moral choice takes just such a consequentialist approach to rights. Therefore it is important to show that it does not founder on either of these two problems, which I will label Scylla and Charybdis.

Scylla is this. Suppose there is a general right to utility maximization. In that case the whole system tends to break down. If utility maximization is the supreme right the system collapses into utilitarianism. If utility maximization is only one right among others it remains to adjudicate the importance of that right. But we have indicated that in any given situation rights violations should be evaluated by their consequences. And any negative consequence should count as a violation of the general right to maximal utility, if there is such a right. So the right to maximal utility would always control our choice.

Charybdis is this. Surely, one might argue, there are rights - no matter how minor - at issue in every moral judgment. Thus every moral judgment will ultimately become one of minimizing violations of the moral contracts which award rights. If we give priority to respect for moral contracts over the maximization of utility in every case of moral choice we will never get to utility maximization as a moral end. Morality will collapse into respect for moral constraints. But if moral constraints are judged by the consequences of infringing them, as I have claimed, we will veer back towards utilitarianism - this time negative utilitarianism, where the utility of a choice is judged by how little harm the choice does rather than by how much good it does.
The solution to both objections is to deny their major premise. Utility maximization is not a right. And moral contracts are not violated in every moral decision.

Utility maximization is not a right; but this does not mean that people have no vested rights to any amount of utility. I think no society and no person outside of a dyed in the wool utilitarian has ever recognized the right of society in general to maximal utility; but we often recognize a right to certain benefits. In this society on the legislative level, for example, we recognize a right of elderly workers to social security and a right of poor families with children to government aid, among many other benefits, or entitlements. We also generally recognize a right of any person in danger to rescue if he can be aided without substantial cost to his rescuer, and so on.

What makes entitlements rights is their assignment of specific benefits to individuals in particular situations; and it is that assignment which constitutes the right. That is, the right to a certain amount of utility is a constraint we place on the distribution of utility. There neither is nor can be a right to maximal utility. Such a right does not even make sense, for utility maximization is a social goal, not a constraint on our goals, or means to an end; hence, it cannot even fit the definition of a right.

The principle of conditional utility, therefore, does not collapse into the principle of utility, for rights do not collapse into a right to maximal utility. We can avoid Scylla.

We can also avoid Charybdis: the principle of conditional utility does not collapse into the principle of negative utility,
for moral contracts are not violated in every moral decision.

In certain decisions moral contracts are not violated because contracts are not pertinent to those decisions. Some moral decisions are concerned only with maximizing the good, or preference satisfaction, and no question of indexing rights and obligations to persons in a particular position arises. For example, I may be trying to decide which of my two children most needs a new coat for the winter. I do not recognize a right of either child under some description of position, such as the older or the younger, or the girl or the boy, to a new coat. I merely look at what is called for in terms of preference satisfaction.

Very often, however, our most important moral decisions do involve moral contracts. But this does not imply that they necessarily involve the violation of those contracts. If I give my daughter a good public school education I do not violate a contract made with her entitling her to an education. But if I go beyond that and see that she has music lessons or that she goes to an academically strong preparatory school I can do what is better for her. In this case morality requires that I do what is best for her (all other things being equal), even though I do not violate a moral contract by giving her less. But the morality of this case could not be decided on grounds of contract violations, for there are none. It is decided on the basis of maximal utility.

In this section I have not tried to give a systematic account of the moral contracts, or constraints, we accept in our society, i.e., of the system (or systems) of rights and duties and provisions for waiving them or compensating for their violation which we actually follow. I have only tried to point out what I
take to be the essential relationship in moral decision-making between moral contracts in general and the general good and to illustrate this relationship with specific examples. A comprehensive account of the relationship in our society between actual moral contracts and the maximization of utility would be far too far-reaching a task for a work of this scope. Nevertheless, we can profit from a general account of utility and constraint and the relationship between them. Nor do we need a systematic account of moral contracts, or constraints, to be able to tell in many troublesome cases what our model for moral decision-making would require.

In the next section I will show that even in its very general form the model for moral decision-making I have given makes it possible to reach the intuitively correct decision in paradigmatic cases where the utilitarian model for moral decision-making fails.

IX. Application of the New Model for Moral Decision-Making: The Solution to the Defects of Utilitarianism

A principle of utility whose application is conditioned on the satisfaction of the tenets of the principle of the least evil offers a ready solution to the problems of utilitarian morality with which I began — its insensitivity to means to the end of utility maximization; its corresponding inability to condemn a choice that maximizes utility, no matter how much harm it does; and its failure to distinguish between intuitively good preference satisfaction and bad preference satisfaction. We can see this by choosing cases exemplative of each short-coming.
a. Nozick’s objection from bad means to utility maximization

With respect to the first claim, Robert Nozick has argued that utilitarianism justifies any means to the end of maximal utility. Nozick notes, as have many other philosophers, that average utilitarianism countenances the (painless and unexpected) murder of all but the most happy person if the person who remains will be happier alone than the previous population was on the average. But Nozick goes further than this. He points out that total utilitarianism is not immune on this score, since it too will approve the murder of any number of people. Nothing more is required than that the murdered people be replaced by happier people who would not otherwise have existed. Murder is, however, by anyone’s lights, the violation of a general right. On my model for moral decision-making it is morally impermissible to use murder as a means to the moral end of utility maximization. Conditional utilitarianism, or the pursuit of the utilitarian goal through moral means, cannot justify murder in the interest of the general happiness. (16)

b. Parfit’s objection from maximal utility with bad consequences

With respect to the corollary to the claim that utilitarianism allows bad means to a good end, Derek Parfit gives an example in which utilitarianism requires us, counter-intuitively, to choose a maximally beneficent alternative that includes bad consequences over a less beneficent one that produces no bad consequences.

In Parfit’s case we have a choice between two kinds of energy policy, a Risky Policy and a Safe Policy. If we choose the Risky Policy we will build a number of nuclear plants. The total
and average utility generated by this policy will be higher than if we had chosen the Safe Policy, even though under the Risky Policy there will predictably be an explosion at one of the plants in some future generation that will kill and injure thousands of people. The Safe Policy will have no bad consequences.

Both total and average utilitarianism say that, as a matter of morality, we should choose the Risky Policy because, even after the costs of the suffering it causes are subtracted, it generates more utility than the Safe Policy. But this does not seem to be the moral choice.

But on the conditional utilitarian model for moral decision-making the Safe Policy is morally preferable to the Risky Policy, for the Risky Policy violates the right we attribute to people to be secure from preventable death and injury. The Safe Policy, with a slightly lower quality of life but without the explosion, violates no moral contract. Conditional utilitarianism gives the intuitively correct judgment about this case, where utilitarianism gives the incorrect one. (17)

The Risky Policy case is more complex than it appears, however, for it involves future generations, and future generations present notoriously difficult problems in the theory of moral choice. Therefore, it is worthwhile to explore that case in more detail, for it gives support to the view that only a theory of the type I have presented can satisfy our moral intuitions in all cases.

Parfit and other philosophers have argued persuasively that who actually exists in the future depends in part on what we choose today. (18) If we choose the Risky Policy this will set in motion a train of events which after two hundred years - when the nuclear
disaster is expected to occur - an entirely different population will exist than if we choose the Safe Policy. Our choice of the Risky Policy thus helps determine who will actually live; it is an essential part of the cause of the existence of those who live and die once we have chosen it. If the Risky Policy is not chosen the Risky Policy people will never exist; if it is chosen they will exist and, by hypothesis, have lives worth living. Since they will have lives worth living, we must say the Risky Policy benefits those who exist because of it more than it harms them; and, since they would not otherwise exist, it benefits them to the full extent of the worthwhile lives they enjoy and would not otherwise have enjoyed. (19)

Parfit then argues that from a utilitarian point of view even the people who die and are maimed in the nuclear explosion because we choose the Risky Policy have no complaint. Not only is that choice maximally beneficent on the whole, but it is more beneficent than harmful to those who are harmed, since they have lives worth living which they otherwise would not have had. Thus we can predict that they would not regret the choice. Parfit asks whether it can 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?" (20)

On a utilitarian conception of the good the Risky Policy is better than the Safe Policy; and on a utilitarian conception of the right it cannot be wrong to bring the Risky Policy people into existence, even though we harm them. For the utilitarian measure of right and wrong, as we have seen, is whether or not an action
conduces to maximal beneficence. If we wish to reach the intuitively correct answer in the Risky Policy case - namely, that we should choose the Safe Policy over the Risky Policy - that answer cannot come from utilitarianism.

This conclusion does not, of course, imply that there is no other theory of morality on which it is wrong to harm others even though they will not regret our action and our action will be better for them than anything else we could have done. It only implies that if we make the claim that there is such a theory we must be prepared to substantiate it; and that is what my theory of moral decision-making is peculiarly suited to doing.

The Risky Policy case gives us reason to choose the conditional model of moral decision-making over the old utilitarian model, for with the conditional model we can claim that the rights of the future people affected by the Risky Policy are violated without appropriate compensation even though that choice is maximally beneficent, whereas no one's rights are violated on the Safe Policy.

Again, however, we run into the complication noted before, namely the fact that the people who are affected by the Risky Policy are people who live in the future. We can see why Parfit's argument that it cannot be wrong to make a choice no one will regret fails as a counter-argument against the new model for moral decision-making. It fails because regret is irrelevant to rights. But Parfit has another argument against a solution to this case through rights.

Parfit argues that we cannot make an appeal to unfulfillable rights. It is a condition of the Risky Policy people's being brought into existence that they will die in the nuclear explosion.
Things could not be otherwise for them, since if the Risky Policy is not chosen they will never exist. (Parfit 1984, p. 371ff.)

I think we can answer Parfit, however, given the nature of the theory of moral choice I have proposed. On that theory rights are rights. They are impersonal and atemporal and are accorded to anyone in a specific set of circumstances, whether those persons and those circumstances exist in the present or in the future. The Risky Policy people's rights cannot be fulfilled once they are realized, it is true. But the rights exist and can be adduced in advance of the choice that will bring the Risky Policy people into existence to determine whether, if those people are realized, they will lack a right we ascribe to everyone, namely the right to be secure from harm. Respect for rights requires us to choose a course of action that will not deny rights to people exactly as we choose courses of action that will not violate people's rights. This is possible precisely because rights, like duties, are not assigned to named individuals, or tokens, but to whoever occupies a given set of circumstances, i.e., to types of persons.

Nevertheless it might be objected that the right not to be harmed or maimed by the Risky Policy is a perfect example of the sort of right a future person would waive. Parfit might even reply that this is what he means by saying that a right whose infringement no one would regret cannot be used to justify a choice. We can reply to this objection.

Existence cannot be a counter in the bargain over rights for two reasons. First, there is no one to contract with in return for existence. Once the Risky Policy people exist they have already been denied a right we accord everyone else - the right to
protection from harm. Second, we would not withhold this right from anyone who already existed. The real ground for withholding the right to harm from the Risky Policy people is the fact that they live in the future. But we have said that position in time is not a relevant ground for denying a right: rights are atemporal and impersonal. This ground is therefore nothing more than temporal prejudice. The objection fails.

With the conditional utilitarian theory of moral decision-making - and, I believe, only with that theory or a very similar one - we can reach the intuitively right answer in the Risky Policy case. But we still need to see whether our model for model choice allows us to solve the problem of bad preference satisfaction.

C. Sen's objection from intuitively bad preference satisfaction

Amartya Sen has presented a case in which utilitarianism must call for the satisfaction of preferences whose satisfaction is intuitively bad over the satisfaction of preferences whose satisfaction is intuitively good. In his example a miserable policeman derives more satisfaction from the torture of a happy romantic dreamer than the dreamer derives dissatisfaction. This being so, utilitarianism must call the state of affairs with torture better than that without. Yet a preference for torture is intuitively bad and its satisfaction intuitively wrong. (21)

The conditional model for moral decision-making easily allows us to reach the intuitively sound conclusion in this case: the policeman's torture of the dreamer violates the right we accord the dreamer not to be tortured, but his refraining from torture violates no rights. (22) The state of affairs without torture is
morally better. We should seek to bring it about.

X. The Sufficiency of Conditional Utilitarianism

Finally, the new conditional utilitarian model for moral decision-making allows us to make intuitively right choices without giving rise to the suspicion afflicting both the pure utilitarian theory and pure contraint theories that an important part of what makes a choice moral has been overlooked or misrepresented.

In Sen’s case of Ali, the bashers, Charles, and Donna, it is true that if Donna breaks into Charles’ room to find Ali’s note about his schedule, she will violate his right to privacy. We can judge the degree of the contract violation by the predicted consequences in terms of Charles’ embarrassment. But if Donna does not break into Charles’ room no one will be able to warn Ali in time to prevent his bashing. We also recognize Ali’s right not to be bashed and Donna’s corresponding duty to warn him. Ali’s bashing will predictably be a greater violation of the moral contract that attributes to him a right to be free from bashing, as measured by the consequences to Ali, than the invasion of Charles’ privacy will be a violation of a moral contract with Charles, as measured by his embarrassment. Donna should break into Charles’ room, get Ali’s schedule and warn Ali. We need never get to the quality of the bashers’ utility, or to the question of utility maximization at all, in assessing the morality of this case.

XI. Conclusion

As Nozick’s, Parfit’s, and Sen’s cases show, the classic utilitarian model for moral decision-making is insufficient to capture all our intuitions about moral choice. But, as Sen’s case of Ali and the bashers shows, we cannot correct for the defects of
utilitarianism by switching to a theory of pure means, or rights, or constraints. The reason for the insufficiency of either a pure end-based theory of moral choice — one that ignores the means to our end — or a pure means-based theory — one that ignores the consequences of our actions — is simply that morality is both a matter of ends and a matter of means.

Utilitarianism provides the end of morality, utility maximization. Constraints, or moral contracts, provide the acceptable means to that end. And a sound theory of moral choice must capture both aspects of moral choice in a single fundamental law or model for moral decision-making.

The recognition of the complex nature of moral choice causes us to replace the unconditional principle of utility with a new conditional utilitarian model for moral decision-making. On this model it is immoral to pursue a good end, utility maximization, by bad means, i.e., by violating moral contracts. When choices contain the violation of moral contracts, we invoke the principle of the least evil. Only when the condition posed by this principle has been satisfied do we invoke the principle of utility. The maximization of general utility can thus be overridden in a moral choice out of respect for moral contracts.

If we correct utilitarianism by recognizing the concept of moral constraints and according the minimization of contract violations moral priority over the maximization of utility we can reach an intuitively satisfactory conclusion in Nozick's, Parfit's, and Sen's cases and in many others like them. And I contend not only that we can make moral decisions using the new model for moral choice evaluation but that at our best we do.
1. For support for these claims cf. Chapters II and III.

2. These shortcomings of the utilitarian model for moral decision-making are illustrated in Section IX, where a new model of moral decision-making is used to solve paradigmatic cases.


As I argued in "Preference Utilitarianism," all efforts to define utility in a way that will avoid the endorsement of "bad" preference satisfaction have failed, however; and all must fail in principle. Formal criteria, such as rational criteria, for the exclusion of the satisfaction of certain preferences from morality do not really address the objection, which is substantive. Therefore they allow intuitively bad preference satisfaction to count as utility. And substantive criteria replace the preference
criterion; they do not correct it.

Moreover, even if we could find a way of identifying and measuring bad preference satisfaction, utilitarianism still could not solve the problem it generates. For there are rational problems with excluding bad preference satisfaction from the utilitarian calculus, as Sen and Bernard Williams have pointed out. Specifically, the elimination of certain preferences from the set of a person's actual preferences causes distortions in the value of the preferences that are left; hence it causes distortions in the notion of maximal preference satisfaction. "Introduction," Sen and Williams 1982, p. 9.


6. A variant constraint theory of moral choice has been
given by Robert Nozick. For Nozick rights are simply entitlements derived from property interests transferred in a legitimate way from the original owner to his heirs and there is one basic duty, the duty to respect property rights. (Robert Nozick, *Anarchy, State, and Utopia*, New York, Basic Books, 1974.) I will confine myself to only one argument against this concept of constraint, namely the practical one that over time it becomes impossible to trace the legitimate heirs to property. Even if it were theoretically sound, Nozick's model for moral decision-making would founder on the impracticality of its implementation.

7. Charles Fried offers a rights theory on which some rights are held to be perfect, or inviolable, and others imperfect. Fried distinguishes some rights as intrinsically weightier than others, the weightier rights being those that follow from the perfect duty not to harm people. A perfect duty determines the right if both a perfect and imperfect duty, or duty of beneficence, apply. But the problem of consequences bothers Fried, and he calls for weighing and balancing consequences (without giving a clear-cut account of how to do this) where an imperfect negative duty, or duty of protection from harm, and a positive duty, or entitlement to a benefit, conflict. Also, in catastrophic cases Fried argues that an imperfect duty can override a perfect one. But he doesn't define these cases or show how he would reconcile the reductive and global approaches to moral decision-making in order to come up with an answer in any specific case. (Charles Fried, *Right and Wrong*, Cambridge, MA, Harvard University Press, 1978.) Moreover, conflicts between two perfect duties give Fried a great deal of trouble and
force him into an implausible retreat to the theory of the double effect, on which one's real duty is that which one intends. Donagan ably rebuts Fried on this score.


9. Sen's case is complicated by the fact that that maximal utility will be achieved in an intuitively bad state of affairs, namely where Ali is bashed. Hence, the right action by utilitarian lights is that which leads to this happier state of affairs. This case is actually better suited to our purposes than it may appear, however, because it reinforces the point that all personal utility satisfaction must count as good on any definition of utility developed so far, or likely to be developed, and any action or choice that leads to the maximal satisfaction of personal utility so defined must count as right on a utilitarian theory of right action.


11. In this chapter I will merely claim that moral contracts are held between persons in specific positions rather than between named individuals. I defend this view and discuss its further implications in a paper, "Is Morality Rational?", unpublished.

12. A number of prominent philosophers have taken issue
with utilitarianism over its impersonality. Cf. John Rawls, A
Theory of Justice, Cambridge MA, Belknap Press, 1971, p. 21; Thomas
Nagel, The Possibility of Altruism, Princeton, N.J., Princeton
Williams, "A Critique of Utilitarianism," Utilitarianism: For and
Against, by J. J. C. Smart and Bernard Williams, Cambridge,
Cambridge University Press, 1973, p. 93ff. I think their concern is
somewhat misplaced. It is not the impersonality of utilitarianism
that is at fault per se but its failure to recognize the existence
of contractual obligations that take precedence in moral choice over
the general good. The fact that morality can recognize obligations
to less than the whole of mankind does not entail that morality must
be personal.

13. It is a common mistake of constraint theorists who
argue with the impersonality of consequentialist moral theories to
take too narrow a view of rights and obligations. When moral
philosophers talk about rights they generally assume that the
relationship between a right and an obligation translates in
practice into a relationship between the person (or persons)
affected by a choice and the person who makes the choice, the moral
agent. But a look at our actual morality shows that this view is
false. The indexing of rights to obligations, or the scope of the
moral contract, ranges beyond a simple one to one relation of moral
agent to recipient of an action.

For example, we can imagine two cases. In the first, a
mother sees an automobile racing out of control towards her child,
and she rushes to push him out of the way; in the second a soldier
charges a bunker to save his comrades and gain a bit of ground. We do not consider these moral acts acts of supererogation, or gratuitous bounty. We consider them obligations. A mother is obligated to try to save her child, and a soldier is obligated to try to destroy enemy bunkers. Yet the mother is not the cause of the child’s danger; nor is the soldier the cause of the danger to his comrades. It is true that if the mother loses her life or the soldier loses his in saving the life of another we will consider the act heroic. In that case we may consider the mother’s act an act of supererogation — one that goes beyond duty. In the case of the soldier the act is still an act of duty. The child’s right is not perceived to extend as far as the army’s. But both are rights held against persons who are not responsible or the plight of those in danger. The fact is that there is in morality no perfect one on one correlation of rights — the child’s right to life; the other soldier’s right to protection — and obligations — the mother’s duty to protect her child; the soldier’s duty to storm the bunker. The indexing of obligations to rights in the moral contract is broader than that.

14. I have claimed that people make rights, but I have not said which people make rights. We might want to ask whether rights and duties are accorded simply by individuals or whether they are accorded by societies or by states. I think moral contracts operate on all three levels.

We can distinguish among personal, social, and legal conceptions of rights and duties. That is, in society we can distinguish between legal rights — the rights confirmed by the state
- and obligations - the laws of the state - what are called "moral rights" or "human rights" and "moral duties." Moral rights are just those rights people accord other people irrespective of the sanction of the state. Moral duties are those obligations people are perceived to have regardless of the law. On the contractual view the social definition of rights and duties may or may not differ from the legal definition. Indeed, the contractual theory of rights and duties can explain how the legally right and the morally right differ and can indicate when they will come into conflict. They will conflict when the preponderance of public opinion - or even the opinion of any faction - as to the right, or as to the existence of certain moral contracts, diverges from the law. But public opinion no more determines the right than does the law. The genuine repository of the right is in people's individual perceptions of the right.

There is a great deal of room for the elaboration of the theory of moral contracts left here. I will recognize the need for the elaboration of a complete theory of rights; but I will not attempt here to answer all the questions that naturally arise about the many issues involved.

15. It is evident that a good deal hangs on the interpretation of the term "appropriate." For example, there would be many times when my idea of appropriate compensation for violating someone's right would differ from his own idea of appropriate compensation. On my view morality would actually be achieved only when the ideas of the agent as to the extent of a right and appropriate compensation for breaking it and appropriate
circumstances for suspending it coincided with those of the persons affected by the choice. Also, rights could not morally be violated and appropriate compensation left to be determined later. Again, there are many issues left to be decided. I will not try to decide them here.

16. Nozick 1974, pp. 41-42. I have chosen Nozick's objection because of its directness and simplicity. It is exactly similar in type to all the objections to utilitarianism from means which I acknowledged in Section X, Chapter I.


19. There are many debatable premises in this argument. They are defended in Chapter I, "Moral Choice and Future Generations."


of Philosophy 76, no. 9 (September 1979) 473ff.

22. It may not, however, violate the torturer's conception of his victim's rights; this case suggests that the estimation what is right may differ from person to person, and that in some cases the social view of the right will differ from the individual view, as indicated above. Cf. fn. 14.


Donagan, Alan. "Consistency in Rationalist Moral Systems."  
*Journal of Philosophy* 81, no. 6 (June, 1984), 291-309.


38-61.


Regan, Donald. "Against Evaluator Relativity: A Response to Sen."


Sikora, R. I. "Is It Wrong to Prevent the Existence of Future Generations?," Sikora and Barry 1978, pp. 112-166.


