Economic Analysis, Moral Philosophy, and Public Policy

Daniel M. Hausman • Michael S. McPherson

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521846295

This page intentionally left blank

Economic Analysis, Moral Philosophy, and Public Policy
Second Edition

Economic Analysis, Moral Philosophy, and Public Policy shows through argument and numerous examples how understanding moral philosophy can improve economic analysis, how moral philosophy can benefit from economists' analytical tools, and how economic analysis and moral philosophy together can inform public policy.

Part I explores rationality and its connections to morality. It argues that in defending their model of rationality, mainstream economists implicitly espouse contestable moral principles. Part II concerns welfare, utilitarianism, and standard welfare economics, and Part III considers important moral notions that are left out of standard welfare economics, such as freedom, rights, equality, and justice. Part III also emphasizes the variety of moral considerations that are relevant to evaluating policies. Part IV then introduces technical work in social choice theory and game theory that is guided by ethical concepts and relevant to moral theorizing. Chapters include recommended readings, and the book includes a glossary of relevant terms.

Daniel M. Hausman is Herbert A. Simon Professor of Philosophy at the University of Wisconsin–Madison. He is the author or editor of seven books, including *Causal Asymmetries* (1998), *Essays on Philosophy and Economic Methodology* (1992), *The Separate and Inexact Science of Economics* (1992), both editions of *The Philosophy of Economics: An Anthology* (1984, 1994), and *Economic Analysis and Moral Philosophy* (Cambridge Surveys in Economic Literature, 1996, with Michael McPherson), all published by Cambridge University Press. Professor Hausman has worked on issues involving ethics and economics, foundational questions concerning the nature of rationality, economic methodology, and causation. He is currently studying the relations between health, welfare, and preferences.

Michael S. McPherson, President of the Spencer Foundation in Chicago, is past President of Macalester College (St. Paul, Minnesota) and served as Dean of Faculty and Professor of Economics at Williams College. He has been a Fellow of the Institute for Advanced Study in Princeton and Senior Fellow at the Brookings Institution. Jointly with Daniel M. Hausman he founded and edited the Cambridge University Press journal *Economics and Philosophy*. Dr. McPherson is co-author of five books. His articles have appeared in the *Journal of Economic Literature, American Economic Review, Philosophy and Public Affairs,* and *Ethics*. He has served as a trustee at the College Board and the American Council on Education, and he is a member of the National Academy of Education.

Economic Analysis, Moral Philosophy, and Public Policy

Second Edition

DANIEL M. HAUSMAN
University of Wisconsin

MICHAEL S. McPHERSON
The Spencer Foundation

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521846295

First published in print format 2006

ISBN-13 978-0-511-14492-9 eBook (Adobe Reader)
ISBN-10 0-511-14492-X eBook (Adobe Reader)

ISBN-13 978-0-521-84629-5 hardback
ISBN-10 0-521-84629-3 hardback

ISBN-13 978-0-521-60866-4
ISBN-10 0-521-60866-X

Contents

Preface

This book is a heavily revised and retitled version of *Economic Analysis and Moral Philosophy.* We added “Public Policy” to the title to emphasize the relevance of this book to policy questions. The book is a descendant of a survey essay, “Taking Ethics Seriously: Economics and Contemporary Moral Philosophy,” which we published in the July 1993 issue of the *Journal of Economic Literature.* Though now dated, that survey essay may still be of use to readers for its extensive references to relevant literature. We would like to thank John Roemer for commissioning that essay and for the detailed criticisms he offered of several drafts. Others who were of tremendous help with the first edition were Richard Arneson, Henry Bruton, Nancy Cartwright, Marc Fleurbaey, John Kautsky, Eric Kramer, Philippe Mongin, Amartya Sen, Julius Sensat, Max Steuer, Hamish Stewart, Alain Trannoy, Gordon Winston, students at Williams College and the London School of Economics, and anonymous referees. Harry Brighouse, Henry Bruton, Lester Hunt, Andrew Levine, Patrick McCartan, Jonathan Riley, David Ruben, Larry Samuelson, and Daniel Wikler read drafts of chapters of the first edition and offered valuable assistance. The research and writing of the first edition were supported by a collaborative research grant from the National Endowment for the Humanities, and Hausman also gratefully acknowledges the support of a Vilas Associate award from the University of Wisconsin, Madison.

Since philosophical reflection on ethics continues apace, as does the development of economic concepts and tools that may be of use to moral philosophers, we thought that a new edition was called for. Although we have preserved the overall structure and many of the specific analyses, distinctions, and arguments of the first edition, we have brought the discussion up to date and added examples that we hope will further illuminate the issues we discuss. We aim to reach a large audience of those interested in economics and policy analysis, and we have tried to avoid unnecessary jargon and complexities.

In preparing this revised edition, we were aided by and would like to thank Elizabeth Anderson, Mavis Biss, Richard Bradley, Harry Brighouse, Michel De Vroey, Jeffrey Friedman, Francesco Guala, David Hausman, Joshua Hausman, Bernd Irlenbusch, William Jaeger, Philippe Mongin, Colin Patrick, David Schmitz, Russ Shafer-Landau, William Thomson, Peter Vanderschraft, Joel Velasco, and David Zimmerman for detailed criticisms and suggestions for improvement. Marc Fleurbaey read the entire manuscript and offered extensive and insightful comments. Colin Patrick did a wonderful job preparing the index, and Matt and Vickie Darnell did a superb job copy-editing and typesetting this book. Over the years we have received a good deal of correspondence with suggestions for improvements in the book, and we apologize to those whose help we may have forgotten to acknowledge.

INTRODUCTION

ONE

Ethics and Economics?

Economic Analysis, Moral Philosophy, and Public Policy is concerned with economics and ethics, but it is not about how to behave ethically when doing business or doing economics. We prescribe no code of conduct and preach few sermons. Rather, in this book we try to show how understanding moral philosophy can help economists to do economics better and how economics and ethics can help policy analysts to improve their evaluations of alternative policies. We also hope to show how philosophers can do ethics better by drawing on insights and analytical tools from economics. We are writing mainly for those who are interested in economics and we aim at helping them to do economics, but we think that economics has some important things to offer ethics, too.

This focus may seem a big letdown. Surely it's more important to grapple with life-and-death moral problems! You'll get no argument from us about that. We fully agree, and we're not proposing that people stop asking moral questions. On the contrary, we hope this book will show how important morality is in economic life. But our concern is with economic theory rather than directly with economic life. *Our job will be to show clearly the role that ethics has in economics and policy analysis and to show how knowing moral philosophy helps one do economics and policy evaluation better.*

In our view, the main value of moral theories does not lie in prescribing what to do in particular situations. Moral theories are not cookbooks for good behavior. Their main purpose is to help people to understand what morality is, where it fits into their lives, and why they assign it the importance they do. Moral theories have a practical role in guiding people's reflection on the moral principles they accept and in helping people decide what to do when their moral principles conflict. Similarly, understanding ethics can help economists to think productively about the moral dimensions of policy problems, and it can bolster their confidence in recognizing and dealing with these moral issues. Knowing some ethics can help economists and policy analysts to improve their methods of policy evaluation and to understand how people's economic behavior is influenced by the moral dimensions of their lives.

Moral insights are, to be sure, more important to some parts of economics than others. Though not entirely irrelevant to any human choices, moral ideas are of little help in forecasting the price of wheat or in refining theories of exchange rate determination. Moral ideas will be more important to economists who face problems such as improving the standard of living in poor countries, increasing tax compliance, or helping citizens think through the trade-offs between environmental protection and economic growth.

1.1 What Are Moral Questions and How Can They Be Answered?

Moral questions and moral reasoning can be difficult to understand, and we have found that students often hold very skeptical or even cynical views. One hears claims such as, "It's just a matter of how you feel." "There's no rational way to resolve moral disputes. One can only fight." "Moral claims cannot be true or false." "Morality is just a matter of social convention or prejudice." These views seem to have some foundation.

- It might seem that morality is just a matter of individual feeling and that moral disagreements cannot be rationally resolved, because it is hard to understand how moral claims can be tested, confirmed, or disconfirmed.
- It might seem that moral claims cannot be true or false, correct or incorrect, because moral claims are often prescriptions and concern how things ought to be rather than how they in fact are.
- It is tempting to believe that moral claims are social conventions or relative, because members of different societies disagree about morality.

Yet these skeptical conclusions are exaggerated, and they yield implications that are hard to accept. To see why, let's be more concrete and focus on an example of a genuine moral question that might face an individual. A young woman attending college becomes pregnant and is trying to decide whether to have an abortion. This young woman might not regard this as a moral problem. She might have no doubt that abortion is morally permissible and be concerned instead about whether it would be advantageous for her to continue the pregnancy. But let us suppose that she is genuinely in doubt about whether abortion is morally permissible.

Notice first that hers is not a legal problem. She knows that abortion is in fact legal. But this doesn't tell her whether it is morally permissible. It's legal to be rude to your parents or to pretend to love somebody in order to seduce them, but that doesn't mean these actions are morally permissible. Second, notice that this young woman's question is not one that a sociologist can answer. Even if she reads that 62.37% of her fellow citizens think that an abortion is permissible in circumstances like hers, her problem has not been solved. She still needs to decide whether she ought to have the abortion or not. The third thing to recognize is that hers is a real question. It is something that she might agonize over. Whether reflecting by herself or talking over her dilemma with friends or family or counselors, she will be thinking about *reasons* why she should conclude that abortion is or is not morally permissible. Whether or not one believes that morality is subjective (in some sense of this ambiguous term) or that morality depends in some sense on feelings, there is unquestionably a huge potential role here for argument and judgment. It seems that her moral question is real, that some answers to it are better than others, and that it is possible to think rationally about which answers are better and which are worse.

There are genuine moral questions about social policy, too. For example, the question about whether abortions *should* be legal cannot be decided by ascertaining what the law *is.* The moral question of what the law concerning abortion ought to be must also be distinguished from questions about whether laws permitting or banning abortion are constitutional. Before the Thirteenth Amendment was passed, the constitution specifically permitted slavery. That made slavery constitutional, but it didn't make it just. Questions about what the constitution ought to say are moral questions. One also cannot decide whether abortions ought to be legal by means of sociological research, such as taking a poll. A poll can determine what most people believe, but it won't say whether they're right. Those who believe that abortions ought not to be legal cannot be refuted by results of polls showing that most people believe that they should remain legal. One addresses moral questions instead by making arguments.

Once we recognize these truisms – that moral questions have better and worse answers, and that arguments can sometimes help people find out which answers are better – we can see that the cynical or relativistic conclusions concerning morality are exaggerated and unjustified.

- It is not true that there's no method of resolving moral disagreements and that consequently all one ever gets in morality is disagreement. There is a method: One can make arguments; that is, one can look for premises that others agree on and then use logic to try reaching agreement on the issues in dispute. When people stand to benefit from doing evil, they may be deaf to rational argument. It took a civil war – in addition to the arguments of abolitionists – in order to eliminate slavery in the United States, but without those arguments (to which there were, in fact, no good responses) there wouldn't have been a movement opposing slavery.

- The fact that moral judgments are *prescriptive* – that they say how things ought to be rather than how they are – may imply that moral judgments cannot be literally true or false, but it does not follow that one cannot sensibly consider whether some moral judgments are mistaken. Although there are subtle philosophical questions one might ask about the sense in which prescriptions can be correct or mistaken, clearly some prescriptions are better than others, and there is room for rational argument concerning which are better and which are worse. Moreover, even if moral judgments are not descriptive assertions, the reasons for those judgments often include empirical claims that can be criticized and investigated.
- Though moral questions are not always easy to answer and though difficult questions give rise to persistent disagreement, there is also a lot of agreement in ethics. Few people approve of torture for any purpose, and even fewer approve of it for entertainment.

The claim that morality is "relative" can be confusing, because in one sense morality clearly is relative: what's right depends on (is relative to) what the facts are. Whether it is permissible to knock over a frail old man depends on whether one knocks him over to see whether his bones are brittle or whether one knocks him over to prevent him from being run down by a truck. But to recognize that one does not have a well-defined moral question until one has specified all the facts is perfectly consistent with the idea that well-defined moral questions have better and worse answers.

What people mean by claiming that morality is relative is often something altogether different: that whatever a person (or a society) believes is right is automatically right ("for that person or society"). But when the woman in our previous example is trying to decide whether it is morally permissible to have an abortion, she is not trying to find out what her beliefs already are; she is trying to find out which answer to her question is correct. Similarly, when thinking about whether abortion should or shouldn't be legal, people are not trying to find out what they (or others in their society) already believe but instead what the law concerning abortion should be.

If whatever people believed about ethics were automatically right, then there could be no moral disagreement. To disagree with someone about a moral question commits you to believing that people's ethical beliefs can be incorrect. Similarly, if a social consensus guaranteed its own correctness, then defenders of unpopular views would automatically be mistaken. One wouldn't need to argue with defenders of minority views, since they couldn't possibly be right. But iconoclasts cannot be refuted with polls, and social consensus is not proof of correctness.

We recognize how tempting it is to think that there is no fact of the matter about morality and that, even if there were, people could not know it. Morality seems in large part a human construction, so it is easy to jump to the conclusion that it is mere social convention or, more radically, that individuals determine what is right or wrong by what they believe or feel. But these temptations lead either to moral nihilism – the complete rejection of morality – or to views that cannot be sustained. If you think that anything is right or wrong, good or bad, morally praiseworthy or blameworthy, then you are not a moral nihilist. And if you take any moral claims seriously, wonder if they are correct or incorrect, and sometimes disagree or argue with others, then you cannot believe that all moral views are on a par and that there can never be any reason to accept some and reject others.

Sometimes people feel that it is intolerant or dogmatic to believe that their moral convictions are correct. In some cases they are right, because some systems of morality are dogmatic in maintaining that there is nothing to be learned or debated concerning the one true moral code. But whether tolerance is a virtue and what views and actions should be tolerated are questions *within* morality. Some moralities are tolerant while others are intolerant, just as some people – whether the subject matter be morality, sports, or deodorants – are dogmatic and others are ready to listen and learn. Short of giving up morality altogether, including all concerns about tolerance, there is no alternative to taking one's moral beliefs seriously. People who are genuinely tolerant are not moral skeptics: They believe that tolerance is (nonrelativistically) good and that those who are intolerant are wrong to be intolerant. Tolerance is tied to an appreciation of the richness of different cultures and different life experiences, to a respect for others, and to a willingness to take their perspectives and arguments seriously. It is not a form of skepticism. Furthermore, to believe that there are better and worse answers to moral questions does not imply any unwillingness to listen to the arguments of others or an inability to see one's own limitations. A serious moral commitment to tolerance is a better remedy for dogmatism than is an impossible skepticism.

There is nothing suspect or intolerant about believing that some answers to moral questions are better than others and that rational argument can help one to judge which answers are better. These beliefs are implicit in individual moral judgments and in policy making, and it is hard to deny them without denying that there is any such thing as morality.

1.2 How Is Moral Philosophy Relevant to Economics?

The idea that studying ethics could help people to do economics or policy analysis may seem far-fetched. Why not consult tarot cards instead? Many people – indeed, probably some of you reading these very words – doubt that moral philosophy can help one do *anything* better (except perhaps to spread confusion and cure insomnia). If one is seeking clarity, why look in a swamp?

Our hope is that readers of this book will not come away with the impression that argument in moral philosophy is obscure, unworldly, or boring. It is, to be sure, intellectually demanding, abstract, and often intricate, and we'll not dispel all puzzlement. Like economics itself, moral theory is loaded with controversies and unresolved issues. We want most of all to show that moral reasoning can help people gain a surer grip on serious problems about how to make their lives and our society better.

But even if moral philosophy clarifies morality, why should it clarify economics? Many economists would draw a sharp distinction between evaluative questions and the "positive" science of economics, which is concerned with facts, not values. In the 1930s, Lionel Robbins expressed this view as follows: "it does not seem logically possible to associate the two studies [ethics and economics] in any form but mere juxtaposition. Economics deals with ascertainable facts; ethics with valuations and obligations" (1935, pp. 148–9). Robbins is drawing on a commonsense distinction (which is maddeningly difficult to make precise) between factual claims and evaluative claims. Intuitively, there is a huge difference between describing how many tons of steel the United States imported in 1999 and saying whether it is a good or a bad thing.

Although there is, we believe, no way to draw the distinction between facts and values precisely, it is worth describing how philosophers and economists have distinguished them. Figure 1.2.1 summarizes the contrasts. We should stress that these contrasts are highly controversial and that (as we shall argue in the Appendix) the distinctions are exaggerated and in some cases mistaken.

Factual Claims	Evaluative Claims
Disagreements can be resolved by evidence	No good way to resolve disagreements
Relatively little disagreement	Relatively little agreement
Descriptive: say how things *are*	Prescriptive: say how things *ought to be*
True or false	Not true or false
Objective	Subjective
Independent of evaluative claims	Dependent on factual claims
Help to achieve goals	Help to determine goals

Figure 1.2.1. Exaggerated Contrasts between Facts and Values

In its gross exaggeration of the contrasts, Figure 1.2.1 makes it easy to see why economists have thought it was so important to argue that economics is and should be "value neutral": that it makes (and should make) no evaluative claims. Economists believed that only factual claims can be studied by the methods of science. In this view, economists trespass beyond the boundaries of science when they take stands concerning evaluative matters.

This view of the separation between scientific claims, which are the subject matter of economics, and evaluative claims, which cannot be the subject matter of any science, might be expressed as follows: "Economics is science or engineering. It shows how to arrive at certain goals but, unlike ethics, it does not prescribe what goals one should have. Economics provides technical knowledge that has no more to do with ethics than does geometry or physics. No matter how sensible and well-conceived ethical theories may be, they have nothing to do with economics and cannot possibly help one understand economies."

This entire book is a response to the view that ethics and economics have (and should have) nothing to do with each other. The best case for relating economics and ethics is to show that linking the two subjects is productive in the practice of those disciplines. We begin to make this case in Chapters 2 and 3 by showing through examples how unproductive it is to try to "cleanse" economics of the evaluative content of familiar economic ideas like efficiency, welfare, and freedom. Instead of beginning this book with abstract philosophical considerations concerning the relations between economics and ethics, we have postponed that discussion to an appendix in which we directly challenge the claim that economics and ethics should be sharply separated. There we also criticize explicitly the "engineering" vision that portrays economics as entirely value neutral. However, in the main text of this book we try to be constructive, showing concretely how evaluative and factual matters are entangled in economics and policy analysis.

1.3 Organization

Chapters 2 and 3 present four examples that illustrate ways in which moral questions arise in economics, and our concluding Chapters 15 and 16 return to these examples and apply the concepts, distinctions, and principles developed in the intervening chapters. Those eleven chapters are divided into four parts.

Part I focuses on *rationality.* Like morality, rationality is normative. One ought to be moral and one ought to be rational. One is wicked if not moral and foolish if not rational. Rationality, unlike morality, plays a conspicuous role in contemporary economic theory. Economists usually deny that economic theory presupposes any ethics, but they freely admit that it presupposes a great deal about rationality. However, economists cannot have it both ways. Endorsing their theory of rationality, we will argue, commits them to controversial moral principles. In defending their model of rationality, economists wind up espousing fragments of a moral theory.

It also turns out, we shall argue, that exploring the connections between morality and rationality leads to criticisms of economics, because the moral principles implicit in standard views of rationality are implausible. When these principles are stated explicitly, few people would endorse them. The standard views of rationality held by economists also make it hard to understand how social norms and morality can be rational. Taking ethics seriously in this case leads (or so we shall argue) to serious theoretical criticisms of basic principles of economics.

Parts II and III then zero in on concepts and theories of economic *evaluation.* Which economic policies and institutions are best? How should they be judged? Part II focuses on the standard theory of welfare and on methods of evaluating economic outcomes and institutions in terms of their consequences for welfare, such as welfare economics and utilitarianism. We shall criticize the preference satisfaction view of welfare that economists defend, and we shall argue that welfare is not the only thing that matters ethically. But we do not doubt that welfare is of great moral importance, and a major aim of Part II is to clarify its role.

Part III is mainly concerned with four other notions: freedom, rights, equality, and justice. These notions are important in the evaluation of economic policies and institutions, and moral theories have been built around them. When one thinks about other things than welfare, new vistas appear. Not only are there new questions and new aspects of economic arrangements to consider, but there are also new methods of thinking about morality. We shall in particular say a good deal about *contractualism,* which provides a way of making sense of morality in which the consequences of policies for individual welfare are not necessarily decisive. Whereas Part II uses the concepts of standard welfare economics, Part III presents alternatives to the questions that welfare economists ask and to the terms in which they answer them.

Part IV provides an introduction to some technical work in economics that is directly guided by ethical concepts and is immediately relevant to moral theorizing. The payoffs from knowing something about both economics and philosophy do not go only to economists. Philosophers have lessons to learn from attempts at formalizing moral concepts and exploring their consequences.

The concluding chapters consider the benefits of harnessing the combined powers of economics and ethics in addressing important policy issues. We revisit the four cases discussed in Chapters 2 and 3 and try to show, through these and other examples, how the ideas and tools explained in this book might help identify good policies and principles for citizens and governments to adopt.

At the end of each chapter (except this one and the final two), we provide a brief discussion of relevant literature. A glossary of relevant terms is assembled at the end of the book.

When you are finished with this ramble through the lush forests of moral philosophy and the brushland where it meets economics, we hope you'll see that economics remains partly a moral science. It can't be done without moral presuppositions, and it's hard to do it well without addressing moral issues intelligently. Similarly, moral philosophy can't be done without beliefs about human interactions, and it's hard to do it well without knowledge of the kind that economists seek. Like those who would completely disavow the culture of their parents, economists sometimes try to deny their philosophical lineage. Although they can reform and improve their philosophical inheritance, they can't escape it, and attempting the escape renders their theories hollow. Neither can the philosophical parents of today's social disciplines successfully repudiate their offspring. Moral philosophy and economics have much to contribute to each other.

TWO

Ethics in Welfare Economics: Two Examples

This book will be filled with arguments, but examples help bring them down to earth. One good example may do more to clarify how ethics matters to economics than would a hundred pages of argument. Furthermore, ethics is not just logic. Emotion has its part to play, too, and examples help to engage the emotions. In this chapter and the next, our concern is not to argue that ethics matters in economics but instead to exhibit – through examples – how important ethics is.

In this chapter we will focus on two examples, which will enable us to identify all the main moral assumptions that characterize mainstream *normative economics.* The first example caused an uproar.

2.1 A Shocking Memorandum

In December of 1991, Lawrence Summers (now president of Harvard University, but then the World Bank's chief economist) sent the following memorandum† to some colleagues.

> Just between you and me, shouldn't the World Bank be encouraging *more* migration of the dirty industries to the LDCs [less developed countries]? I can think of three reasons:
>
> (1) The measurement of the costs of health-impairing pollution depends on the foregone earnings from increased morbidity and mortality. From this point of view a given amount of health-impairing pollution should be done in the country with the lowest cost, which will be the country with the lowest wages. I think the economic logic behind dumping a load of toxic waste in the lowest-wage country is impeccable and we should face up to that.
>
> (2) The costs of pollution are likely to be non-linear as the initial increments of pollution probably have very low cost. I've always thought that under-populated

† Quoted in *The Economist,* 8 February 1992, p. 66.

countries in Africa are vastly *under* polluted; their air quality is probably vastly inefficiently low [sic] compared to Los Angeles or Mexico City. Only the lamentable facts that so much pollution is generated by non-tradable industries (transport, electrical generation) and that the unit transport costs of solid waste are so high prevent world-welfare-enhancing trade in air pollution and waste.

(3) The demand for a clean environment for aesthetic and health reasons is likely to have very high income-elasticity. The concern over an agent that causes a one-in-a-million change in the odds of prostate cancer is obviously going to be much higher in a country where people survive to get prostate cancer than in a country where under-5 mortality is 200 per thousand. Also, much of the concern over industrial atmospheric discharge is about visibility-impairing particulates. These discharges may have very little direct health impact. Clearly trade in goods that embody aesthetic pollution concerns could be welfare-enhancing. While production is mobile the consumption of pretty air is a non-tradable.

The problem with the arguments against all of these proposals for more pollution in LDCs (intrinsic rights to certain goods, moral reasons, social concerns, lack of adequate markets, etc.) could be turned around and used more or less effectively against every Bank proposal for liberalisation.

The memorandum is worth reading carefully, and we would urge you to read it a second time before proceeding further. It seems that Summers wrote this memorandum as a provocative exploration of the implications of "economic logic" rather than as a serious proposal for a World Bank program to export pollution to the LDCs. What makes this memorandum particularly worth studying is that its economic logic is exemplary yet its conclusions are alarming to many people. We've seized on this example because it is easier to recognize controversial ethical content in economic theory when it is blatant and possibly repugnant than when it is subtle and intuitively plausible.

2.2 Eight Distinctive Features of Welfare Economics

Summers's memorandum does not purport to be just a study of facts, which bear on policy only in the way that engineering bears on policy. On the contrary, this memorandum makes claims about what the World Bank "should" be doing, and it describes some facts as "lamentable." Summers is clearly making evaluative claims, and his work would be excluded from economics by those who insist that economics must be free of any value judgments. Yet this memorandum obviously seems to be concerned with economics. One way to recognize this – while still insisting on the importance of distinguishing between factual and evaluative claims – is to maintain that there are two kinds of economics: "positive economics," which deals only with matters of fact; and "normative economics," which is concerned with the evaluation of economic states of affairs, processes, and institutions. Summers's memorandum is clearly an instance of normative economics.

We would like to draw your attention to eight features of this memorandum which are, we maintain, typical of mainstream normative economics or *welfare economics.* Each of these features represents a choice: Summers's way of thinking about economic states of affairs and policies is just one of many possible ways. Once one recognizes how this way of thinking about outcomes and policies is distinguished from other ways, one understands a great deal about normative economics.

1. We've already mentioned the first of these eight features: Summers is concerned with *evaluating* economic states of affairs and with recommending how to improve them. Note that his focus is on economic outcomes rather than processes.
2. Summers assumes that there is a single framework for economic evaluation, which he takes for granted. He never states it explicitly, and he never argues for it. Though he wouldn't put things this way, Summers is relying on an ethical foundation that he believes his readers share.
3. Summers's memorandum takes the form of an *argument* in which premises concerning costs and demand are supposed to establish conclusions concerning what the World Bank should do and what facts are lamentable. That argument seems to draw on rather intricate economic reasoning. The ethical conclusions are based on ethical reasoning, not on appeals to emotion or prejudice.
4. The memorandum considers how policies and states of affairs bear on *individuals.* No questions are asked about the significance of their effects on other things such as the environment or local cultures, except insofar as those in turn affect the welfare of individuals. Summers assumes that what matters are the consequences for individual people.
5. The memorandum evaluates economic states of affairs in terms of their consequences for individual *welfare* rather than with respect to any other feature of individuals. Summers is concerned about whether transfers of pollution would be "welfare-enhancing." Owing to the prevalence of this feature, mainstream normative economics is typically called "welfare economics."
6. In measuring welfare, the memorandum implicitly accepts the way that markets evaluate things when (competitive) markets exist.
7. Although the memorandum focuses exclusively on welfare, it does not attempt to add up welfare gains and losses or to compare the welfare of different people. Summers does not claim that trade in pollution

would maximize total or average welfare; he never mentions total or average welfare.

8. In addition to focusing exclusively on the welfare implications of shifting pollution, the memorandum suggests that there is a qualitative difference between the "impeccable" "economic *logic* of dumping a load of toxic waste in the lowest-wage country" [our emphasis] and miscellaneous and unspecified ethical objections in terms of "intrinsic rights to certain goods, moral reasons, social concern, lack of adequate markets, etc." Summers implies that the welfare arguments are rigorous and worth taking seriously while the miscellaneous objections can be disregarded. (Otherwise, Summers's claim that these objections "could be turned around and used more or less effectively against every Bank proposal for liberalisation" should lead him to criticize the World Bank's proposals for liberalisation rather than to make an additional one!) Welfare economists rarely deny explicitly that other moral considerations are relevant to evaluating policies and outcomes, and indeed Summers makes no such argument. However, welfare economists are, like Summers, often suspicious or even contemptuous of other ethical concerns.

Some of these eight features of welfare economics are widely shared in the thought and culture of modern liberal democracies, while others are more distinctive to mainstream economics. Thus the presumption that social policies are subject to rational evaluation through argument, as in items 1 and 3, is widespread (though hardly universal), while the tendency to focus on outcome evaluation and downplay process is more peculiar to economics. Similarly, "individualism" – in the particular sense of item 4 that the only sources of ultimate value are human individuals – is a familiar feature of liberal social theory, while the idea that it is only the *welfare* of those individuals that ultimately counts (item 5) is more specific to normative economics.

These features of welfare economics, even those that are widely shared with liberal social theory more generally, are not inevitable. Each involves a choice, and each feature could be questioned or changed. These choices are both methodological and ethical. Although welfare is obviously very important, so is freedom and so is justice, and normative economics might focus on them in addition to or instead of welfare. *There are alternatives,* and to choose among them requires ethical reflection.

One of our tasks, which will occupy much of Parts I and II of this book, is to lay bare the philosophical commitments underlying welfare economics and then examine to what extent these commitments constrain the character of normative economics. Another task, which will occupy Part III, is to consider alternative approaches. Rather than passing judgment, we aim to reveal the philosophical complexities implicit in policy evaluation as well as the range of choices that economists might reasonably make.

2.3 The Economic Benefits of Exporting Pollution to LDCs

The last three of the eight features of the memorandum we highlighted are all concerned with the economic benefits of transferring pollution to less developed countries. In Part II we shall have a good deal to say about the notion of an economic benefit or cost, but a few words here will help clarify Summers's view. Air and water pollution lessen the quality of life in many ways, yet most kinds of pollution have no market prices. One cannot go to the hardware store and purchase for $19.95 a 20% decrease in the toxic chemicals in the air one breathes.

Three factors explain why there are few markets in pollution limitation. First, it is not possible to locate all the sources of air and water pollution. (If your neighbors decide to dump leftover weed killer down their drain, who will ever know?) Second, even if it were possible, it would be prohibitively expensive to strike a deal with all the polluters in order to improve your air or water. Third, any deal you strike with polluters will affect your neighbor and vice versa: while walking to the corner, you've got to breathe the same air your neighbor breathes. Any effective deal will require cooperation among your neighbors. Thus some collective action is often needed in controlling pollution.

One way economists can help with the problems of controlling pollution is by *imputing* costs to it. The hope is to figure out what pollution costs would be if there actually were markets where pollution could be bought and sold. Economists may attempt to impute pollution costs by examining housing prices in communities that are much the same – apart from their air quality. They can draw inferences from how much people pay for air filters, water filters, or bottled water. They can collaborate with biologists in determining and assessing the costs of damage to health caused by pollutants. In such ways, economists may be able to estimate how much people in developed countries would be willing to pay to lessen pollution in their environment and how much people in LDCs would have to be compensated in order to be willing to accept more pollution.

People might be ignorant of the harms caused by pollution. People in an LDC might be willing to accept toxic wastes for very little compensation if they are unaware of the contents of the wastes, the harms they might do, or the prospects of the wastes escaping and poisoning ground water. Any willingness to accept more pollution that is based on such ignorance would not truly reflect what serves people's interests. Summers avoids this difficulty, because he argues that the willingness of those in LDCs to accept pollution for less compensation than those in developed countries rests on differences in the consequences of pollution in LDCs and developed countries.

Economists employ complicated techniques to determine economic costs when things are not literally bought or sold, and it will simplify the discussion here if we engage in some make-believe. Suppose pollution could be bought and sold on the market. Since pollution is a "bad," consumers would want to buy its *absence.* Let us think of reductions of pollution that consumers could buy as units of a single hypothetical commodity, "environmental quality." We can then think about what the price of units of environmental quality would be in this make-believe world in which individuals could separately purchase and consume environmental quality while firms could separately produce and market units of it.

One can then take Summers to be arguing that – were it possible for individuals easily to buy, sell, and transport "environmental quality" – there would be active trading between the developed and less developed nations of the world, and pollution would be pouring out of the developed nations and into the less developed nations. Wealthy people would be buyers of environmental quality and poor people would be sellers for the same reasons that wealthy people are buyers of housecleaning services and poor people are sellers of them. But units of environmental quality cannot ("lamentably"?) be individually appropriated, bought, and sold, and it is hard to transport pollution between nations. Consequently, the World Bank can enhance world welfare by simulating what a hypothetical market would accomplish by facilitating transfers of polluting industries to LDCs.

2.4 Summers's Argument and a Further Feature of Welfare Economics

The three points in Summers's essay show that the amount that individuals in developed countries would be willing to pay in order to lessen their pollution is more than the amount that individuals in LDCs would demand in order to accept more pollution; also, there would be more pollution in LDCs if pollution were easily exchangeable. But why should one conclude that it is "lamentable" that pollution is not easily exchangeable? Why should one conclude: "Clearly trade in goods that embody aesthetic pollution concerns could be *welfare-enhancing*?" [our emphasis]. What makes it an improvement to approximate more closely how things would be if pollution could be easily traded? How is one supposed to reach the conclusion that "the World Bank [should] be encouraging *more* migration of the dirty industries to the LDCs"? How does one get from claims about hypothetical markets or willingness to pay to claims about welfare and from claims about welfare to claims about what the World Bank ought to do? We need to clarify the logic of Summers's argument.

We suggest that Summers's argument can be spelled out as follows.

1. Rational agents in LDCs would accept pollution from developed countries for less compensation than rational agents in developed countries would be willing to pay to get rid of the pollution. In other words, for some compensation *C* – which lies between the least that agents in LDCs will accept and the most that agents in rich countries will offer – all rational individuals, whether in developed countries or in LDCs, would prefer to transfer pollution from a developed country to an LDC (premise).
2. Whatever individuals prefer makes them better-off or increases their welfare (premise).
3. Shifting pollution to LDCs from developed countries and paying some compensation *C* makes everyone better-off (from 1 and 2).
4. One should adopt policies that make people better-off (premise).
5. One should adopt policies that shift pollution to LDCs and pay compensation *C* (from 3 and 4).

If one assumes that the jobs and revenues provided by dirty industries are "reasonable compensation," then this reconstruction captures what Summers intends when he claims that "the World Bank [should] be encouraging *more* migration of the dirty industries to the LDCs." In fact, as Summers recognizes, many transfers of pollution are not feasible, and individual units of pollution typically cannot be bought or sold. Some economists might argue that it would be a good thing to shift pollution to LDCs regardless of whether there is any compensation – on the grounds that there is a "net benefit" in such shifts. But it takes more controversial moral premises to establish this stronger conclusion, and we will not comment on arguments that rely on the notion of a net benefit until Chapter 9.

Statements 2 and 4 appear to be moral premises, and in later chapters we shall examine them at length. Premise 2 is particularly important in forging the link between market evaluation and welfare by identifying welfare with the satisfaction of preferences. This feature lies at the very core of mainstream normative economics. It is through this identification (plus the connection that positive economics establishes between preferences and

1. What should economists appraise?
   a. *Outcomes*
   b. Processes
2. What method(s) of appraisal should economists use?
   a. *Single method of appraisal*
   b. Multiple ethical perspectives, depending on problem
3. What matters about outcomes?
   a. *Consequences for individuals*
   b. Consequences for groups, or the environment
4. Which features of outcomes for individuals matter?
   a. *Welfare*
   b. Freedom
   c. Rights
   d. Justice
5. What is welfare?
   a. *The satisfaction of preferences*
   b. Some mental state, such as happiness
   c. "Objective" goods – e.g., achievements, personal relations, health
6. How does welfare (as preference satisfaction) bear on the evaluation of outcomes?
   a. *Market evaluation and the Pareto concepts*
   b. Add up preference satisfaction
7. What role do other ethical notions play?
   a. *Independent; important, but not a concern of economics*
   b. Their importance is derivable from their consequences for welfare
   c. Must be integrated into the economic appraisal
   d. Of no importance

Note: Responses in *italics* are those of orthodox welfare economists.

Figure 2.4.1. The Moral Framework of Normative Economics

market prices) that Summers is able to link premises about costs and demands to conclusions about what outcomes are good or bad and what policies will enhance welfare.

We especially want to emphasize that the tone of the memorandum (which is addressed to fellow economists, not to the public at large) is misleading. It suggests that the three numbered paragraphs make a "scientific" case, with the last paragraph dismissing wishy-washy moral objections. But the moral content does not wait for the last paragraph to make its appearance. The three numbered paragraphs are part of a *moral* argument, and Summers's economics is saturated with ethics.

Figure 2.4.1 lists some of the alternatives that normative economists face and the specific choices orthodox welfare economists have made.

2.5 Is Summers Right? Should the World Bank Encourage Migration of Dirty Industries to LDCs?

The uproar caused by this memo suggests that many people are not willing to accept its conclusion. But this may be a thoughtless first reaction. Why shouldn't the World Bank encourage migration of dirty industries? Since the argument is logically valid, those who reject the conclusion must reject at least one of its premises. Though we cannot speak for all those who find this conclusion unacceptable, here are five possible objections (we will consider others in Chapter 15).

1. Premises 1 and 3 tacitly assume that the consequences of pollution are local and that the amount of pollution is independent of where the pollution takes place. Yet both of these assumptions are questionable. Encouraging dirty industries to migrate to LDCs might easily lead to more total pollution. Developed countries have both the incentive and the administrative capacity to enforce pollution controls, whereas LDCs (for the reasons Summers gives) have less incentive to limit pollution, and they may have a harder time enforcing pollution regulations. Since many of the effects of pollution are global, the consequences of encouraging migration of dirty industries to LDCs may be very bad indeed.

Although the first objection – that migration of dirty industries will increase worldwide pollution – is a serious one, it need not challenge Summers's moral framework, and we shall have little to say about it. To the extent that the effects of pollution are not localized, Summers's arguments concerning the net benefits of shifting pollution do not go through. If the harms of pollution are not mainly local, then one cannot conclude that the migration of dirty industries will be beneficial – no matter how attractive the arrangement may be to particular traders.

2. Even though people in both developed economies and LDCs would be happy to shift pollution to LDCs for some reasonable compensation, the exchange is unfair. Consider an agreement between a billionaire and a beggar whereby the beggar agrees to work 16-hour days in exchange for gruel and a straw mattress. Both parties may be rational and well-informed and may enter the agreement "voluntarily." (As we will see in considering involuntary unemployment in the next chapter, to maintain that any exchange not directly involving physical force is voluntary is to make a controversial moral claim.) But the billionaire nevertheless is exploiting the beggar. No matter that the beggar may agree without reservations, the arrangement seems to most people unjust. This analogy may actually understate the injustice of dumping pollutants in LDCs, because some of the poverty

in LDCs has been caused by the developed countries. For example, agricultural subsidies in rich nations have devastated agriculture in poor countries. Exploitation seems more egregious if the wealthy party caused the other's poverty.

This second objection would not surprise Summers. He alludes to such objections in his last paragraph, but he offers no response apart from the rather alarming observation that such objections "could be turned around and used more or less effectively against every Bank proposal for liberalisation." This unfairness objection shows that premise 4 – that policy should make people better-off – needs qualification. It is plausible to maintain that, other things being equal, it is a good thing to make people better-off. But other things are not always equal. It may not be a good thing to make people better-off if, for example, doing so involves severe unfairness.

In our view, to point out that shifting pollution to LDCs might be exploitative is not enough to show that it is wrong. Most people are prepared to put up with additional injustice in exchange for a sufficiently large increase in welfare. Before deciding what is best, one needs to know how large the welfare benefits will be and how much unfairness will result. But normative economics only talks about welfare; it has nothing to say about how to consider trade-offs between justice and welfare. One misleading feature of this memo is its suggestion that, apart from some moral bellyaching, the policy implications are simple.

3. Summers's analysis compares only one possible alternative to the status quo: shifting pollution to LDCs. But there may be other policies that would be better still. Notice in particular that Summers's case depends on the huge income disparities between rich and poor countries; without those disparities, why would people in (say) Africa pay less to avoid pollution than people in the United States? Should this status quo income disparity be simply taken as given? Transferring wealth from rich to poor countries might well be much more welfare enhancing than transferring pollution.

4. Satisfying preferences does not automatically increase welfare. People may prefer things that wind up being bad for them. Voluntary exchange need not always be mutually advantageous.

The basic point in this objection is hard to deny, but Summers can argue that – given the actual consequences of increased pollution in LDCs – rational individuals in LDCs should be willing to accept pollution for less money than individuals in developed nations should be willing to pay. The reasons are that wages are lower in LDCs, low levels of pollution do little harm there, and poverty in LDCs makes clean air and water relatively less important. The greater willingness to accept pollution does not derive from misunderstanding or miscalculating the effects of pollutants. The standard reasons why one might question whether satisfying people's preferences would make them better-off do not apply.

5. What about premise 1, that all rational and well-informed agents would prefer to make the exchange? This premise is itself the conclusion of an argument from the fact that the (economic) costs of pollution are lower in LDCs than in developed countries. But do the economic costs and benefits capture what is morally relevant? Must rational and well-informed individuals accept the market's evaluation of pollution's consequences? Isn't premise 1 a controversial moral premise, too?

The moral relevance of market prices is questionable, even when they do not reflect any ignorance or irrationality. Individuals need not accept the market's evaluation of the consequences of pollution. If there were markets on which all varieties of environmental quality could be bought and sold, then there would be massive transfers of pollution to LDCs. But those transfers would not necessarily increase human well-being. Let's see why.

People in LDCs are willing to pay relatively less to avoid pollution, because pollution causes relatively less harm in LDCs. It causes relatively less harm (as indexed by willingness to pay) because wages and productivity are lower, because people are more likely to die of other things before they can be harmed by pollution, and because there are other pressing needs upon which individuals will spend their money first. But is willingness to pay the right way to measure economic benefits? Are economic costs and benefits a good guide to what is harmful and beneficial? Given the current unequal distribution of wealth, people in rich countries will pay much more to prevent crippling injuries than people in poor countries will. But the moral significance of crippling injuries should not depend on whether the victim lives in a wealthy country. A middle-income person in a rich country might pay $50 to acquire a third DVD player for his household, while a poor person in an LDC might decline (or be unable) to pay the local equivalent of $50 to fix his broken leg. Willingness to pay seems a very dubious guide to what *ought* to be done. To regard costs and prices, as determined by actual markets, as markers of moral significance is itself to make a moral presumption – and a dubious one at that. For a telling example, see Figure 2.5.1.

In fact, economists do not typically identify the value of a human life with the loss of expected earnings or with the differing amounts different people would pay to prevent a death, and in a serious argument for a World Bank initiative, Summers would probably not have done so. But why not? If economic costs and benefits are a good guide to what is harmful and beneficial, then they should likewise be a good guide to the allocation of risks

> In some cases market value seems to differ drastically from what most people would regard as moral value. The drug eflornithine is a highly effective "miracle" cure for sleeping sickness, which infects tens of thousands of people in Southern Africa. Until 1999 eflornithine was produced by a U.S. subsidiary of the Aventis company, but when eflornithine proved ineffective against cancer (its intended target), Aventis stopped making the drug, which is expensive to synthesize, and gave the production license to the World Health Organization. People who contract sleeping sickness are very poor, so there were no profits to be made in producing a cure for the disease. Only in early 2001, when stocks of the drug were almost exhausted, was the WHO able to find drug companies to manufacture it – and then only because the companies hoped to profit from marketing eflornithine in developed countries as a cream that removes facial hair. It is currently sold under the brand name Vaniqa. By market criteria, eflornithine is of value because it removes facial hair but is not worth producing as a cure for sleeping sickness.

Figure 2.5.1. Market Value versus Moral Value

of death and injury. Yet if they are not a good guide to the allocation of risks of death and injury, then why should one believe that they provide an acceptable way to measure benefit or harm?

Summers's memorandum is an amalgam of economic analysis and moral philosophy. Or perhaps it would be less misleading to say that the economic analysis in this memo already contains an inextricable dose of ethics. We do not pretend to have proven than ethics is mixed up in all of economics and cannot be separated out; indeed, in the Appendix we shall consider proposals for isolating a pure "value free" science of economics. Our point here is illustrative. Economics of the kind exemplified by Summers's memorandum is shot through with controversial ethical commitments. Figure 2.4.1 illustrated this point by listing a number of questions that must be faced by anyone concerned with the evaluation of policies and by presenting some of the alternatives to the answers mainstream economists have generally given.

2.6 School Vouchers

Our second example concerns a leading controversy in primary and secondary education: the proper role of competition and market incentives in the provision of schooling.

Economists distinguish between the case for public *funding* of some product or service and the case for direct public *provision* of that item by government workers (Schultze 1977). The argument for public funding is usually cast in terms of some public purpose in expanding production and consumption of the service above what the market will provide, and this purpose is distinguished from the question of the mechanism used to accomplish it. Regarding the mechanism, economists often argue that markets have advantages in terms of efficiency and freedom compared to direct provision. That markets possess these admirable features does not imply that it is always best to employ them, both because markets sometimes fail (as in the case of pollution) and because, even when they succeed, markets might undermine other values. However, economists are on the whole inclined to believe that markets are neutral with respect to other values and to take market failures as the only justification for government interference. Economists typically regard markets as serving given preferences: as means to whatever ends are desired. This is a controversial and substantial moral assumption that is more reasonable in some circumstances than others, as the case of public elementary and secondary education illustrates. Can it be right to rely on preferences when evaluating an institution, such as schooling, which so strongly shapes preferences?

The question of public provision versus public funding has arisen in a particularly sharp way in the context of elementary and secondary schooling in the United States. Milton Friedman (1962, ch. 6) argued that the social interest in ensuring that all young people receive an education did not imply that there needed to be government-operated schools. He proposed that support for schooling should be provided instead in the form of vouchers. Vouchers are certificates that families can use to purchase schooling; schools then redeem the vouchers for cash.

There are many possible variants on a voucher system. Vouchers could be a fixed amount per student, or they could vary with family income or the differing costs of educating students with special needs. Some voucher proposals would permit parents to supplement the value of the voucher, while others restrict families' spending on schools to the amount of the vouchers. Some voucher proposals would permit schools to select among their applicants, while others would require them to accept all applicants. The main attraction of vouchers, as an alternative to governmentally provided schools, is that they would induce schools to compete for students and thereby (it is argued) provide education of better quality at lower cost.

To consider the case for or against vouchers, it helps to step back and ask why governments need to encourage elementary and secondary schooling beyond the amounts and types families would choose themselves. What makes education different from "private" goods like groceries or automobiles? We shall distinguish three reasons that support government funding or provision of primary and secondary education: paternalistic, distributive, and political.

Paternalist justification. The most obvious reason for public support of education rests on the fact that primary and secondary education are crucial to people's life prospects. They determine the range of career opportunities as well as cultural and personal horizons available to people. Furthermore, the recipients – children – are not capable of making informed choices themselves. Parents generally take an interest in the future well-being of their children, and at least in developed societies that interest gives parents a strong reason to see to it that their children get educated. But the interests of parents and children are not the same, and the children's prospects should not depend entirely on the responsibility, interest, and capacities of their parents. Government has a (paternalistic) responsibility to children.

Distributive justification. By itself, the paternalist reason for public concern might be satisfied by requiring parents to school their children, leaving it to the market to supply the demand occasioned by that requirement. But this line of thought leads immediately to a distributive concern, for parents may lack the resources to provide adequate schooling for their children. Allied to this may be a broader distributive issue: a concern for distributive fairness or equality of opportunity demands an adequate level of education for children from disadvantaged families.

Political justification. Both the paternalistic and distributive arguments focus on the influence of education on individual outcomes and well-being. A third strand of argument is directly political. In his case for public funding of education, Friedman stresses that how well children are educated has consequences for society at large. A society's schools benefit many others in addition to those whom they educate by, for example, helping to protect and to enrich the society's culture. As we saw in the discussion of the Summers memorandum, there is a case for interference with the market when activities have costs or benefits that individuals are not paying for. So there is a narrowly economic case for a government role in ensuring that children are educated.

Among the possible benefits and costs of education to others which Friedman emphasizes are its effects on political institutions and decisions. Education plays an important role in shaping the values and attitudes of future citizens. Both the moral dispositions of members of society and their civic values are matters in which society as a whole has a stake. The question of just what are the "correct" moral, social, and political values to encourage – as well as the question of how best to encourage those values – is controversial. For example, people disagree about the extent to which fundamentalist

Christian or Islamic schools undermine democratic institutions. One important reason for public concern with elementary and secondary education lies in their consequences for moral and civic development.

How do these three types of reasons – paternalistic, distributive, and political – bear on the question of public funding versus public provision of schooling? A paternalist who worries that parents would not spend enough on schooling might also worry that parents will make poor choices among schooling options. Governments, it could be argued, are better able to make schooling choices than parents, who lack expert information and are not always well motivated. But in most societies, relations between parents and children are governed by a strong presumption in favor of parental authority. This presumption rests on the reasonable grounds that, in matters (such as education) that deeply concern parents, they'll try hard to figure out what's best. Parents know their children better than governments do and are typically more actively concerned with their welfare. Though there is an important case to be made for regulatory limits on parental discretion, paternalistic concerns do not provide a strong case for government provision. The fact that children cannot decide for themselves justifies limits to parental authority, but it does not undermine the case for decentralized market provision.

Distributive concerns lie behind other arguments for government provision. Many have argued that a voucher system would wind up being unfair to poor families. Vouchers that were equal for children of all family backgrounds and that parents could supplement would tend to reinforce inequalities, although this problem lessens as the size of the standard voucher grows. On the other hand, such a system would tend to undermine political support for generous vouchers, since these are less advantageous to the relatively affluent, who have the greatest political influence. As unfair as this variant of a voucher system would likely be, it is doubtful that it would be as unfair as the current system of funding public education in the United States, which – thanks to its reliance on local taxes that generate more revenue in wealthier areas – devotes a substantially larger share of public funds to educating children of those who are well-off than to educating poorer children. Since a voucher system would probably be funded at a state rather than a local level, and since it would be difficult to justify providing a larger voucher to richer children than to poorer, even a uniform and supplementable voucher system might well turn out to be fairer than the current system in the United States. Voucher policies that direct resources toward those who are in more need of help would be fairer still. Whether any such systems are politically feasible is a separate question.

The most challenging and the philosophically most interesting difficulties facing voucher schemes relate to the last reason for public support of schools: their influence on the development of moral and civic "character" among students. As we have observed, judgments about what kinds of beliefs and values to promote among schoolchildren are political and moral judgments. At the most abstract level there might be a reasonable degree of agreement on the values to be promoted – tolerance, mutual respect, support for democratic participation – but schools will be pressed to make many decisions that are more fine-grained and controversial. Should Muslim girls be allowed to wear head scarves in publicly funded schools in France? To what degree should American history emphasize the contributions of members of minority groups? At what age (if any) is sex education appropriate, and what should children be told?

Questions like these, easily multiplied, are important for two reasons: first, they concern the content and not simply the amount of education children receive; second, the content questions are ones on which the society, as a political entity, has a central interest (insofar as civic and moral education are important aims of public subsidy of schools). Thus, while it is plausible on grounds of both liberty and efficiency to say that a child's parents are as well positioned as anyone (within wide limits) to judge which educational investments are best for their child, there is no similar case that parents should have a uniquely important role in deciding what kind of civic preparation their own child should receive. The question is not what is best for the individual child but what is best for the community. Even though a liberal state aims ultimately to serve the interests of individuals, it can do so only if it can engender support for its institutions. The degree to which parents should have authority over the civic education of their children is itself a controversial political question.

At the same time that we find these considerations persuasive, we also find them worrisome because what some people call political education, others call political indoctrination. To the degree that the civic purposes of schooling are taken seriously, the question of who decides what will be taught becomes crucial. Economists who regard markets as neutral with respect to values other than efficiency and freedom would argue that these questions, as important as they may be, are independent of the question of whether there should be government provision as well as government funding of education. Either way, one faces the same questions concerning the content of civic education and concerning who is responsible for determining that content. In a voucher system, couldn't government require whatever civic education is judged appropriate? But a voucher system is not neutral. It is bound to diminish the influence of local school districts, whose children will have wider choices, and it is bound to diminish the influence of teachers, who as employees of private firms will have fewer common interests, less employment security, and consequently less responsibility and power. To what degree can elements of decentralization and competition be introduced into public schools compatibly with the appropriate extent and kind of democratic control?

It is of course too simple to characterize the options as simply traditional public schools versus complete market provision through vouchers. Governments can hire private contractors to provide schooling without offering families choice, and a variety of devices, including charter schools, can provide elements of decentralization and parental choice in systems that do not feature profit-seeking firms providing schooling.

However, it would be naive to opt for thoroughgoing market provision through vouchers and then assume that government regulation will protect the character of democratic education. Unless the development of students' moral and civic values is unimportant, which is hard to believe, decentralization and competition raise serious concerns. The view of the market as a morally neutral mechanism is not innocuous (Gutmann 1987, 2000).

Whether to shift from a system of government provision to a voucher system is a momentous social choice, since doing so could significantly change the character of individual, social, and political life. Individuals do not interact only via the market, and political life does not consist only of voting. Recent discussion of vouchers, while often citing civic preparation as a basic rationale for schooling, focuses primarily on the effects vouchers would have on the efficiency of schools, with some secondary concern about the distribution of education. In empirical studies, including those based on social experiments, the assumption is sometimes maintained that scores on standardized tests are an adequate proxy for the "output" of schooling, with the consequence that questions about the content of schooling are either ignored or reduced to the influence of differences in content on test scores.

Although questions of efficiency and distribution are surely important, reducing questions about vouchers – or about different systems of school governance more generally – to these matters begs important questions. In addition to the issues concerning civic education, what are the effects of dismantling publicly provided education on social solidarity and on the existence of a common culture? Classrooms in many countries are already segregated by class, race, and religion, and voucher systems have the potential to exaggerate that segregation. Although many parents want their children to get to know children from different backgrounds, other parents would prefer that their children associate only with others of the "right" class, religion, or race. Overt discrimination by schools can be prohibited, but self-selection is difficult to control; the possibility that vouchers will help bring about a future in which children grow up wearing the blinders of narrow communities is not far-fetched. Such a system might be splendidly efficient, and possibly even fair, while at the same time causing the dissolution of community or the collapse of national identity. Society does not consist solely of market exchange within political constraints, and markets are not morally neutral apart from their contributions to freedom and efficiency. Awareness of the ways in which markets compete with other institutions and of their bearing on values such as justice, solidarity, and community can help economists contribute more thoughtfully to debates about schooling.

2.7 Conclusions

Our purposes in discussing Summers's memorandum and school vouchers are, we repeat, only illustrative. Our goal is not to pronounce on whether school vouchers are a good idea or on whether pollution ought to be exported from wealthy countries to LDCs, and it is certainly not our intention to maintain that welfare economics is wicked or misguided. Indeed, it is the unabashed clearheadedness of Summers's and Friedman's applications of familiar economic ideas that makes the examples powerful. The point of the examples is instead to show that mainstream normative economics rests on substantial and controversial moral premises and thereby to make plausible our claim that knowing ethics can contribute to doing economics.

Suggestions for Further Reading

General overviews of the relations between economics and ethics can be found in Buchanan (1985), Hamlin (1986), and Sen (1987a).

There has not been any scholarly discussion of Lawrence Summers's memorandum, although it raises general issues concerning cost–benefit analysis, which have been extensively discussed. See Chapter 9.

The classic statement of the case for vouchers as a mechanism for public financing of education appears in Friedman (1962, ch. 6). For an influential empirical study see Chubb and Moe (1990). Criticisms of vouchers from the standpoint of democratic theory are advanced in Gutmann (1987, 2000) and from the standpoint of justice in Brighouse (2000).

THREE

Ethics in Positive Economics: Two Examples

Although some economists have envisioned the possibility of a value-neutral normative economics that merely investigates the consequences of policies for the satisfaction of preferences, most would be willing to concede that moral judgments play a role in normative economics. For that reason, they might also concede that there would be some benefit in thinking about ethics. But when it comes to *positive* economics – the attempt to predict and explain economic outcomes and processes – few economists see any role for ethics.

Although it is sometimes possible (even in economics) to investigate features of the world without evaluating them, it is often very hard to do so. When economics bears strongly and immediately on people's interests, those interests and the moral commitments that are relevant to them are likely to influence the questions economists ask and the answers they defend. It is not obvious that economists should aim for moral neutrality, even if they could achieve it. It is better to understand the values that are at stake than to pretend to transcend them.

Familiarity with moral philosophy can help here. Although values will be most obvious in debates over policy questions that bear directly on people's interests – such as debates over energy policy, prescription drug benefits, agricultural policy, or tax policy – moral commitments can have a much broader influence. We shall show how using two examples. Both concern matters of fact about economies that bear on policy questions, though the second is less immediately relevant to policy than is the first. The first example, which concerns whether there is any involuntary unemployment, is likely to seem odd to readers who are not familiar with economics. How, one might ask, could anybody doubt that some people are unable to get a job? We shall shortly see. Whether workers are unemployed voluntarily or involuntarily seems to bear on the question of what sort of unemployment compensation they should be paid, but its normative implications may go deeper. The second example, which involves Paul Samuelson's overlapping generations model, might appear to have no bearing on policy; but as we shall argue, its normative implications have driven what purported to be a theoretical debate in positive economics.

3.1 Is Unemployment Involuntary?

"Macroeconomics" was born during the worldwide economic depression of the 1930s with the publication of John Maynard Keynes's *General Theory of Employment, Interest, and Money.* Among the central concerns of macroeconomics have been explaining and figuring out how to remedy mass involuntary unemployment. More recently, influential economists have argued that this whole line of inquiry rests on a mistake: that there is no such thing as involuntary unemployment; or that it is at most a transitory occurrence of no interest to economic theory and an inappropriate focus of economic policy.

In everyday language, a person is "voluntarily" unemployed if she quits her job and chooses not to seek another one. A person is involuntarily unemployed if she loses her job and is unsuccessful in finding another one. People who are voluntarily unemployed are "at leisure": they choose not to be employed because they prefer some other set of activities to working. Indeed, in the way economic statistics are kept, people in the latter category are not counted as unemployed at all – to be considered unemployed you must be looking for work.

The disputes between those who seek to explain and those who seek to dismiss the phenomenon of involuntary unemployment involve difficult matters of economic theory and methodology, but the moral weight of the idea of "voluntariness" is significant as well. Most Keynesian economists, who believe involuntary unemployment is a real and important phenomenon, are also inclined to believe that activist monetary and fiscal policies are essential in managing modern economies. Thus, Shapiro and Stiglitz write, "To us, involuntary unemployment is a real and important phenomenon with grave social consequences that needs to be explained and understood" (1985, p. 1217).

Those who are skeptical of involuntary unemployment tend to believe that individual activities coordinated by markets work out well and that government action is more likely to cause than to cure economic difficulties. Thus Robert Lucas suggests that unemployment is like any other economic choice:

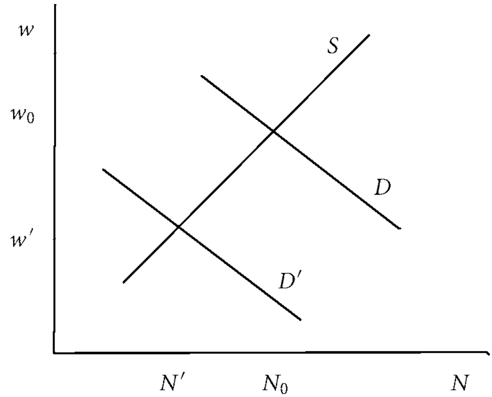

Figure 3.1.1

the unemployed worker at any time can always find *some* job at once, and a firm can always fill a vacancy instantaneously. That neither typically does so *by choice* is not difficult to understand given the quality of the jobs and the employees which are easiest to find. Thus there is an involuntary element in *all* unemployment, in the sense that no one chooses bad luck over good; there is also a voluntary element in all unemployment, in the sense that however miserable one's current work options, one can always choose to accept them. (1978, p. 354)

But Lucas emphasizes the voluntary aspect: "By treating all unemployment as voluntary, this work has led to the examination of alternative arrangements which firms and employees might choose to adopt for dealing with fluctuations in product demand" (p. 356).

The analytical and methodological issues raised by this dispute are deep and complex, and we can only sketch them here. The logic of standard microeconomic theory states that if demand for some product or service dropped, producing an excess supply of it, then the price would promptly drop until the excess supply was eliminated. So, viewing labor as such a service, there should be no excess supply of labor or unemployment. Keynesian analyses claim that the self-correcting mechanisms that smoothly adjust for fluctuations in supply and demand in commodity markets somehow go wrong at the level of the market for labor in the economy as a whole.

Simplifying drastically, one might depict the market for labor in a given industry as in Figure 3.1.1. The supply curve, S, shows how much labor workers are willing to supply at each wage. We are oversimplifying and equating this with how many workers want to work. The demand curves D and D' show how much labor (how many workers) firms want to employ at each wage. The demand curves slope downward, because firms are more eager to hire when wages are lower. The supply curve slopes upwards because workers are willing to work more at higher wages. Here D shows the demand for labor before the shocks that gave rise to the Great Depression, such as the Wall Street stock market crash of 1929. At the "equilibrium" wage w_0, the number of workers whom firms want to employ and the number of workers who want to work both equal N_0, and there is neither unemployment nor excess demand. There is then a shock that lowers demand to D'. Though there will (in reality) be confusion and transitory unemployment, the theory abstracts from this and shows that equilibrium will be restored, though at a lower wage of w' and a lessened employment N'. Fewer workers are employed, but none are unemployed. Everyone who wants work at the new wage w' can find work.

Keynes argued that in the transition from an individual market to the system as a whole – from micro- to macroeconomics – this story of how markets equate supply and demand breaks down. Unlike in an individual market, a drop in the wages of labor across the economy lowers workers' purchasing power and may reduce aggregate demand. Exactly where and how the analogue of the microeconomic argument breaks down and whether there exist any mechanisms within the market system that will correct the imbalance in aggregate supply and demand are questions that have shaped the evolution of macroeconomic theory for three quarters of a century. But, at least in Keynes's view, unemployment cannot be attributed to the unwillingness of workers to accept lower wages, and it falls to government to provide the proper cure, which is to generate higher aggregate demand.

However, as Michel De Vroey has shown (2004), finding a satisfactory way to incorporate involuntary unemployment into a coherent choice-theoretical account of the economy has proven to be enormously difficult.

As a simple illustration of this challenge, suppose one redraws the supply curve to represent the reluctance of workers to accept employment at wages below w_0 (see Figure 3.1.2). This has been proposed as one way of characterizing the assumption of "rigid wages," which on some interpretations is a key assumption of Keynesian economics. The unwillingness of workers to accept employment at wages below w_0 is depicted by drawing the supply curve as horizontal to the left of the old labor market equilibrium (at w_0 and N_0). Because workers refuse to accept jobs for wages less than w_0, employment drops all the way to N^*, where the supply curve intersects the new demand curve. Orthodox theorists did not believe that a supply curve such as the one shown would continue to reflect workers' preferences, believing rather that workers' expectations that they would find work at w_0 would collapse.

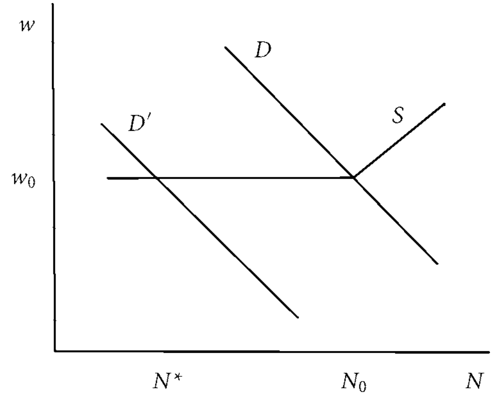

Figure 3.1.2

Although some theorists attempted to explain involuntary unemployment by means of a horizontal supply curve, it seems instead (as De Vroey argues) that, along the horizontal segment of S, workers must be indifferent between working and not working. They prefer not working to accepting jobs at wages below w_0, and they prefer accepting jobs at wages above w_0 to not working. The intersection of D' and S is a point of market *equilibrium,* and at w_0, workers do not prefer to supply more or less labor than N^*. Suppose, for example, that $w_0 = \$8.00$ per hour. What the diagram says is that any individual worker willing to work for \$7.99 per hour could readily find a job. The decisions of such workers to decline an offer of a penny less per hour would appear to be voluntary. Since workers are doing what they prefer to do and since nothing prevents them from accepting lower wages, Figure 3.1.2 does not depict a situation involving involuntary unemployment. (This pertains to the labor force as a whole. At a more detailed level, it is possible that some workers coercively prevent others from accepting jobs at lower wages.)

Indeed, it is arguable that in Figure 3.1.2 there is no unemployment at all – at least as unemployment is conventionally measured. It is not the case that the $N_0 - N^*$ workers who were employed before the shift in demand and who are not now employed would strictly prefer working at the going wage to not working. Since they do not prefer working to not working at w_0, they have no reason to seek work at w_0 and thus, on one plausible construal of what it means to be unemployed, they are not literally *unemployed* at all! They are instead at leisure. Moreover, within the framework of economic theory, the situation depicted in Figure 3.1.2 is "efficient," too. Workers and firms are behaving as they prefer. Nothing in the graph captures the fact that something has gone wrong. The enormous human tragedy of the Great Depression has disappeared. The attempt to comprehend involuntary unemployment depicted in Figure 3.1.2 is neither plausible in itself, because workers' expectations will adjust, nor adequate to the theoretical problem, because workers are not indifferent between working and remaining unemployed.

In order for there to be involuntary unemployment, it must be the case that workers would strictly prefer to be working at the market wage but are unable to find a job – or, in other words, that the jobs and wages available to those who have lost their jobs are substantially worse than those of workers with similar qualifications who are still employed. As De Vroey and others have maintained, that means that actual employment at the market wage must be off the supply curve. In Figure 3.1.1, for example, the wage must somehow get "stuck" above w', even though workers are willing to work for less.

But what keeps the wage from falling? Except as a bit of transitory confusion, from which economic analysis abstracts, when demand shifts as it does in Figure 3.1.1 wages *must* fall unless (as in Figure 3.1.2) workers are indifferent between working and not working at the old wage w_0. Simply to assert that wages are "sticky" does not explain why. Apart from totalitarian control over labor, how is it possible to be off the labor supply curve? However "obvious" it may be that a good deal of unemployment is involuntary, when one steps into the theoretical world of economics and focuses on graphs like the two shown here, it is hard to understand how involuntary unemployment is possible.

This is the point that Lucas and other members of the "rational expectations" school focus on. Their view – which is shared by many economists of other schools, too – is that an adequate analysis of economic activity must be built up from proper "micro-foundations." Theorists must assume that every actor is acting rationally based on the information available to him and that every actor invests rationally in acquiring information. Unemployment is then a voluntary consequence of rational search. In searching for a new job, an unemployed worker is viewed as continually weighing the quality of the best option she has found against the likelihood of finding a better option by searching longer. Periods of high unemployment occur when for some reason more people are searching and the average search is longer. Lucas, however, hastens to counter any "blaming the victim" accusation that might follow such a claim: "treating unemployment as a voluntary response to an unwelcome situation does not commit oneself to normative nonsense like blaming depressions on lazy workers" (1978, p. 356).

Whether some version of the rational expectations program can ultimately succeed in reducing macroeconomics to microeconomics is a hotly disputed question. But even if a worker's unemployment can be shown to be rational from her point of view, is that enough to make it "voluntary"? Both Keynesians and "rational expectations" theorists agree that, in depressions, the alternatives facing workers become sharply worse. And both agree that the project of explaining in detail how labor markets work in these circumstances is a difficult problem, not fully resolved. In his denial that he is blaming the victim, Lucas shows his awareness that the word "voluntary" in this context carries significant normative freight: if unemployment is voluntary, then it is the worker's own fault and of lesser social concern. Similarly, in speaking of the "grave social consequences" of involuntary unemployment, Shapiro and Stiglitz ally themselves with a critical and activist social agenda. In our view, a better understanding of ethical theory has a good deal to contribute here.

Both in everyday and in philosophical discourse, whether a choice is "voluntary" depends both on the quality of the alternatives and on the circumstances bringing about the occasion for choice. The choice posed by the robber who demands "your money or your life" is involuntary both because the alternative to parting with one's purse is totally unacceptable and because the choice is brought about in an illegitimate way. The obvious fact that the choice is rational is not enough to make it voluntary. Robert Solow captures this understanding of "voluntariness":

> You can take the definition that is actually there in the General Theory, and loosen it up so that it says that someone is involuntarily unemployed if he or she would be prepared to work at a job that he or she knows how to do, at the going wage, and cannot find employment. That is good enough for me. You will notice that by this definition a person can be involuntar[il]y unemployed and employed at the same time. If you take an inferior job you can still be an involuntar[il]y unemployed skilled worker. (Snowdon and Vane 1999, p. 287)

From this point of view, it is apparent that *not all rational choices are voluntary.* To claim that people are involuntarily unemployed is not to claim that they cannot find any job at all. Unemployment is involuntary when people do not have any job opportunities that are reasonably good by the standards of their society. Since whether the alternatives to unemployment are "reasonably good" is clearly a normative question, the concept of involuntary unemployment is itself, in part, a normative concept. (The normative criterion in Solow's definition is that a "reasonably good" alternative for one who is unemployed is a job–wage package that others with the same qualifications are actually enjoying.) Those who claim that – even in a depression – unemployment is voluntary might then be understood in one of three ways. One is that they have made a mistake that could have been prevented by a greater familiarity with ethics: they are confusing "voluntary" with "rational." The available jobs – selling apples, shining shoes – are worse than not working, and so folks are doing their best, being rational, by being unemployed. But rationally refusing such poor alternatives does not imply that people are voluntarily unemployed. The following quotation from David Colander may illustrate this kind of confusion about what "voluntary" means:

> The same unemployment can be viewed as voluntary or involuntary depending from which perspective you look at it. What I think is important is whether there are some policies that can eliminate it. Let's say we accept all unemployment as voluntary. Does that mean we should see that unemployment as something good? In my view no. Say I put a gun to your head and say: "Your money or your life", I suspect you would voluntarily give me your money. But that does not mean that society should allow such actions. Given institutions all results are voluntary to the degree that people accept those institutions. (Snowdon and Vane 1999, p. 233)

A second possibility is that Lucas and others disagree about whether the alternatives to continuing to search are reasonably good, particularly as temporary expedients. In this case, the dispute about whether unemployment is involuntary would be almost entirely a normative matter. The third possibility – which is related to the second – is that those who regard unemployment as voluntary are saying that spells of unemployment, even occasionally severe and widespread spells, are part of the normal workings of a market economy and (viewed perhaps in a lifetime perspective) having to endure such spells is a reasonably good alternative to arrangements that would eliminate the risk of unemployment. To the extent that unemployed workers previously faced choices between secure jobs with no stretches of unemployment and less secure jobs that were otherwise more attractive, they had a reasonably good alternative and may be deemed to be unemployed voluntarily.

"Voluntariness" is at root a moral notion, closely related to judgments of freedom and personal responsibility. If unemployment is involuntary, then a depression threatens economic freedom as well as wasting resources and lowering welfare. Emphasizing that workers are not responsible for their fate invites a search for governmental remedies. Robert Skidelsky's (1992) biography of Keynes describes the importance Keynes attached to these emphases as he developed his arguments through the 1930s. If unemployment is voluntary – a matter of searching longer rather than accepting an available job – then it is easier to see it as part of the normal workings of a market economy, as relatively unthreatening to basic social values, and as less obviously somebody's responsibility to fix. If workers did not have mistakenly high expectations, the unemployment would disappear.

The roots of these contrasting ethical perspectives lie in differences in attitudes toward market economies and government intervention, which align plausibly (though not necessarily) with differences in theoretical view. Thus Keynes himself and Keynesian followers have been inclined to assume that widespread unemployment reveals a basic flaw in both market economies and microeconomic theory, and they have argued that a different kind of theory is needed for the economy as a whole than for its parts. On the other hand, those who see unemployment as voluntary generally believe that market economies perform well without government interference and that macroeconomic analysis will ultimately prove to be consistent with every individual choosing optimally under conditions of rational expectations.

The links we have noted between judgments about the voluntariness or otherwise of unemployment, about the adequacy or inadequacy of microeconomic equilibrium theory to explain macroeconomic phenomena, and about the role government should play in reducing unemployment are not logically tight. One can believe that unemployment is both involuntary and the source of grave social consequences yet still think there is nothing government can or should do about it. In contrast, one can believe that unemployment is involuntary (in the moral sense just defined) and still think that a suitably developed theory of rational expectations equilibrium is the best approach to explaining it. Still, the links among these views are not coincidental, and the energy that goes into defending alternative positive theories of unemployment derives in large part from the normative views that frequently accompany particular theoretical positions. Though disagreements would obviously remain, a better understanding of ethics (in particular, of voluntary choice) would help to refine them.

3.2 Overlapping Generations

One striking and fruitful theoretical contribution of the 1950s was Paul Samuelson's model of overlapping generations. In his celebrated paper, "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money" (1958), Samuelson addresses the following problem: Suppose individuals want to save for their old age, when they cannot produce anything, and suppose nothing lasts from one period to the next. All people can do if they don't want to starve is to strike a bargain during their working years so that those who are younger will support them when they are retired. In a world of endlessly overlapping generations of workers and retirees, what will the pattern of interest rates be? This appears to be a factual and not an evaluative question, and obviously it is a very abstract issue.

To isolate the effect of this desire to provide for one's old age from other factors that influence the rate of interest, Samuelson formulates an extremely simplified model in which everyone lives exactly three periods. In each of the first two periods of their lives, people produce one unit of a completely perishable consumption good, but people produce nothing in the third period of life. To simplify further, Samuelson considers first an absolutely unchanging economy and then one growing eternally with an unchanging rate of growth and an unchanging rate of interest. In these models a "biological" rate of interest equal to the rate of growth clears all markets and is optimal (in the sense to be discussed in Section 5.2). For example, in the case of no growth at all, individuals can "save" one third of their output in each of the two periods in which they are productive and then maintain the same level of consumption (2/3 unit) during retirement.

But is this outcome socially feasible? Remember that output is perishable and cannot literally be saved. The one third of their output that young and mature workers do not consume constitutes the 2/3 unit of the good that retired workers consume. What is here called "savings" is in fact a transfer to retired workers. But why should young or mature workers transfer anything to the retirees? After all, the retirees never gave anything to the current workers, and the retirees are in no position to retaliate if the young or mature workers refuse to "save." It seems as if the biological interest-rate solution (which in this simple example with a zero rate of growth involves a zero interest rate) is impossible. All that remains – if people are to avoid starving when retired – is for the mature workers to transfer goods to the young workers in return for repayment in the next period when this period's mature workers will be retired. Hence it seems that, on a competitive market, the only bargains people can strike will result in lavish living during the first period of life and penury during the remaining parts.

At this point Samuelson's essay takes a surprising turn. He simply drops the question he began with concerning the effects on the rate of interest of the desire to save for one's retirement! Instead his focus shifts to a theoretical demonstration that people would be better-off with the "biological" rate of interest (equal to the rate of growth) than with the competitive solution and to an exploration of how such a biological interest-rate regime could be achieved. He argues that an arrangement like the one just described, whereby everybody consumes 2/3 unit in each period, could be achieved by means of a social contract in which individuals collectively agree to an obligation to support retirees or if people agree to accept "fiat money" from retirees. ("Fiat money" has value only because it is common knowledge that individuals are willing to accept it as payment for goods; Samuelson argues that fiat money is, in fact, very much like a social contract.) Samuelson treats his conclusions as important normatively, but an outsider might wonder why anybody should care about the efficiency of a biological interest rate in an utterly fictitious economic model.

The immediate responses to Samuelson's essay by Abba Lerner (1959a,b) and William Meckling (1960a,b) are also surprising. They purport to find analytical mistakes in Samuelson's work (erroneously – see Samuelson 1959 and 1960), but in fact they are both driven by normative concerns. From a utilitarian perspective (see Chapter 7) and from the "left" of Samuelson, Lerner denies the ethical desirability of a rate of interest equal to the rate of growth. Lerner points out that maximum average utility in any given period is achieved if consumption is equal across the three generations. In the case of a zero interest rate, Lerner's utilitarian plan corresponds to the distribution that would result from a biological interest rate, but the two distributions differ for any positive rate of interest. Lerner argues that Samuelson is mistaken in regarding provision for the retired as savings because, in Samuelson's model, nothing can be invested or held over from one period to the next. Social security should instead be regarded as a program in which workers transfer income to retirees. Normative policy disagreements concerning the attractiveness of utilitarianism and concerning whether social security can be distinguished from transfer programs are central to the controversy between Samuelson and Lerner.

Meckling complains from a more conservative perspective that, since no one can repay what workers transfer to retirees, the only way to save for retirement involves mature workers making loans to young workers, who consume more than they produce their first year. This solution should not be regarded as suboptimal, because Samuelson's biological rate of interest is impossible.

Why? Meckling's main argument against the possibility of a social contract is that it is against the self-interest of young workers to transfer income to retirees and that they will not voluntarily do so: "the zero-interest-rate equilibrium can prevail only if the sheriff is retained on a permanent basis" (1960b, p. 84). But the sheriff is just as necessary if there is no social contract, for whether in repayment or as required by the social contract, it is never in anybody's self-interest (in this model) to transfer anything to retirees. Contracts to do so require enforcement. Furthermore, if an individual believes that later generations will respect the social contract or accept fiat money only if she does so, then it is in her interest to honor the social contract and accept the fiat money. So it is doubtful whether a very tough sheriff is needed after all.

What's really bothering Meckling, as well as Lerner and Samuelson, is a normative rather than an analytic issue. In this hypothetical instance, perfect competition does not lead to a good result. As we shall see later, perfect competition is widely regarded as an ideal both by economists who support government intervention to remedy market failures and by economists who believe that government ought not to interfere in economic affairs. That is why Samuelson is so interested in demonstrating that the competitive solution without a social contract or fiat money is suboptimal and why Meckling is so disturbed by the demonstration. Lerner, in contrast, doubts that markets are the right way to make provision for retirement, and he seeks to shift the question to one of establishing welfare-maximizing levels of transfer payments to retirees. What appears on the surface to be a dispute about a highly theoretical question in positive economics actually turns on the evaluative commitments that economists have toward competitive markets.

3.3 Conclusions

That the arguments economists make about important matters include an ethical component should be neither a surprise nor a cause for regret. As our second example shows, this ethical dimension is present even in a controversy concerning abstract overlapping generations models, let alone in contexts (such as unemployment or school vouchers) fraught with policy implications. But it would, we believe, be a great mistake to conclude that economists are at fault for allowing value judgments to "intrude" on supposedly morally neutral analyses. So long as economists deal with socially consequential subjects, entanglements of fact and value are unavoidable and proper.

If there is any criticism of the economists in the examples discussed in this chapter, it is that they have not always achieved the same level of analytical clarity and critical self-awareness in the parts of their work that connect with ethics as they have sought in the rest of their empirical and analytical work. Our aim in the following chapters is to clarify ethical questions and to help readers to think about them rigorously.

Suggestions for Further Reading

Our discussion of involuntary unemployment was heavily influenced by De Vroey (2004). The debate about whether unemployment is involuntary is connected to the larger question of whether the transactions that compose a market economy are voluntary. A classic defense of that view is in Friedman (1962, ch. 2). An economist's argument that market systems have coercive elements can be found in Okun (1975). On voluntariness in market transactions see also Peter (2004).

Paul Samuelson's (1958) overlapping generations model has been much discussed. For an overview of its influences and of the permutations the basic model has undergone, see Hausman (1992, ch. 7).

PART ONE

RATIONALITY AND MORALITY

When we say that it is morally right to relieve famine victims, we are expressing our approval of famine relief and we are at least suggesting that people ought to come to the aid of those in danger of starvation. Morality is both *expressive* and *normative* as opposed to merely descriptive.

So is rationality. When we say that it is rational for individuals to have medical insurance, we are expressing approval of doing so and suggesting that people ought to make sure they are insured. Similarly, to characterize a choice as irrational is usually to condemn it, and not simply to describe it.

Not only are morality and rationality alike in these ways, but "rational" is often used (as in the previous example) as a synonym for "prudent," and prudence is a morally admirable character trait or moral virtue. Yet morality and rationality are of course not the same thing.

How are morality and rationality related? Is it always rational to be moral? These general philosophical questions are critical to understanding the relations between ethics and economics, because economics is built around a theory of rationality. We explore the relations between economics and rationality in Chapter 5 and between rationality, norms, and morality in Chapter 6.

But first, Chapter 4 presents the standard theory of rationality and its extensions to circumstances of risk and uncertainty; it also considers some of the objections to which the theory has been subject. Chapter 4 is the only technical chapter in the first three parts of this book.

Chapter 5 examines the role of the standard theory of rationality in both positive and normative economics, and it explains how the theory ties positive and normative economics together and makes both appear to be more plausible than they otherwise would be.

Chapter 6, the last chapter in Part I, deals with questions raised by the effects of people's moral commitments on their behavior. How (if at all) can economic theory take account of these effects? What is the relation between morality and preference? Can one (or should one) develop a conception of economic rationality that permits a distinction between acting morally and maximizing utility? If the two diverge, is it then sometimes irrational to act morally or immoral to act rationally?

FOUR

Rationality

Economics portrays agents as choosing rationally. Many generalizations in economics concerning how people do in fact choose are also claims about how agents *ought rationally* to choose. This fact distinguishes economics from the natural sciences, whose quarks and polymers do not choose at all and whose theories have no comparable normative dimension.

The theory of rational choice that dominates economics derives from an everyday theory of human choice, which has been called "folk psychology." This theory takes actions to derive jointly from beliefs and from a wide array of motivational factors such as urges, emotions, habits, and commitments. So, for example, when one rainy Friday night a hungry student named Ellen takes a frozen pizza out of the refrigerator, unwraps it, puts it in a stove, and turns knobs on the stove, we folk psychologists explain Ellen's action by Ellen's beliefs – including especially her beliefs that turning the knobs will cause the stove to heat the pizza – and by her desire to eat hot pizza.

This sort of explanation is familiar but not very satisfactory. Ellen might also like to eat her pizza frozen, or she might also have a desire to reheat some leftover meatloaf. Or she might rather skip dinner and keep studying decision theory. What explains her action is not merely wanting to eat hot pizza (plus possessing the requisite beliefs) but also wanting to do this as much or more than she wants to do any of the feasible alternatives.

One way to tighten up the folk-psychological account of action is to replace the noncomparative notion of a "desire" with the comparative notion of "preference." One can then explain the interaction between Ellen and her stove in terms of physical constraints, Ellen's beliefs about the outcomes of the alternative actions she can undertake that Friday night, and her overall ranking of those outcomes. One explains the actual pizza warming by showing that Ellen ranks it at least as highly as any feasible alternative.

In taking "preferences" to incorporate everything relevant to choice, one has arrived at the view of preference and choice that economists have generally adopted. Economists regard choices not as mental determinations but rather as actions that arise from *constraints, preferences,* and *expectations* (or beliefs). Economists typically take preferences to be predetermined or "given" facts about individuals and not themselves in need of explanation or subject to rational appraisal. Economic analyses begin with an individual's preferences, whatever they may be.

Rationality enters the picture as soon as one examines the relations among choices, preferences, and beliefs. Choice is rational when it is determined by a rational set of beliefs and preferences. The rationality of sets of preferences and beliefs is defined within "utility theory." Although this is not a technical book, our discussion cannot proceed without some discussion of the technicalities of utility theory.

4.1 Certainty and Ordinal Utility Theory

Suppose one is concerned with the choices, preferences, and beliefs of an agent named "Q." In circumstances of complete certainty, Q chooses rationally if her preferences are rational and if there is nothing available that Q prefers to what she chooses. Except in the case of ties, one can simplify: Q is rational if her preferences are rational and she chooses what she most prefers among those things she can obtain.

Q's preferences are rational if they are transitive and complete. Q's preferences are *transitive* if and only if, for all objects of choice (or "alternatives") x, y, and z, if Q prefers x to y and y to z then Q prefers x to z. Similarly, if Q is indifferent between x and y and between y and z, then she is indifferent between x and z. It is plausible to require that rational preferences be transitive. Suppose:

1. Q has intransitive preferences (she prefers x to y and y to z, but she prefers z to x);
2. Q already has some of y; and
3. Q is willing to pay a penny to trade what she has for something she prefers.

Q will then pay a penny to trade y to get x, another penny to trade x to get z, another penny to trade z to get y, another penny to trade y to get x, and so on until Q realizes that her preferences leave her vulnerable to manipulation. Hence Q will not cling to intransitive preferences. Since rational choices should further one's ends, intransitive choices are not rational. This argument showing the irrationality of intransitive preferences is called "the money pump argument." It is persuasive, though some would question it (Schick 1986).

Q's preferences are *complete* if, for all options *x* and *y*, either Q prefers *x* to *y* or Q prefers *y* to *x* or Q is indifferent between *x* and *y*. If Q's preferences are complete then Q is never unable to rank *x* and *y*. If asked to choose between two sealed shoeboxes, one of which contains $10 and the other $1, Q will not be indifferent; but if she doesn't know which box contains the $10 then she may not be able to form a preference. Such problems will not come up in conditions of certainty, and completeness may be a reasonable simplification when there are no uncertainties. But it is questionable whether completeness is a condition of rationality (Levi 1980) or a precondition for the capacity to choose rationally. Note the difference between being indifferent and being unable to rank. Indifference is ranking equally, not an inability to rank.

We regard Q's preferences as a subjective state of Q: Q's evaluation of states of affairs with respect to everything relevant to choice. Q's preferences and Q's beliefs then jointly explain her choices. But many economists would disagree. In accordance with the theory of "revealed preference," they instead attempt to *define* preference in terms of choice. Provided that consistency requirements are met, choosing *x* when one might have had *y* at a lower cost reveals a preference for *x* over *y*.

Preferences, as we conceive of them, coincide with revealed preferences when the alternatives are restricted to just those things among which an agent chooses. In the example, Ellen's preferences for putting the pizza in the oven and turning its knobs are revealed preferences. These preferences stand in a one-to-one relation to specific actions. Their influence on action is not mediated by belief. On the contrary, Ellen's beliefs and her more general preferences for eating pizza versus meatloaf have already done their work in determining her preferences for putting the pizza in the oven and adjusting its controls. If economists never needed to refer to people's evaluative attitudes over things other than those among which they are directly choosing, then the theory of revealed preference would be a convenient simplification. In everyday consumer choice – the realm in which the theory of revealed preference is most often used – this condition is sometimes reasonably well met.

But economists need to be able to predict how people's choices will change when they acquire new information, and in order to do this they need a more general and less behavioristic notion of preference. When corporations report large unexpected quarterly losses, economists predict that the price of their stock will go down. Why? To say that people now reveal a preference to sell this stock simply shifts the question. To explain why people want to sell the stock, economists must cite the preference for profitable investments. But this is not a revealed preference. Unlike a revealed preference, its connection to action is via belief (about which investments are profitable) rather than direct. The same kind of point can arise in the context of consumer choice if, say, evidence of a link to heart disease reduces demand for foods containing trans fat. In Chapter 14 we shall see that preferences could not play the role that they do in game theory if they were revealed preferences.

Though choices are crucial evidence concerning preferences, we think that revealed preference theory is mistaken to identify preference and choice. Keeping them distinct is useful (1) for clarity about what preference and choice mean, (2) to permit preferences to explain choices and to show how choices depend on beliefs, and (3) to leave space for the possibility that agents may choose actions that do not maximally satisfy preferences (see Sen 1973, 1977).

Completeness and transitivity together establish a *weak ordering* of any finite set of alternatives. What this means is that one can put the alternatives in a long list, with the alternatives one prefers in higher rows and the alternatives among which one is indifferent in the same row. Completeness implies that every alternative goes into some row, and transitivity ensures that alternatives go into only one row. Given such a list, one can assign numbers to options such that preferred options get higher numbers and indifferent options get the same number. If Q prefers a particular portable computer to \$500 and \$500 to a Big Mac, then economists can represent Q's preferences by assigning numbers to these three alternatives. For example: computer, 12; \$500, 7; Big Mac, 2. But any other three numbers in the same order – such as (100, 97, 4) or (2,000,000, 92, 91) – would do just as well. This seems bizarre, because people are used to attaching significance to magnitudes, not just comparisons. But the sole purpose of the numbers here is to show the ranking. Any such assignment of numbers is an *ordinal utility function.* "Utility" here does not refer to usefulness or pleasure. A utility function is only a way of *representing* a preference ranking – that is, a ranking of alternatives with respect to everything relevant to choice. The *ordinal representation theorem* states that if an individual's preferences are complete, transitive, and satisfy a further "continuity" requirement, then they may be represented by a continuous real-valued ordinal utility function (Debreu 1959, pp. 54–9). Continuity is a technical condition that we shall not discuss.

Program A	200 people saved
Program B	600 people saved with probability 1/3 No one saved with probability 2/3

Figure 4.1.1

The theory of rationality in circumstances of certainty can thus be restated: Agents are rational if and only if their preferences may be represented by ordinal utility functions and their choices maximize utility. We intentionally avoided saying that they act "in order to maximize utility," for in contemporary economic theory, utility is merely an indicator. *Maximizing utility is just doing what one most prefers to do.* Although the utility language was inherited from the utilitarians – some of whom thought of utility as a sensation with a certain intensity, duration, purity, or "propinquity" (Bentham 1789, ch. 4) – there is no such implication in contemporary theory. To speak of individuals as "aiming to maximize" utility or as "seeking more" utility may misleadingly suggest that utility is an object of choice, some good thing that people want in addition to healthy children, lower taxes, or kiwi fruit. But the theory of rational choice says nothing about *what* people want.

Because the theory of rational choice says nothing about what people want, it does not imply self-interest. Someone who ranks the well-being of others very highly is no less a utility maximizer than is the individual who is indifferent to the welfare of others. When economic theories treat individuals as self-interested, as they often do, they must add *substantive* claims concerning what people want to the standard theory of rationality.

Utility theory lays down formal conditions that choices and preferences ought to satisfy. It is not a positive theory because it says nothing about the extent to which people are rational, and it is not merely a model or definition because rationality is itself a normative notion. To define what rational preference and choice are is ipso facto to say how one ought rationally to prefer and to choose. For example, the psychologists Amos Tversky and Daniel Kahneman (1981) presented the following hypothetical story to two groups of experimental subjects: "Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed." The first group, consisting of 152 subjects, was told that the exact scientific estimates of the consequences of alternative responses were as shown in Figure 4.1.1, and 72% preferred program A. The second group, consisting

Program C	400 people will die
Program D	No one will die with probability 1/3 600 people will die with probability 2/3

Figure 4.1.2

of 155 subjects, was instead given the information shown in Figure 4.1.2, and 78% preferred program D.

But the choices differ only in how they are described, so it is irrational to prefer A to B and at the same time prefer D to C. In saying this we are not denying that many people have these preferences, for that's what the experiment shows. Nor are we simply stipulating that the term "rational" should not be used for preferences like these. In saying that these preferences are irrational, we are *criticizing* them. Tversky and Kahneman show that decision making is being influenced by factors that ought not to have an influence.

Since utility theory says nothing about what individuals prefer, it has a much wider scope than economic theories, which depict agents as seeking more commodities or larger net returns. Utility theory is employed by psychologists, statisticians, philosophers, and sociologists as well as by economists. To hold that choices are rational if determined by preferences and that preferences are rational if they are complete and transitive does not seem to demand much. It apparently makes it easy to choose rationally and to have rational preferences. Yet many theorists maintain that it is still too demanding, and there are a variety of weaker concepts of rationality that we cannot discuss here (see McClennen 1990, ch. 2; Sen 1997, ch. 4).

One view, which is increasingly influential, holds that rationality is *bounded* (Simon 1982, 1997; Rubinstein 1997; Gigerenzer 2000). People need to make decisions even when they lack information and are unable to formulate estimates of the probabilities of various outcomes. Not only is information limited, but so are time and the ability to reason. Maximization in such circumstances may not be feasible, and the attempt to determine which among the feasible options one most prefers might be self-defeating. Simon argues that people do – and should – employ a variety of shortcuts. Among other things, they may set a certain "aspiration level" and accept the first alternative that meets their aspirations rather than seeking the "best" solution.

4.2 Expected Utility Theory

In a good deal of traditional economics, theorists abstract from the problems that arise when agents are not certain about the consequences of their actions. Frank Knight (1921) suggested that one speak of circumstances of *risk* when all the alternative outcomes and their probabilities are known. Someone playing roulette is in a circumstance involving risk. In circumstances of uncertainty, on the other hand, some of the probabilities and perhaps even some of the possible outcomes are unknown. Building a nuclear power plant involves uncertainty, because no one knows all the possible outcomes or their probabilities. In speaking of known or unknown probabilities, we are talking about so-called objective probabilities, such as relative frequencies. For example, the probability of a coin landing heads can be regarded as the limit of the ratio of heads to total flips as the number of flips grows ever larger. As we shall see shortly, there is also a subjective notion of probability as "degree of belief" that is important in the theory of decision.

In situations of risk and uncertainty – when actions do not lead with certainty to any particular outcome – actions may be regarded as *lotteries* with outcomes as prizes. Actions undertaken in circumstances of certainty can also be regarded as lotteries, though of a particularly boring sort, since they always pay off the same prize. To extend the theory of rational choice to circumstances involving risk and uncertainty, standard normative theories assert that preferences among lotteries are complete, transitive, and continuous. As noted before, completeness is not very plausible as a condition of rationality in circumstances of uncertainty. Second, economists postulate some technical conditions relating complex lotteries, which have lotteries as prizes, to simple lotteries.

Finally, one needs one additional requirement of rationality, which is known in the literature as the "independence condition" or the "sure-thing principle." Suppose that Q, like most people, prefers (other things being equal) more money to less and that Q has a choice between two gambles concerning the flip of a coin. Gamble 1 pays off $10 if the coin comes up heads and nothing if the coin comes up tails. Gamble 2 pays off $9 if the coin comes up heads and nothing if it comes up tails. It seems to be a condition of rationality that Q prefer the first gamble with the higher prize, though there is no violation of completeness, transitivity, or continuity if Q prefers the gamble with the lower prize. The further condition of rationality that is violated by a preference for the second gamble is called the "independence" or "sure-thing" principle. It says that if two lotteries differ only in one prize, Q's preferences between the lotteries should match her preferences between the prizes. It is called the sure-thing principle because Q is at least no worse-off no matter what the outcome of the lottery, and with one of the outcomes Q is better-off.

Axioms such as completeness, continuity, and independence are more controversial in the case of uncertainty than in the case of risk. Indeed one might argue that talk of probabilities and of lotteries is obscure when probabilities and even the range of possible outcomes are not known. So-called Bayesians are more comfortable applying these axioms in circumstances of uncertainty than are non-Bayesians. For Bayesians believe that the probabilities that are relevant to decision making are *subjective* degrees of belief. People may have degrees of belief about events for which relative frequencies cannot be defined, such as the outcome of a particular World Cup soccer competition. A particular World Cup is played only once, and the probability that a particular team will win is not the limit of the frequency of its victories in repeated World Cups. Bayesians interpret a claim such as "The odds that Brazil will win again are 3 to 1" as expressing instead a subjective degree of belief. An argument resembling the money-pump argument shows that degrees of belief should obey the axioms of the mathematical theory of probability. Whether one should apply expected utility theory in conditions of uncertainty remains controversial, however.

Indeed, the characterization of actions as lotteries is itself controversial, because it is arguable that what matters when an agent is choosing among alternative actions are the probabilities that the actions will *cause* various outcomes and not simply the conditional probabilities of the outcomes given the actions. Robert Nozick (1969) provided a memorable hypothetical problem ("Newcomb's problem") to illustrate what is at issue here. Suppose there are two boxes on a table in front of you. One is transparent and contains $1,000. The other is opaque. You can either take the opaque box and leave the transparent box, or you can take both boxes. The opaque box contains $1,000,000 if and only if an almost infallible predictor predicted that you would choose only the opaque box. The monetary expectation of choosing one box is thus close to $1,000,000, while the monetary expectation of choosing both boxes is close to $1,000. If one thinks of the actions as two lotteries, it seems as if choosing one box is the better bet. Yet whether the $1,000,000 is in the opaque box is already determined; it does not depend causally on whether you choose one or two boxes. Whether the million dollars is in the opaque box or not, you always come away with $1,000 more by choosing both boxes. Complications such as these have led some decision theorists to argue for what they call "causal decision theory," wherein our decisions among actions depend on a consideration of their causal consequences rather than simply on conditional probabilities (for further discussion see Gibbard and Harper 1978; Skyrms 1980; Lewis 1981; Eells 1982). We cannot go into this controversy further, and we shall have nothing more to say about causal decision theory.

If Q's preferences satisfy all the axioms, then they can be represented by a *cardinal utility* function. In calling it a "cardinal" utility function, economists mean that the magnitudes of the utility numbers, rather than just their order, are significant. The (cardinal) representation theorem says that if an agent's preferences are complete, transitive, and continuous and if they also satisfy the independence condition (and other technical conditions concerning compound lotteries), then those preferences may be represented by a utility function that has two special properties (see Ramsey 1926; von Neumann and Morgenstern 1947; Herstein and Milnor 1953; Savage 1972; Harsanyi 1977b, ch. 3).

First, this utility function possesses the *expected utility* property: the (expected) utility of any lottery is equal to the utilities of its outcomes multiplied by their probabilities. Suppose Q is offered a gamble over the roll of a fair die. Q will win \$60 if the die comes up 1 or will lose \$12 if the die comes up with any other number showing. The *expected monetary value* of the gamble is 1/6 times \$60 minus 5/6 times \$12, or \$0. The *expected utility* of the gamble to Q is 1/6 times the utility to Q of winning \$60 plus 5/6 times the utility to Q of losing \$12. Depending on how rapidly Q's utility increases as a function of money, the expected utility of the gamble may be positive, negative, or zero. If Q's preferences satisfy the axioms, then there is a way of representing Q's preferences so that the expected utility of a lottery is the sum of the utilities of the prizes weighted by the probabilities of winning them. One sense in which expected utilities are "cardinal," rather than merely ordinal, is that the precise utilities of gambles can be calculated from probabilities and utilities of the outcomes.

Second, the cardinal representation theorem states that if Q's preferences satisfy the axioms then an expected utility function (a utility function with the expected utility property) that represents them is "unique up to a positive affine or linear transformation." To say that a function U is "unique up to a positive affine or linear transformation" is to say that any $U' = aU + b$, where a is a positive real number and b is any real number, will do just as well as U itself. If a function U represents an agent's preferences and has the expected utility property then so will all positive affine transformations of U. Conversely, no expected utility function that is not a positive affine transformation of U will represent the agent's preferences, though there will be other merely ordinal utility functions (without the expected utility property) that will also represent Q's preferences.

To understand what this means, it helps to compare the measurement of expected utility to the measurement of temperature. Alternative temperature scales, such as the Fahrenheit and the Celsius scales, are positive affine transformations of one another ($T_F = (9/5)T_C + 32$.) Because of this, the decision to measure temperature on a Fahrenheit as opposed to a Celsius scale is arbitrary. A claim such as "temperature differences between night and day are larger in deserts than in wet climates" is true regardless of which scale one uses, provided that the scales are positive affine transformations of one another.

For the case of ordinal utilities, in contrast, differences between utilities are entirely arbitrary, since any numbers in the same order represent the preferences just as well. The fact that the difference between the ordinal utilities $U(x)$ and $U(y)$ is larger than the difference between $U(c)$ and $U(d)$ tells one *nothing* about the agent's preferences. The relative magnitudes of the numbers are just an accident of the particular utility function chosen. But things are different with cardinal utilities. The cardinal representation theorem establishes that comparisons of utility *differences* (like comparisons of temperature differences) are not arbitrary; they do not depend on what scale is chosen. If $U(x) - U(y) > U(c) - U(d)$ and if U' is a positive affine transformation of U, then $U'(x) - U'(y) > U'(c) - U'(d)$.†

Although expected utilities are cardinal and thus comparisons of utility differences are in this way "objective," nothing said so far permits one to make comparisons between the utilities of different individuals. If Q's cardinal utility is 100 and M's is 89, one cannot draw any conclusion about whose preferences are better satisfied. Since any positive linear transformation of these numbers will represent the preferences of the individuals just as well, their relative magnitude reflects only an arbitrary choice of one particular utility function for each.

As in the case of utility theory concerning choice under certainty, economists relate choice to preference by asserting that individuals choose whatever available option is highest in their preference ranking – or, in the case of ties, one of the options tied for first place. Unlike ordinal utility theory, which is a theory of the rationality of preference and choice only, expected

† *Proof:* We have $U'(x) - U'(y) = aU(x) + b - aU(y) - b = a(U(x) - U(y))$. Similarly, $U'(c) - U'(d) = a(U(c) - U(d))$. So $U'(x) - U'(y) > U'(c) - U'(d)$ if and only if $a(U(x) - U(y)) > a(U(c) - U(d))$. And since a is positive, this will be true if and only if $U(x) - U(y) > U(c) - U(d)$.

utility theory has implications concerning the rationality of *belief*. If the axioms of expected utility theory are conditions of rationality, then people whose degrees of belief concerning uncertain outcomes do not satisfy the axioms of the probability calculus are irrational. Expected utility theory is thus not only a theory of rational preference and choice; it is also a partial theory of rational belief.

4.3 Questions about Utility Theory

Expected utility theory demands more than does utility theory under certainty, and it is much more controversial. Psychologists and economists have found ways of testing whether people's preferences satisfy the conditions of expected utility theory, and they have found that preferences often appear not to. If people don't behave the way that expected utility theory says they ought to, then it may be the case that people are irrational. People may break the rules, even if they are good rules. But if choice behavior that violates the axioms of expected utility theory is common and seems sensible, then one may question whether expected utility theory is a correct theory of rationality. Some empirical results do suggest that expected utility theory is flawed as a theory of rationality. These empirical anomalies, along with general critiques of expected utility (Allais 1952; Ellsberg 1961; Sen 1985b; Levi 1986; McClennen 1990), have stimulated the formulation of alternative theories of rationality under uncertainty.

Testing a normative theory of rationality is less straightforward than testing a positive theory of actual choice, and controversy can easily arise and persist. We cannot survey the controversies concerning expected utility theory and its competitors here, but we will give one famous instance that is of some importance to moral philosophy as well as to decision theory. This example provides some sense of the character of these controversies.

In the early 1950s, Maurice Allais formulated the problem shown in Figure 4.3.1 (Allais and Hagen 1979). A ball is drawn from an urn containing 1 red ball, 89 white balls, and 10 blue balls. So the probabilities are known. Depending on the color and the choice of A or B in problem I or of C or D in problem II, the player receives one of the payoffs listed in the figure.

For example, if our friend Q faces problem I and chooses A, then she gets $1 million for sure. If Q chooses B then she has a 1% chance of getting nothing (if a red ball is drawn), an 89% chance of winning $1 million, and a 10% chance of winning $5 million. Many people are inclined in problem I to prefer option A to option B and in problem II to prefer option D to option C. Even the prominent Bayesian statistician Leonard Savage at first

Payoffs

Problems	*Choices*	Red (1)	White (89)	Blue (10)
I	A	\$1 million	\$1 million	\$1 million
	B	\$0	\$1 million	\$5 million
II	C	\$1 million	\$0	\$1 million
	D	\$0	\$0	\$5 million

Figure 4.3.1. Allais's Problem

expressed these preferences (1972, p. 103). But these preferences violate the independence principle, for the only difference between the pair (A, B) and the pair (C, D) is in the magnitude of the payoff if a white ball is drawn, which should be irrelevant.‡

Yet many people still prefer A in problem I and D in problem II. Intuitively the reason is that in choosing B one is giving up a million dollars for sure,

‡ Here's why: Let L be the lottery that pays off \$5 million with probability 10/11 and nothing with probability 1/11. One can then recast Figure 4.3.1 as follows.

Payoffs

Problems	*Choices*	White (89)	Red or Blue (11)
I	A	\$1 million	\$1 million
	B	\$1 million	L
II	C	\$0	\$1 million
	D	\$0	L

The independence condition says that A is preferred to B if and only if \$1 million is preferred to L, and it says that C is preferred to D if and only if \$1 million is preferred to L. So A is preferred to B if and only if C is preferred to D. For those who prefer some algebra, let V be the utility of \$5 million, U the utility of \$1 million, and 0 the utility of \$0. Then $\mathrm{EU(A)} = U$, $\mathrm{EU(B)} = 0.89U + 0.1V$, $\mathrm{EU(C)} = 0.11U$, and $\mathrm{EU(D)} = 0.1V$, where EU denotes expected utility. If A is preferred to B, then $\mathrm{EU(A)} > \mathrm{EU(B)}$. Hence $U > 0.89U + 0.1V$, or $0.11U > 0.1V$. So if A is preferred to B, then C must be preferred to D.

while in choosing D one is not. This reason is in conflict with the independence condition, which says that the prize if a white ball is chosen is irrelevant. It is possible to reconcile a preference for A and D with the independence condition by arguing that the outcomes have not been properly described. The outcome of choosing B if a red ball is chosen is not \$0, but \$0 plus severe regrets. ("I could kill myself. I gave up a million dollars for sure!") Although some decision theorists (see Broome 1991b, ch. 5) have defended the independence condition in this way, others have worried that such redefining of the outcomes makes expected utility theory empty, since apparent disconfirmations can always be explained away by redescribing the outcomes. For example, in Tversky and Kahneman's Asian disease example, one could say that there is no irrationality in people's choices, because 200 out of 600 people living is a different outcome than 400 people out of 600 dying. Broome's response is to argue for substantive principles of rationality that specify whether it is rationally permissible to distinguish between the outcomes when a red ball is drawn when one chooses B (and thus might have had \$1 million for sure) and when a red ball is drawn when one has chosen D. (Such principles would deny that it is rationally permissible to distinguish 200 out of 600 dying from 400 out of 600 surviving.) Most decision theorists and economists do not follow Broome here, and they regard preferences for A in problem I and at the same time for D in problem II as being inconsistent with the independence condition.

If one takes the choices of A in problem I and D in problem II to be inconsistent with the independence condition, then are these choices evidence against the independence condition or rather evidence of human irrationality? It is hard to see how to make conclusive arguments in favor of one of these alternatives, but we believe that either the payoffs need to be redescribed or there is a case to be made against the independence condition. Observe that (1) there is no other evidence of irrationality among those preferring A and D, (2) some distinguished decision theorists themselves choose A and D, and (3) plausible alternative theories of rationality have been defended that rationalize choosing A and D. Together, (1)–(3) provide reason to question whether the independence condition is a requirement of rationality.

Apparent anomalies such as Allais's example are relevant to moral philosophy as well as to economics, because expected utility theory is involved in arguments for moral conclusions – such as Fleming's (1952), Harsanyi's (1955), and Vickrey's (1945) purported proofs of utilitarianism. As we will discuss in Chapter 7, utilitarianism is the view that actions and policies are morally permissible if and only if they lead to no less total "utility" than any alternative. There are also nonutilitarian ethical systems that are built around expected utility theory, such as David Gauthier's (1986) bargaining theory of justice (discussed in Chapters 12 and 14).

The issues concerning expected utility theory are not merely theoretical, because alternative theories may yield different recommendations about what is the "rational" thing to do. For example, in the case of nuclear power (which obviously involves uncertainty rather than merely risk), should policy makers rely on their best guess as to the future consequences and maximize expected utility? Or should they, as Isaac Levi (1980) urges, recognize their ignorance of relevant probabilities and use some other principle of choice?

In addition to the controversies concerning whether expected utility theory is too narrow or too demanding, some moral philosophers have maintained that there are such things as irrational preferences quite apart from any questions about consistency. This is a classical view found in Aristotle and Plato. More recent defenders include Thomas Nagel (1970) and John McDowell (1978). Derek Parfit, for example, holds that it is "irrational to desire something that is in no respect worth desiring" (1984, p. 123). He gives the example of a person who suffers from "future-Tuesday indifference." This individual prefers a greater pain on a future Tuesday over a lesser pain on a Wednesday, simply because he does not care what happens on future Tuesdays. Parfit remarks: "the fact that the agony will be on a Tuesday is no reason for preferring it. Preferring the *worse* of two pains for *no* reason, is irrational" (1984, p. 124). Although the decision theory most economists accept makes no substantive claims about what preferences are rational or irrational, the economic theories they accept typically identify rationality with material self-interest and hence implicitly judge that preferences for things that are advantageous for oneself are rational and that other preferences are irrational (Walsh 1994).

John Broome argues that expected utility theory itself must rely upon judgments that some preferences are intrinsically irrational. Consider, for example, an instance of intransitivity: Q prefers x to y, y to z, and z to x. If Q wanted to make trouble for the decision theorist, she could argue that the apparent intransitivity results from misdescribing the objects of her preferences. She might say that "z when compared to y" is a different thing altogether than "z when compared to x." If one calls the first z_1 and the second z_2, then Q may have a fully transitive preference ordering such as z_2, x, y, z_1. Such a reply to the charge of having intransitive preferences can always be given, and so transitivity turns out to be an empty condition of rationality unless one is willing to argue that it is *irrational* to prefer z_2

(z when compared to x) over z_1 (z when compared to y). Thus expected utility theory has no "bite" unless one is committed to the irrationality of such preferences. David Lewis, Susan Hurley, and John Broome have argued in this way that expected utility theory presupposes substantive principles of rationality.

Although it is good to know something about the ways in which expected utility theory may be too strong or too weak, what's most important for our purposes is to understand clearly what the different varieties of utility theory have to say about rationality. In simple terms, utility theory sees rational choice as choice that is determined by rational preference and belief, and it sees rational preference as demanding only that individuals be able to rank all options. It is remarkable that such a bare-bones formal theory of rationality has given rise to so much controversy and (as we shall see) that it has such striking ethical implications.

Suggestions for Further Reading

The basic model of rationality and its extension in expected utility theory have been widely discussed. Some of the sources that we have found particularly accessible and helpful are Sen (1971), Harsanyi (1977b, ch. 3), Loomes and Sugden (1982), Machina (1987), Gärdenfors and Sahlin (1988), McClennen (1990), and Hollis and Sugden (1993).

For philosophical discussion on the plausibility of restricting questions about rationality to questions about consistency and about the rationality of means, see: Hume (1738, 1748); Nagel (1970); McDowell (1978); Parfit (1984, pt. 2); Gauthier (1986, ch. 2); Griffin (1986, chs. 3, 4); Hurley (1989, ch. 5); Broome (1991b, ch. 5); Hampton (1998); and Sen (2002).

For empirical investigations of possible irrationality, see Lichtenstein and Slovic (1971); Kahneman, Slovic, and Tversky (1982); and Camerer (2003).

FIVE

Rationality in Positive and Normative Economics

Positive economics is concerned with the explanation and prediction of economic phenomena, while normative economics is concerned with evaluating economic policies, practices, and states of affairs from a moral standpoint. Rationality is a normative notion concerning how people ought to choose, prefer, or reason, so it may seem surprising that it plays a large role in positive economics. Since rationality is different from morality, it may also seem surprising that rationality plays a large role in normative economics. But in fact rationality is ubiquitous in both positive and normative economics.

5.1 Rationality and Positive Economics

People's preferences are rational if they are complete and transitive, and people choose rationally if their choices are determined by their preferences. If one adds to this theory of rationality the generalization that consumers are to some extent rational and that they prefer more commodities to fewer, then one has the central principles of the positive theory of consumer choice. Similarly, the traditional theory of the firm maintains that firms or entrepreneurs are rational and that they combine inputs so as to maximize the difference between revenues and costs. Though the theory of consumer choice and the theory of the firm make additional claims, they both take the theory of rational choice to be the theory of actual choice.

Positive economics on both the consumer and the producer side can be formulated without using the word "rational." Rather than first defining "rational" and then stating that individuals are rational, one can assert that the preferences of individuals are complete and transitive and that individuals choose whatever affordable bundle of commodities they most prefer. On the production side, one can simply assert that firms maximize profits.

But the identification of what is actual with what is rational remains. It does not depend on any particular formulation. It is instead a reflection of the fact that economics simultaneously provides a theory of the *causes* and *consequences* of people's economic choices and a theory of the *reasons* for them.

As discussed in Chapter 4, folk psychology takes human actions to result from the beliefs and wants of agents. In explaining actions, we sometimes cite only the agent's beliefs or only the agent's wants, since the other is obvious from the context. For example, when asked why he robbed banks, Willie Sutton replied: "Because that's where the money is." His explanation makes us laugh because, unlike most of us, Sutton thinks that the desire to steal is too obvious to require any mention. The pattern of explaining action in terms of beliefs and desires is ubiquitous.

Economic explanations of choices in terms of utility maximizing fit this pattern. Economists typically talk in terms of "preferences" or "utility" rather than "wants," and they often do not mention explicitly the agent's beliefs because they assume that the agent has perfect knowledge and thus believes whatever the facts are. But these are minor complications. In the main, explanations in mainstream economics of the choice behavior of individuals conform to the pattern of folk psychology. Most economists are more interested in the market consequences of individual choices than in the choices themselves, but individual choices are the causal intermediaries that connect "shocks" such as crop failures or a new tax to their consequences for prices or quantities exchanged.

The relations between microeconomics and folk psychology are worth noticing because folk psychology is an account of the *reasons* for actions. Indeed, folk psychological explanations in terms of reasons are often contrasted to causal explanations of behavior in terms of conditioning, post-hypnotic suggestion, or brain chemistry. A number of philosophers in the 1950s argued that explanations for actions in terms of beliefs and desires were exclusively reason giving and could not be causal. If they were right, then there would be a radical difference between explanations in microeconomics, which cite reasons, and explanations in the natural sciences, which cite causes.

But in our view, which is now common among philosophers, reasons that explain actions must also cause the actions they explain. Donald Davidson (1963) supplied the crucial insight here. An example will clarify Davidson's argument. Consider a real-estate agent who attends church regularly. One reason is that the agent believes in God and wants to express devotion to God. Another reason is that the agent meets potential clients at church. Both of these are in some sense "good" reasons to go to church, though most people find the first reason more admirable than the second. Davidson points out, however, that identifying good reasons does not automatically explain actions. One needs to know which reasons actually "led" the agent to go to church. There are many possibilities: (1) It may be that both reasons explain the church going. (2) It may be that the agent attends because of piety and only incidentally profits from attending. (3) It may be that the agent is a hypocrite and attends church only as a way of attracting clients. Or (4) it might be that the "real" reason is that the agent loves to listen to the church organist. Davidson's point is that, in order to explain an action, one needs to know which reasons were *causally responsible* for it.

One special feature of reasons is that people evaluate or assess them. Piety is a *good* reason to attend church. "Good" here seems to mean both "(morally) admirable" and "sensible." The desire to make lucrative contacts is a sensible reason to go to church, though not as admirable. The desire to reverse the greenhouse effect is an admirable aim though hardly, by most people's theology, a sensible reason to attend church. The desire to resurrect Joseph Stalin is not a good reason to go to church, for it is neither sensible nor admirable.

Reason-giving explanations of human behavior invite questions about whether the behavior is justified by the reasons and whether the reasons themselves are justified. Thus, in microeconomics courses, word problems often take the form of asking students to figure out the hidden rationality behind an instance of seemingly irrational behavior, such as an airline requiring a Saturday night stay for a reduced ticket price. The explanation – that it permits airlines to charge business travelers higher fares – shows why the behavior may be profit maximizing for the firm. Such explanations may help justify a firm's behavior. If the airline has a good reason to charge those who stay Saturday night less, then it has the beginnings of a response to the accusation that charging different prices to different customers is unfair.

Every theory that takes beliefs and preferences to be reasons that explain choices must incorporate some theory of rationality. Since desires cannot function as explanatory reasons if they do not induce at least a loose ranking of what is better or worse and thereby influence choices, it follows that preferences must not be radically incomplete and that they must have a large measure of consistency. To say, as economists often do, that rationality requires that preferences be complete and transitive and that choice maximizes preference satisfaction is one simple way of meeting these constraints on reason-giving explanations. The standard economic theory of actual choice must also be a theory of rational choice if explanations in economics are to cite both causes and reasons. To say this, however, is not to offer a defense of the standard theory of rationality. There are alternatives, such as Herbert Simon's theory of bounded rationality, which also permit one to explain choices in terms of reasons.

Although some philosophers and psychologists have argued against folk psychology, it is hard to imagine humans abandoning this way of understanding their actions. That the theory of actual choice is simultaneously a theory of rational choice thus helps to shore up the theory of actual choice. Suppose one finds evidence that people's preferences are not complete and transitive. Psychologists claim to have found such evidence. For example, with a bit of experimental trickiness, one can induce individuals to judge that a bet or investment *J* is definitely better than another bet or investment *K* yet to express a willingness to pay *more* for the inferior bet, *K* (Lichtenstein and Slovic 1971, 1973). This "preference reversal" phenomenon is well established in experimental settings, and it follows that people do not have complete and transitive preference rankings. Such solid evidence disconfirming central theoretical propositions of positive economics ought to be worrisome.

But most economists are not worried. Why not? One reason is that choice phenomena such as preference reversals are irrational and so can be exploited. Consequently, irrational preferences tend to be unstable. In one study, Chu and Chu (1990) sold *K*-bets to experimental subjects for the price they stated they were willing to pay, had them exchange the *K*-bets for the *J*-bets (which the subjects claimed to prefer), and then purchased back the *J*-bets for the lower price that subjects claimed the *J*-bets to be worth. At the end of the series of transactions the subjects were, of course, poorer. If the "money pumping" cycle is this transparent (see Berg, Dickhaut, and O'Brien 1985), then it does not take long for subjects to figure out that they are being exploited and to adjust their stated preferences or the prices they are willing to pay for the bets. The exploitation of such irrationalities will tend to limit departures from rationality.

A defender of the standard theory might then go on to claim that, although not perfectly correct, the standard theory of choice is a good first approximation, because a theory that portrayed choice as irrational would typically reveal opportunities for exploiting this irrationality. Since people will learn not to be exploited, acting on such a theory would undermine it. Similarly, profit-maximizing firms will tend to drive those that behave otherwise out of business, and corporate firms that use their assets inefficiently will be targets for takeover. By a process akin to natural selection, surviving firms will be profit maximizers.

We have serious reservations about these arguments, but it nevertheless matters to the appraisal of mainstream economic theory that the positive theory of choice is simultaneously a theory of rational choice and thereby serves to evaluate even as it predicts and explains agents' conduct. The intermixture of positive economics and the theory of rationality makes positive economics more acceptable to many economists and harder to disconfirm.

5.2 Preference Satisfaction and Pareto Efficiency

Rationality is a normative notion. One *ought* to be rational; one is foolish or mistaken if one is not rational. But it might reasonably be contended that rationality is not a moral notion. One can be a rational villain. What one ought rationally to do need not coincide with what one ought morally to do.

The standard "thin" theory of rationality as utility maximization is not itself a moral theory. Some of those who are wicked, like some of those who are just, might be rational and others irrational. But the standard theory of rationality does not stand by itself. On the contrary, it is embedded in both normative and positive economics and thereby smuggles crucial elements of moral theory into positive economics. Before we see how, let us examine what part rationality plays in normative economics.

Positive economics and the theory of rationality jointly determine the character of normative economics. Moreover, one can virtually derive mainstream normative economics from the theory of rationality and components of *positive* economics! Here's how.

Start with the theory of rationality and add a common assumption of positive economics: that individuals are exclusively self-interested. If nothing but self-interest affects *S*'s preferences, then *S* prefers *x* to *y* if and only if *S* believes that *x* is strictly better for *S* than is *y*. Rational and exclusively self-interested individuals always prefer what they believe to be better for themselves over what they believe to be worse.

Add then a second common assumption of positive economics: that individuals have perfect knowledge. Self-interested individuals with perfect knowledge prefer *x* to *y* if and only if *x* is *in fact* better for them. Hence one can identify how well-off an individual is with how well satisfied an individual's preferences are. Orthodox normative economics consequently identifies welfare with preference satisfaction. Economists frequently make this identification in a flash without noticing that they are defending a controversial philosophical theory. We will discuss the importance of this identification later in this chapter and in Chapter 8.

Once well-being is identified with the satisfaction of preferences, the central features of standard normative economics follow naturally. All they need is one innocuous moral principle of *minimal benevolence*: other things being equal, it is a morally good thing if people are better-off. Indeed, one might argue that this principle follows from the near tautology: other things being equal, it is a morally good thing if there is more good. Because of the "other things being equal" clause, minimal benevolence is uncontroversial. It does not, for example, say whether it is better or worse to make some people better-off at the cost of increasing inequality. It only says that if all other morally relevant considerations (such as equality) are a toss-up, then it is a morally good thing to make people better-off. One premise in Chapter 2's reconstruction of the argument in Lawrence Summers's memorandum came to grief precisely because it did not contain such a *ceteris paribus* clause. Even if shifting dirty industries to LDCs makes everyone better-off, it might be morally unacceptable because unjust. This observation is fully consistent with minimal benevolence.

Those who accept minimal benevolence and identify an individual's welfare with the satisfaction of that person's preferences will judge that, other things being equal, it is a morally good thing to satisfy an individual's preferences. The main issue in standard normative economics is accordingly to what extent economies enable individuals to satisfy their preferences. From this follows the importance of "Pareto optimality."

A Pareto optimum (also called a "Pareto efficient" allocation) is typically defined as a state of affairs in which it is impossible to make anyone better-off without making someone worse-off, but this purported definition is misleading. It is more accurate to say that *R* is a "Pareto improvement" over *S* if nobody prefers *S* to *R* and somebody prefers *R* to *S*; then *R* is a Pareto optimum if and only if there are no Pareto improvements over *R*. Since economists identify preference satisfaction with well-being, they often describe a Pareto improvement as a change that makes some people better-off without making anyone worse-off. There will typically be many Pareto optima, and many of these will be ethically unattractive. Consider a state of affairs *R* in which millions of people are starving. If there is no way to make them better-off without making some affluent person worse-off, no matter how slightly, then *R* is a Pareto optimum.

If one accepts minimal benevolence then, *other things being equal*, Pareto improvements are moral improvements, and states of affairs that are not Pareto efficient are morally undesirable. At this point, some formal theorems of welfare economics become very important. What is usually called "the first welfare theorem" establishes that the allocation resulting from any perfectly competitive equilibrium among self-interested agents is Pareto optimal. Perfect competition obtains when there are (1) no interdependencies among people's utility functions, (2) markets for all goods and services (and thus no externalities), (3) no barriers to entry or exit from any market, and (4) so many traders in every market that no one trader can influence prices. A perfectly competitive equilibrium obtains when there is perfect competition and there are no excess demands on any market. Given this first welfare theorem, one can conclude that, *ceteris paribus,* perfectly competitive equilibria are morally desirable and market imperfections that interfere with the achievement of competitive equilibria are morally undesirable.

This is a theoretical defense of perfect competition, not a defense of actual markets or of a laissez-faire policy, and it is only a defense of perfect competition *other things being equal*. Given this vague and extensive qualification, it is best to express the virtues of Pareto optimality negatively. A Pareto optimal state of affairs, and thus (according to the first welfare theorem) a perfectly competitive equilibrium, is not subject to one sort of moral criticism – that is, it is not possible to make things better without going against somebody's preferences. This negative formulation captures the full – and quite limited – merits of Pareto optimality.

A consideration of the "other things" that are morally relevant to the appraisal of economic outcomes, institutions, and processes leads to ethical controversy, which many economists would prefer to avoid. Indeed, we conjecture that mainstream economists rarely argue for markets on grounds of individual liberties and rights mainly because they believe that arguments based on such premises are more philosophically controversial and ambitious than the benevolence argument.

Justice is one of the "other things" that may or may not be "equal," and a Pareto improvement that leads to distributional injustice may be morally undesirable. But the argument may be continued, for there is a second welfare theorem. This theorem shows that all Pareto optima can be obtained as competitive equilibria from the "right" initial distribution of endowments. In other words, one can achieve any desired Pareto optimal outcome by first redistributing the resources that individuals bring to their market interactions and then letting individuals trade freely under the highly idealized conditions of perfect competition. One can employ this second welfare theorem to argue that all other moral concerns, including concerns about justice, can be satisfied by adjusting initial holdings and that perfect competition is thus a moral ideal. The conclusion does not follow, however, because it could be the case that one will have to settle for an outcome that is not Pareto optimal in order to satisfy the other moral constraints. This would be the case, for example, if one held on moral grounds that certain goods or services should not be marketed – for example, body parts. The practical force of the second welfare theorem is further limited by the inevitable efficiency-impairing distortions and the grave political difficulty of redistributing initial endowments.

The welfare theorems do not help with the problem of comparing states of affairs that improve some people's levels of preference satisfaction while harming others. This severely limits the range of social policies on which the Pareto notions can find purchase. Normative economics will not be very helpful unless there are other ways of comparing alternative policies. In Chapter 9 we shall see how standard welfare economics has found ways to say more.

Whether they are defenders of laissez-faire or of extensive government intervention to address market failures, most economists share a moral commitment to the ideal of perfect competition. It is this commitment that gives point to the analysis of market failures. (Why should they matter if market successes are not a good thing?) The fact that this commitment appears to presuppose nothing more controversial than minimal benevolence helps resolve a paradox: On the one hand, most economists do not see themselves as moral philosophers, and they attempt to steer clear of controversial ethical commitments when doing theoretical welfare economics. Indeed, some economists have supposed that theoretical welfare economics was independent of all value judgments. Yet when welfare economists address policy questions, they speak with apparent moral authority. They purport to know how to make life better.

What explains how economists with these views can feel themselves possessed of moral authority without the trouble of doing moral philosophy? The answer, we think, is that welfare economists do not usually regard the identification of well-being with the satisfaction of preferences as a controversial ethical judgment; it seems to be just a part of the standard view of rationality. Once one accepts this identification, one need only add the uncontroversial principle of minimal benevolence to derive the argument for perfect competition sketched here. As we discuss in Chapter 8, the identification of well-being with preference satisfaction is actually quite controversial.

5.3 Rationality and Ethics in Positive Economics

As you sort through the complicated story told in the previous sections, you can begin to see how rationality can function as a Trojan horse smuggling ethical commitments into the theoretical citadel of positive mainstream economics. At a relatively superficial level, commitment to a theory of well-being as the satisfaction of preferences places constraints on the character of positive economics. It is only plausible to identify well-being and the satisfaction of preferences if (1) individuals are rational, (2) individuals are self-interested, (3) individuals are well-informed, and (4) individual preferences are not formed or deformed in odd ways. If, for example, people come to prefer things because of manipulation by advertisers or because of "sour grapes," then it is harder to believe that satisfying their preferences automatically makes them better-off. Hence one reason, among many others, to avoid incorporating theories of preference formation into positive economics is that doing so could threaten the accepted (moral) theory of individual well-being and thus possibly upset the conclusions of welfare economics. The fact that there has been little work on preference formation helps to make plausible the identification of well-being and preference satisfaction, and commitment to this view of well-being discourages research on preference formation.

But the standard theory of rationality – together with standard assumptions about the objectives of consumers and firms – introduces evaluative commitments at a deeper level. A telling illustration of this tendency is the economic theory of international trade and the attachment so many economists feel to the doctrine of free trade. The core idea behind the classical and neoclassical theory of international trade derives from the model of production and trade between two rational individuals. Obviously, if Jill is better than Jack at hunting and Jack is a better fisher than Jill, it will make sense for Jill to hunt, for Jack to fish, and for the two to trade (assuming they both enjoy both meat and fish). Indeed, it is easy to show that trade results in a Pareto improvement over a no-trade equilibrium. The key insight that David Ricardo saw clearly is that trade will be valuable even if Jill is better than Jack at *both* hunting and fishing, provided only that the two differ in their relative skill in the two activities. If Jill hunts three times as well as Jack and fishes twice as well, then it will still make sense for Jill to hunt, for Jack to fish, and for the two to trade. Jack has a *comparative* advantage in fishing and Jill in hunting, and in that circumstance trade can produce a Pareto improvement relative to a no-trade situation. The assumptions of rationality and minimal benevolence are thus sufficient to warrant advocating trade between the two. Yet even in the case of individuals this argument only works other things being equal. For example, large differences in resource endowments or in power between the two traders may result in an equilibrium that, while Pareto optimal, is highly unjust.

In its simplest form, international trade theory reproduces this analysis of individual exchange by substituting two countries for the two individuals in the example. So, in comparing Japan to the United States, the former has a comparative advantage in fishing and the latter in raising cattle for beef. It makes sense (mad cow disease aside) for Japan to concentrate on fishing and the United States on cattle raising and for the two to trade. This simple argument, with a variety of complications added, has been overwhelmingly compelling to most economists for two hundred years, and economists as a group have consequently been resolute opponents of tariffs and other trade barriers except in very narrowly defined circumstances. Many details change, of course, in the move from individual persons to countries, and some of these details pose significant theoretical and empirical challenges. Yet in the context of "positive" economics, the explanatory power of the simple idea of comparative advantage, transposed from individuals to countries, is enormous.

However, the move from individual traders to countries is of great moral significance. Initiating trade between Jack and Jill is a Pareto improvement, but changing the rules of trade will virtually never be Pareto improving. Although facilitating trade in beef and fish between the United States and Japan may make average incomes in both countries rise, it is quite likely to harm Japanese cattle raisers and U.S. fishers. Many economists would argue that there are more efficient ways to address the costs to the losers than to impose tariffs, but that case cannot be made on the basis of the second welfare theorem. One needs to address the political realities. Much stronger normative assumptions than rationality and minimal benevolence would be required to justify the simple free-trade argument in these circumstances. Whether tariffs are a good or a bad idea cannot be decided by rationality, minimal benevolence, and the simple analytics of comparative advantage.

There is no reason in principle why this slide from explanation to justification need occur – welfare economists determined to limit their judgments to those of Paretian welfare economics could simply be silent on questions of whether free trade is, all things considered, desirable. We would suggest that there are at least two quite different kinds of reasons why mainstream economists have typically not followed this path. One is a powerful (and often justified) sense that those who oppose free trade simply "don't get it": they either don't understand the economist's "positive" argument about how trade works, or they deliberately appeal to misunderstandings in order to promote the interests of a particular industry. "If people only understood some economics, they would see how stupid and harmful tariffs are."

We can mention here two obvious misunderstandings that economists struggle to combat. First is the argument that it is impossible for industries in poor countries to compete with much more productive industries in rich countries. This argument may simply confuse absolute and comparative advantage – with better skills and more equipment, U.S. workers may indeed be absolutely better at producing many things than workers in poor countries, but the whole point of trade is to import those things that you produce *comparatively* less well. Second, critics of free trade sometimes talk as if exporting is all benefit and importing is all cost, so nations should block imports and encourage exports (a view often labeled "mercantilism"). But obviously nations can't export if they don't (eventually) import. Moreover, their residents can consume the imports, which is a presumed benefit. Seeing these mistakes tempts economists to believe that if only their critics understood the analysis, they would agree with the economists' conclusions.

A second reason, in our view, why economists so frequently advocate free trade – even though it cannot be defended as a Pareto improvement – is that they hold further moral beliefs which support their advocacy but which they often have not developed explicitly. Thus, for example, many economists may plausibly believe that tariff barriers imposed by rich countries on poor countries wind up causing harm to people in poor countries that outweighs (in moral terms) the gains to the protected workers in rich countries. Or economists may believe that the political voice of workers in industries threatened by imports is greater than that of more diffuse sets of workers who will benefit from more exports, and this gives the first group an unfair advantage in shaping trade policy. These are both reasonable arguments, at least in some circumstances, and in our view the advocates of free trade would do well to voice them more thoroughly and openly.

Note though that these arguments, even if adequate to defeat "mercantilist" views, do not suffice to make a blanket argument for free trade. We can further illustrate this point in terms of the lively current controversy about apparel manufacturing. Makers of well-known brands are often accused of exploiting workers in poor countries by operating "sweatshops" that pay low wages, impose unreasonable hours and working conditions, and use child labor. On the explanatory side, one may ask: Why do firms treat U.S. workers so much better than those abroad? An explanation would be sought in the very different labor supply conditions in different places and in the theory of comparative advantage. Workers are paid much less in poor countries because, owing (among other things) to having less education and less capital to work with, they are much less productive per hour of work. Yet it pays companies to employ them – to have the work done by these relatively unproductive workers – because of comparative advantage: in relative terms, they are better at doing this than at doing other things American workers do. Economists, then, may be inclined to argue that these workers in poor countries are better-off than they would be if American companies were not allowed to buy their labor, and economists may see this point as an argument for free trade. Citing this explanation, manufacturers might defend their practices on the ground that improving working conditions or wages in their plants would make the plants unprofitable. Thus – or so goes the justification – the net effect of enforcing requirements for better wages and working conditions on these firms would be their withdrawal and hence the unemployment of the firms' workers, leaving them still worse-off. Here an argument for the economic rationality of a firm's labor practices turns into a justification for its conduct.

It is likely, however, that this argument goes too far. The labor practices that maximize the profits of a firm are not ipso facto justified. Granting that better wages and working conditions would reduce the firm's profits, this would only result in the firm's withdrawal from this market if (a) conditions were so competitive that there were no excess profits or if (b) the firm had an inflexible policy to allocate its resources across countries so as to maximize profits. (A further argument might be made that the capital market would force a takeover of the firm if it didn't follow the policy in (b).) A defender of free trade may well be correct in arguing that the sweatshop workers would be harmed by the withdrawal of American companies (and perhaps correct as well in suspecting that American unions who criticize these labor practices are more concerned with American jobs than with the fate of workers in poor countries). But the claim that any restriction – beyond those imposed by the market – on the firm's labor practices would necessarily result in the firm's withdrawal is a strong empirical claim, one not supported by a general argument that international regimes with trade are better than those without.

Indeed, it is plausible that firms could agree to (or have imposed on them) restrictions on working hours, child labor, and the like without making their operations unprofitable – especially if all firms observe similar restraints (since that would relieve some of the competitive pressures they would otherwise face). The search for reasonable restrictions of this kind can be helped by more explicit attention to the values at stake as well as to the empirical consequences that would follow from such restrictions.

We are not criticizing those economists who believe that their "positive" analyses carry normative implications. Improving economic policy and practice is an excellent reason for working on economic problems, and such improvement is inherently a normative endeavor. Our concern is that economists should be self-aware and self-critical when they take up problems where the normative and positive elements are entangled (to borrow a phrase from Putnam 2002).

5.4 Self-Interest and Moral Motivation

Explaining and predicting human actions rely on generalizations about what motivates people. Suppose one is studying tax compliance. People have been known to cheat on their taxes. The extent of cheating varies across countries and among groups of taxpayers. In studying this cheating, many economic analyses focus on the factors that influence the gains from cheating (marginal tax rates, most obviously) and the costs (probability of detection and probability of legal penalty if detected). These cost–benefit considerations provide reasons of a certain kind for or against cheating on taxes. But are they the only important reasons? Not everyone will cheat just because they expect it to be profitable. Some people think it is wrong to cheat. Others may cheat even when there is no expected benefit, believing (say) that the government will put their taxes to immoral uses.

Economists who analyze the self-interested reasons for cheating and who explore their implications while ignoring or dismissing "moral" reasons may, perhaps unintentionally, wind up justifying the practice. If there is sufficiently large expected return, then cheating is rational from the point of view of an individual concerned only with personal net income. From the perspective of such a materially self-interested individual, the only thing wrong with cheating is the risk of getting caught. On the assumption that everyone is materially self-interested, those found evading their taxes are either incautious or unlucky.

One might put the point in this way. The search for an explanation typically starts with a puzzle, with some phenomenon that is surprising given one's expectations. Should one regard tax compliance or tax evasion as surprising? Beginning, as economists so often do, with a simplified model of individual self-interest, it is puzzling why people comply with the tax law to the extent that they do, given the many opportunities to cheat. A view of tax compliance that runs entirely in terms of self-interested calculations about the costs and benefits of cheating may suggest the absence of other reasons to comply, endorsing through silence a pessimistic view of the impotence of moral constraints on selfishness. Analyses that discount other, moral reasons for complying with the law may implicitly endorse a view that whatever is rational from a self-interested point of view is right.

Those who grant the existence of moral and public-spirited reasons for paying one's taxes might instead think it puzzling that people cheat on their taxes. They are part of a society that depends on public services financed by taxes, they had the opportunity (assuming a democratic state) to participate in a public process that determined the taxes they owed, et cetera. Why in the face of obvious good reason to pay taxes – reasons reinforced by the state's police powers – do people cheat? Material self-interest surely has a great deal to do with cheating, but it is one factor among many.

Explanations of choices in terms of preferences for larger bundles of commodities or greater net revenues are often entirely adequate: no heavy moral lifting is involved in predicting a firm's decision about the color schemes for next year's new cars on the basis of its expectations concerning sales and ultimately profits. But when economists proceed to analyze morally problematic choices in the same manner, larger issues are at stake. It is one thing to assume, as Adam Smith did, that market incentives would determine behavior "within the rules of justice" (1776, bk. II, ch. 4). It is quite a different thing to assume that self-interested calculations will determine whether people obey those rules or observe any moral limitations on their choices.

For example, suppose that a firm knowingly dumps its toxic wastes into a stream and poisons an area of marshland. If the costs of the dumping in terms of possible lawsuits or loss of reputation were less than the costs of disposing of the wastes in some harmless way, then the firm increases its net revenues by dumping the wastes and poisoning the marshland. Some economists might cite this fact to explain the dumping even if the firm would have had good earnings in any case. An observer might be led to wonder whether moral reasons carried any weight with the firm's managers. Perhaps they didn't realize the damage was serious, or perhaps they were in more desperate straits than initially supposed. To the degree that economists assume that the only reasons to be sought in explaining the firm's behavior are self-interested reasons – like those justifying car color choices – their analysis will tend to excuse or justify the firm's behavior and to locate responsibility for the pollution not with the firm but entirely with the government for failing to set the firm's incentives properly. It's as if the only thing wrong with Willie Sutton's response about bank robbing was that he underestimated the risks of getting caught or the benefits of legal occupations.

A more persuasive argument an economist might offer for ignoring moral restraints on a firm's polluting behavior would be that the firm had little choice because competitive pressures do not permit firms the luxury of moral scruples: If the polluting firm took the high road then it would be driven out of business, and the less scrupulous surviving firms would continue to pollute and thereby cause similar damage.

The argument that competitive pressures exclude moral considerations (or that firms that take morality into account have already been eliminated) is a complicated one that is seldom carefully presented. Notice first that the conclusion does not follow that taking the high ground never does any good. Even if competitive pressures were to drive the firm out of business if it refused to pollute the marsh, those pressures may not lead any competitor to open a factory at the same location or to cause the same pollution. All the competitive pressures argument can show is that the alternatives to pollution are not merely lower profits but also bankruptcy and a risk that the pollution will occur just the same.

At this point, moral judgment re-enters. To take avoiding bankruptcy as an understandable and unremarkable reason for polluting a marsh is to presuppose a limit on the weight that owners or managers give to moral considerations. This judgment is less objectionable than is the judgment (or presupposition) that moral scruples have no weight at all when there is a conflict between profit and social good, but it likewise tends to excuse or justify the firm's behavior. If a firm facing bankruptcy hired gangsters to kill the managers of a competitor, few economists would show the same attitude toward an explanation in terms of economic pressures.

There are also serious questions concerning the argument that the price of morality is bankruptcy. First, note that the argument does not go through if all firms show the same scruples, for then none is at a competitive disadvantage. So, in giving the argument, one is taking for granted that some firms do things such as poisoning marshes in order to increase profits. Second, the strength of competitive pressures varies enormously. For example, from the time of its formation in 1901 to 1991, the market share of the United States Steel Corporation constantly decreased, yet it survived (Mueller 1992, pp. 309–10). Third, this selection argument rests on the empirical claim that firms observing moral constraints will lose out in competition to less scrupulous firms, and it is not obvious that this claim is true.

A reputation for honest dealing with customers, for decent treatment of employees, and for pursuing worthy objectives may all increase a firm's profitability. Consider data such as the following gathered by Robert Frank (2003): In 1989, starting salaries in the most prestigious U.S. public interest organizations, such as the American Civil Liberties Union, peaked at \$29,000, while prestigious private firms paid up to \$83,000 (p. 83). Students in a study at Cornell University claimed that they would have to be paid $10,000 more to work as a lawyer for the National Rifle Association than as a lawyer for the Sierra Club or $15,000 more to write advertising copy for Camel cigarettes than for the American Cancer Society (pp. 87–8). There is thus evidence that "moral" firms can save on wages, and they are likely to have less shirking and lower costs from pilfering or vandalism.

A firm's moral behavior *may* result from a calculation of its good effect on the firm's profitability, but it need not and we suspect that often it does not. A manager or owner may follow certain policies simply because they seem right. The overall effect on the firm's profitability of acting in this un–self-interested moral way may well be more positive than the effect of a more calculated approach involving decisions about whether doing the right thing would have the desired long-run effects on reputation and profitability (see Frank 2003). Such considerations raise doubts about the general argument that competitive pressures compel firms to neglect moral concerns.

Economists should take these issues seriously for at least two reasons. First, if moral considerations do affect firms' and consumers' choices, then economists can strengthen the explanatory and predictive power of their analyses by including such factors. And second, resting content with self-interested explanations of immoral behavior can have the effect of excusing that behavior. To see nothing remarkable in destructive behavior motivated only by the prospect of additional profits is to endorse (or to be blind to) a pernicious moral cynicism. Equating rationality with self-interest – and then explaining and implicitly justifying morally questionable behavior solely on grounds of rationality – relies on questionable empirical assumptions; should such explanations become widespread, they risk becoming self-fulfilling prophecies. Such approval or excuse should be offered only reluctantly.

We may draw two implications for the explanation of behavior that is prima facie harmful or immoral. First, any argument that the behavior is mandated by competitive forces should be made explicit and scrutinized; there can be no blanket presumption that such arguments are sound. Second, if competitive pressures don't offer a satisfactory explanation, then economists should ask whether material acquisitiveness is a good reason in the particular case. Again, there can be no blanket presumption that moral considerations are irrelevant in explaining morally problematic behavior. Economists have shown increasing interest in studying morally freighted activities like theft, violent crime, environmental pollution, and trustworthiness. When assessing explanations in such areas it is important to recognize that moral as well as self-interested reasons can help explain behavior and that a precommitment to explaining all such behavior in terms of self-interested rationality tends to excuse or justify it. Most of us regard a desire for more commodities as a justification for working longer hours, but not as a justification for robbing banks. These moral and prudential judgments are reflected in our satisfaction with the explanation of the longer hours worked and our demand for further explanation of the bank robbing. Judgments of prudence and morality are implicit in attitudes toward explanations of human behavior.

5.5 Conclusions

Positive and normative economics are linked via the theory of rationality in complicated ways. Because of these links, ethical commitments have an unavoidable role within positive economics. Ethical commitments have some influence in all branches of science, at least on the mores of scientific inquiry and on the questions that are asked. But in the physical sciences the behavior of the objects under study is not (we trust) influenced by the objects' values and beliefs. In the social sciences, in contrast, beliefs and norms often play a significant role in explaining the behavior under study – a point we explore further in the next chapter.

Positive economics, like many other social sciences, explains choices in terms of reasons, and reasons are subject to evaluation. This does not mean that economists must evaluate them; they can cite the preferences and beliefs that explain an individual's choice without appraising them or judging the person who has these preferences and beliefs. But whether or not the evaluative questions are answered or even explicitly mentioned, the questions are always there; and it is sometimes important to ask them. In addition to the intrinsic interest of finding out what drives tastes or perpetuates myths, the assessment of beliefs and preferences may suggest further topics for empirical research. So if, for example, economists explain some market phenomenon by a widespread belief among market participants that interest rates will remain steady, *and* if economists go on to evaluate that belief as unsupported by data and theory, then they will want to pursue the question of how these unsupported beliefs have propagated and persisted. The theory of "rational expectations" in finance and macroeconomics emerged from just this kind of inquiry.

Although some economists are willing to cite people's moral beliefs in explaining some of their choices, many are strongly inclined to cling to the "solid ground" of material self-interest in constructing economic explanations and to leave no room for moral constraints or purposes in explaining behavior. This tendency may be at its strongest in the teaching of economics, where there is an interest in building up students' skill in seeing the hidden hand of self-interest behind behavior that is not obviously self-interested. In combating naiveté, economists may have cultivated cynicism rather than wisdom.

Suggestions for Further Reading

An excellent discussion of the relationship between facts and values, with special attention to the views of the economist Amartya Sen, is Putnam (2002). There has been a good deal of discussion of so-called folk psychology by philosophers of psychology. Good contemporary discussions can be found in Stich (1983), Dennett (1987), Dretske (1988), and Churchland (1989).

For discussions of reasons versus causes and of the possible difficulties involved in causal explanation of human behavior, see Winch (1958), Melden (1961), von Wright (1971), Rosenberg (1976, ch. 5), Davidson (1980), and Rosenberg (1995, ch. 2).

There is a considerable literature concerning preference reversals. The phenomenon was first discussed in Lichtenstein and Slovic (1971), and the first discussion by economists is Grether and Plott (1979). More recent views can be found in Tversky, Slovic, and Kahneman (1990) and Tversky and Thaler (1990). An overview of economists' reactions can be found in Hausman (1992, ch. 13).

The two main theorems of welfare economics are proven in Arrow and Hahn (1971) and discussed in Bator (1957), Graaff (1957), and Koopmans (1957). Kelman (1981) argues that an economistic framework for assessing environmental issues can affect how people behave toward the environment. Various aspects of the relationship between economic rationality and morality are discussed in Frank (2003).

The argument in Section 5.3 concerning the evaluative implications of accepting explanations in terms of reasons was suggested by some remarks of Frank Knight (1935).

SIX

Rationality, Norms, and Morality

Every society has moral norms, although none can boast of perfect compliance. People approve of behavior that conforms to the dictates of their morality and disapprove of behavior that violates established moral norms. Those who violate moral norms typically experience negative feelings of guilt or shame. Moral norms enable people to coordinate their actions more efficiently than would be possible without a shared morality. It is in everyone's interest to live in a society governed by moral norms requiring that people tell the truth, keep promises, refrain from actions of violence toward others, and so forth.

Yet it can be costly to do what is morally right. If one is on a sinking ship without a life jacket, there are obvious advantages to stealing one from another person. One might feel guilty about it later, but people can get over guilt; they cannot get over death. The central character in Woody Allen's *Crimes and Misdemeanors,* a distinguished physician who has his mistress murdered to protect his reputation, is surprised at how easily he represses his guilt. To go to one's death in order to comply with a moral norm against theft might even appear to be *irrational.* Although self-interest and morality may often march in step, at a certain point their paths seem to diverge.

How compatible are morality, self-interest, and rationality? This is an old philosophical question that we cannot hope to settle, but we can explore its ramifications for economics. In this chapter we shall argue that moral considerations deserve a place in economic modeling. To support this conclusion we show that moral norms have important economic consequences and that they cannot be explained away in terms that make no reference to morality. We shall also argue that moral commitments should not be regarded by economists merely as individual preferences or sociological facts. But first some clarification is in order.

6.1 Rationality and Self-Interest

People often use the word "rational" as a synonym for "prudent." To say that it's not rational to take a midnight stroll through the shrubbery of New York's Central Park (even with the recent drop in crime rates) is to say that it is not prudent to do so. Because of the risks, it is not in one's *interest* to take such a stroll. There is a close association between rationality and self-interest. In his classic work, *The Methods of Ethics,* Henry Sidgwick (1901) attempts to derive utilitarianism from intuitions concerning the nature of rationality, but in the end he concludes that what he calls "the method of egoism" – that is, self-interested conduct – is just as rational. Robert Frank, in his contemporary classic, *Passions within Reason,* writes: "I will use the terms 'rational behavior' and 'self-interested behavior' to mean the same thing" (1988, p. 2n).

The connection between rationality and self-interest is not unequivocal. Though most people would say that it is rational to steal a stranger's life jacket to save one's life, most would not say that it is irrational to refuse to do so, and very few people would regard as irrational all actions that apparently conflict with self-interest. Parents who sacrifice their planned Carribean vacation in order to pay for music lessons for their children are not necessarily irrational. So it seems as if people are inclined to regard both self-interest *and* self-sacrifice as rational.

Economists take preferences to be rational if they are complete and transitive, and they take choices to be rational if they are determined by preferences. This theory of rationality places no constraints on *what* a rational individual may prefer. So on this theory someone who prefers honesty to life itself and who accordingly does not steal another's life jacket may be perfectly rational. Rationality and self-interest appear to have nothing to do with one another, and there seems to be no conflict between morality and rationality.

Moral and altruistic preferences are ruled out of many economic models not by axioms of rationality but rather by the assumption that all people care about are bundles of commodities to be privately consumed. There is thus no room for altruistic or moral considerations. For example, a successfully stolen Porsche is the same commodity as one that is purchased or received as a gift, and the agent's preference for the Porsche will be the same if the objects of preference are commodities. But this does not imply any incompatibility between rationality and altruism or morality.

It is easy to interpret the standard theory of rationality mistakenly as a theory of self-interest. One might reason as follows: According to utility theory, it is rational to choose whatever one prefers. Rational choices are determined by one's own preferences rather than by anyone else's. So it sounds as if rational choices are self-interested and thus, insofar as it is rational, altruistic or moral behavior is actually self-interested. The individual who drowns rather than stealing another's life jacket, like the individual who has no such scruples, does what he most prefers and hence acts in his own interests. So the person who steals someone else's life jacket is no more self-interested than the one who chooses death over immorality.

This conclusion is bizarre, and the argument is fallacious. To be self-interested is to have preferences directed toward one's own good, not simply to act on one's own preferences. What distinguishes people who are self-interested from those who are altruistic or malevolent is *what they prefer*, and utility theory says nothing about what the content of rational preferences ought to be. Similarly, accepting utility theory as a correct description of people's actual choices and preferences leaves open the question of the extent to which individuals are self-interested. Hence it appears that there is no conflict between morality and the standard view of rationality. However, we shall argue that some moral phenomena cannot usefully be accommodated within utility theory.

6.2 The Influence of Moral Norms on Economic Behavior

Before we can determine whether morality and the standard theory of rationality are compatible and whether (and how) moral factors can be incorporated into economic models, we need to say something about the nature, causes, and consequences of specifically moral behavior. Social scientists and philosophers have different starting points when they study moral norms. Social scientists are concerned with the causes and consequences of norms, and especially with their role in predicting and explaining social phenomena. Thus, as Max Weber insisted, one can treat the norms present in a community as objective and causally significant matters of fact quite apart from the question of whether they are worth endorsing. Economists adopting this perspective may want to know how the moral norms prevailing in a certain community influence the behavior of the community's members and how moral norms are established, modified, and sustained. Philosophers, on the other hand, are mainly concerned with the validity and justification of norms. One may thus usefully distinguish *philosophical* inquiries into the validity of moral norms from *sociological* inquiries into their consequences and causes – even though, as we shall see, these inquiries are not completely separate.

Moral norms are a subclass of social norms, which we take (following Elster 1989b, p. 113) to be prescriptive rules regarding behavior that are shared among a group of people and that are partly sustained by the approval and disapproval of others. Social norms include items such as standards of dress and rules of etiquette. Specifically, moral norms are marked out by their subject matter (interpersonal interaction where significant benefits and harms are at stake), their weight (they typically override other considerations), and the sanctions – both internal (guilt) and external (blame) – attached to their violation. (See Elster 1989a; Pettit 1990; Goldfarb and Griffith 1991.)

Moral norms would be of little interest unless they did influence behavior. This point marks an important difference between moral claims and other kinds of assertions. An argument about how shotguns work can be framed and defended independently of what they are pointed at. But a coherent argument for a moral norm against murder must provide reasons why people should not shoot guns at their neighbors. Since any adequate theory of morality must include an account of how human beings can come to be motivated by moral considerations, moral philosophers must take some account of the sociology and psychology of human beings.

Some social scientists, including many economists, doubt that moral norms play an independent role in explaining people's behavior. They insist that satisfactory explanations require showing that behavior that is apparently influenced by norms is really ultimately self-interested – that people follow norms only when doing so serves their interests. This is then a challenge to those who claim an explanatory role for norms: they cannot simply take for granted that moral norms influence behavior.

As a way of gaining clarity about the role norms can play in economic analysis, we shall consider an example in which moral norms are invoked to explain "efficiency wages." Usually economists maintain that individuals are paid more because they are more productive. But paying people more may also induce them to work more productively. The idea of an efficiency wage is that, in certain circumstances, simply raising the wages of a group of workers will motivate them, other things being equal, to work more effectively. Efficiency wages are wages above the market-clearing rate that motivate workers to be more productive, not high wages that reward high productivity.

Efficiency wage theory has attracted a great deal of attention because it apparently helps to explain two important features of industrial economies. The first feature is involuntary unemployment, an issue we discussed in Chapter 3. If there are unemployed workers who are willing to work for lower wages, why don't wages fall to a level consistent with employing the entire labor force? Efficiency wage theory answers that, since high wages induce greater productivity, reducing wages could cost firms more from lost productivity than they gain from the savings in wage payments. The second feature that efficiency wage theory apparently explains is the existence of so-called dual labor markets. Some observers characterize the U.S. economy as having both a "casual" labor market characterized by high turnover and low (market-clearing) wages and a "career" labor market characterized by lower turnover, higher wages, and rationed entry through queues of applicants for a limited number of jobs. The presence of efficiency wages in firms whose jobs are in the career labor market category may be part of the explanation. Firms could hire equally well-qualified workers at lower wages, but at lower wages the workers would not be motivated to work as productively, and higher turnover would also be costly.

A variety of attempts has been made to explain efficiency wages in models that assume rational self-interested behavior on the part of both firms and workers. For example, Bowles and Gintis (1993) point out that it is difficult for firms to monitor and control the performance of their workers. If the only sanction firms have is the threat of firing, then firms are not going to be able to motivate their workers unless firing is costly to the workers, and firing will not be costly to workers unless firms pay more than market-clearing wages. Although the unemployment that results is also beneficial to firms, it is not something within their control and not a reason why they choose to pay wages above the market-clearing rate. Although Bowles and Gintis are Marxists and argue for the advantages of worker cooperatives, their account of efficiency wages relies only on self-interested maximizing choices by workers and employers. It does not rely on norms.

George Akerlof, in contrast, has offered a socioeconomic explanation of efficiency wages in terms of what he calls "partial gift exchange" (1982). Akerlof revisits a case study developed originally by the sociologist George Homans. Homans examined "cash posters" at the Eastern Utilities Company, whose job was to record customers' payments on ledger cards at the time of receipt. It was easy to measure the workers' output and to compare it to the company's performance standard. Of the group of ten workers, two barely met the standard of 300 postings per hour, but the average number of postings per hour (353) was 17.7% above the standard. Despite considerable variation in performance (the range was from 306 to 438 postings per hour), all workers received the same wage (above the market-clearing wage), differential promotion was not an issue, and there were no serious efforts to punish or threaten the slower workers. These facts are hard to reconcile with standard economic theory. If workers prefer gossiping, coffee drinking, or daydreaming to recording payments on ledger cards, then those who post more than the standard 300 should either slack off or demand more pay. If the company, for its part, wants to increase its returns, then it seems that it should raise its performance standards – at least for the more productive workers. If these workers exceed the standard without any reward then they should be willing to increase their rate of postings further in exchange for additional compensation. Higher compensation for more productive workers should benefit both the firm and the workers.

Akerlof argues that norms explain these facts. Both the firm and the workers adhere to a norm of "gift exchange" for effort and wages above the minimum expected. Workers receive a gift of more than the market-clearing wages and in return they give more than the minimum standard of effort. These transactions are governed by the norm of "a fair day's work for a fair day's pay." To explain the absence of pay differentials and the firm's willingness to maintain low-performing workers in their jobs, Akerlof appeals to further norms of fair treatment among workers. Introduction of pay differentials or efforts to dismiss the less productive workers might result in withdrawal of some of the extra effort. (Akerlof reports evidence that slowdowns occurred in the past when the company tried to toughen up.) In the presence of a "gift" from the workers of 17.7% greater output, the firm has reason to respect the workers' norms of fair treatment.

The fact that in this example the norms governing workers' effort decisions are sensitive both to wages and wage differentials introduces a basic difference between the workings of a pure market and a market with gift exchange. Akerlof shows in a formalization of the cash posters' story that the observed phenomena of wage equality and effort above the minimum required are consistent with the firm maximizing profits and the workers maximizing utility in the presence of this relationship between norms and wages.

It is hard to judge whether Akerlof's explanation of the cash posters' case is correct because, as Akerlof notes, it is possible to construct many subtle neoclassical explanations of these facts which avoid any appeal to norms and rely instead on turnover costs, measurement difficulties, and the like. The sociological explanation is plausible: norms regarding gift giving are familiar in the United States, and the "ties of sentiment" that develop among workers and even between workers and their employing firm are a plausible ground for the emergence and maintenance of such norms.

The norms that figure in Akerlof's explanation raise further questions. How are they generated? Why does the relation between work norms and perceived fairness take precisely this form? How stable are these norms in response to changes in other factors? Are there other norms whose observance would be more efficient? These are, in our view, serious questions to pursue. However, some might argue that these further questions raised by Akerlof's explanation of the cash posters show that the explanation is unsatisfactory, because it is not sufficiently "fundamental." Many economists would suggest that a fuller picture would ultimately reduce the norm-guided behavior to sophisticated self-interested choices.

Akerlof's account raises questions about the origin and determinants of social norms, questions that for now have no entirely satisfactory answers. It is, however, a huge leap from recognizing these questions to claiming that good answers to them will reduce everything to self-interest and so produce an explanation of the sort mainstream economists prefer. Whether there is an adequate account in terms of economic self-interest is precisely what is at issue here. It begs the question to insist that there must be one.

As this example from Akerlof illustrates, sociological and normative aspects of work relations may influence the macroeconomics as well as the microeconomics of labor markets. In the view of some, the "efficiency wage" story offers a promising route to explaining the existence of involuntary unemployment in equilibrium. (In addition to Akerlof, see Solow 1981, Bowles 1985, Blinder and Choi 1990, and De Vroey 2004.) However, an adequate macroeconomic explanation of unemployment grounded in efficiency wage theory requires solving further puzzles. Why don't unemployed workers undercut the prevailing above–market-clearing wage by offering to work for less than those with jobs? Why don't employers actively seek such workers, perhaps differentiating the workforce to pay more to those with established jobs and less to the eager newcomers? Solow (1990, ch. 2) and Weibull (1987) suggest that further social norms come into play – a norm among workers that discourages bidding against their fellows for jobs and a norm among employers against seeking to undercut their existing workforce.

There are many other areas of economic life in which it appears that moral norms are influential. For example, the regulation of negative externalities like air pollution and littering is accomplished partly through legal sanctions, but norms against such antisocial behavior also play a significant role. The same may be said about the provision of public goods in general and about voting. Public transport in many European cities operates on an honor system with only sporadic ticket checks. One cannot assume, however, that norms will always do the work. Honor systems with sporadic checks work smoothly in much of Germany, Austria, and Switzerland, whereas cheating is rampant in France.

A particularly central example of a public good whose provision relies on moral norms is the existence of a legal system, including a system of property rules (Arrow 1974). As Adam Smith stressed in the eighteenth century, no legal system – certainly not one that allowed widespread commercial freedom – could rely purely on criminal sanctions for its enforcement. A widespread conviction that people ought to obey the law is essential to the maintenance of social order. Moral norms do apparently matter enormously. Unless these appearances can be shown to be misleading or the persistence of the norms explained in terms of rational self-interest, moral norms must be recognized as significant causal factors that are sometimes of major importance.

6.3 How Do Norms Motivate and What Sustains Them?

Automobile manufacturers sometimes initiate costly recalls of their product even when they are not required to do so. One possible explanation is that there are norms requiring that manufacturers repair flaws in merchandise. Thus, one might explain a particular voluntary recall by General Motors or Toyota by mentioning the norm and citing the generalization that, other things being equal, agents follow norms. Few economists would be satisfied with such an explanation. They would insist that one must explain why the management of Toyota or General Motors decided to comply with the norm and why the norm persists. As we mentioned before, many economists would go further still and deny that one has an explanation until one has demonstrated that the action is in the "material" or "economic" self-interest of the agents. The qualifiers "material" or "economic" are needed to give some teeth to the constraint. If being in compliance with the norm is itself counted as in the interest of the agent, then it is trivially true that norm following is self-interested. What most economists would like to show is that norm following increases the profits of firms or the consumption or leisure of individuals. Although it is not, in our view, reasonable to insist that only material self-interest motivates, the other challenges are serious. Norms do not descend from the heavens, as the Ten Commandments are alleged to have done. They are somehow created, modified, sustained, and enforced by people, and sometimes they break down. An explanation of behavior that merely cites a norm – without considering what sustains and enforces it – is shallow.

Let us shift to an example that involves individuals and where a moral norm explanation is more persuasive than in the case of an automobile recall. In a classic experiment, researchers left wallets containing identification and a small amount of cash on the streets of New York City (Hornstein, Fisch, and Holmes 1968). Almost half were returned with the cash intact. It is hard to understand how returning the wallets could have been advantageous for those who found them. It is troublesome to pack up a wallet and mail it (particularly in New York City, where there are long lines at the post offices). Presumably the reason why most people returned the wallet was either a direct altruistic concern for the owner of the wallet or a moral norm against appropriating other people's property. Since altruism tends to be weaker when one has no vivid impression of the other person, it seems reasonable to assume that the norm against theft is the main explanation for why people returned the wallets (Batson 1993).

Just how does a norm motivate individuals to comply with it? One possibility is that the moral reasons justifying the norm themselves motivate the compliance. People's desire to act on principles that no rational agent could reasonably reject (Scanlon 1998) – or their desire not to free ride but rather to do their part in practices of which they approve – can motivate them to follow norms, even when doing so will not benefit them materially. Explanations like this may not be congenial to most economists, but we believe they are often correct. And if such explanations are correct, then the justification for a set of norms, and not merely the fact that the norms are accepted, may be important empirically. If people comply with some norms because they accept the moral reasons that justify those norms, then a critique of those reasons may have important behavioral consequences. And if reasons have such empirical importance then economists may, for some explanatory purposes, need to understand how norms are justified.

Economists and sociologists influenced by economists have tended to focus instead on the individual *sanctions* motivating compliance with norms. These sanctions may not be material and hence they may be suspect to some economists, but they are very much the stuff of which everyday life is built. If one returns the wallet, one feels good about oneself, and the receipt of a thank-you note makes one feel even better. One can tell family, friends, and associates, and in response one gains their trust and praise. Indeed, in this way it is possible to garner some material advantage, though usually not enough to compensate for the trouble of returning the wallet.

On the other hand, if one does not return the wallet then, if one has been "well brought up" (and hasn't studied too much economics!), one feels guilty. Moreover, if this failure to return the wallet becomes known to others, one may be censured or may suffer from knowing that one has lost their esteem, even if their censure is never expressed (Pettit 1990, p. 740). It is important to be careful about viewing these guilt feelings and perceived loss of esteem as "sanctions." These reactions have an important cognitive aspect and may flow directly from the belief that we did wrong. The idea that these are like penalties others impose on us treats conscience as if it were something external, like some sort of immaterial ball and chain that prevents us from doing what we want and punishes us when we lapse (Nussbaum 2001). That said, it is surely true that sanctions, both internal and external, play a significant role in explaining compliance with norms: Individuals comply with norms in large part because of the individual benefits attached to doing so and because of the costs of not doing so. These reasons are important, but – as we have already said – we do not think this is the whole story.

To say that individuals comply with norms because of sanctions does not get one far. Sanctions have to be applied by people, and it is typically no more materially advantageous to punish or reward others for complying with norms than it is to comply with norms in the first place. Given that others (including one's "superego") will punish violations and reward compliance, it may be advantageous to comply with norms. But this only explains why one person should comply with norms if others do. It does not explain why people generally comply with norms, nor why the norms persist.

One might attempt to explain why norms persist in terms of their social benefits. Groups with strong norms of truth telling and promise keeping do better than groups without such norms. But in any sizeable group, no individual can have much effect on the norms obtaining in the group or on average compliance. So although the social benefits of a system of norms may provide a moral reason for complying with them, these benefits do not generally provide a self-interested reason. On the contrary, one often does better by free riding on the norm following of others. Since there are few sanctions attached to not sanctioning others, rational individuals who are concerned only about their own profits, consumption, or leisure will rarely trouble themselves to punish or reward others for violating or complying with norms. So if one knows that one's fellows are rational and self-interested agents of the sort found in economic models, then one should have little to fear in the way of sanctions against violating norms and little to expect in the way of reward for complying with them; and so it seems as if norms cannot be effective. But norms are effective, and hence people's commitment to them cannot be explained in terms of economic self-interest.

We do not maintain that this argument is an ironclad demonstration that norms cannot be generated or sustained by a community of rational and self-interested economic agents. Involuntary attitudes of approval or disapproval (which are not themselves rationally self-interested) may create the needed incentives to conform to norms (Pettit 1990; Brennan and Pettit 2004), or it may be that the costs and benefits of sanctioning sometimes make sanctioning rational (Coleman 1990, ch. 10). But unless norms are of no economic importance, which is very hard to believe, economists need to find a place within their theory of economic behavior for a discussion of when norms will form, of how much influence they might have, and of what their consequences will be. If the existence and influence of moral norms depend on the strength of moral reasons (as we have argued), then economists will not be able to treat moral norms as brute sociological facts that need not be understood or assessed.

In recent years, there has been a great deal of interest concerning the role that both biological natural selection and broader conceptions of social evolution may play in the development and persistence of moral norms. This work both draws on and informs economics. While not assuming that individuals are psychologically self-interested, these lines of inquiry focus on the complex interplay between the advantages to groups and to their members that shared norms provide and the advantages that may accrue to individuals if they are able to free ride on the norm following of others. One way to achieve compliance is to build it in biologically, and practitioners of "evolutionary psychology" have argued that receptiveness to following norms and even certain more specific normative commitments – such as caring about children or being disposed to act reciprocally toward those who treat us well – have a basis in biological evolution. These arguments are controversial, and the line between scientific evidence and interesting conjecture can be difficult to draw in this arena. Obviously, natural selection has so far been compatible with the moral capacities and proclivities of human beings. But it is no easy task to determine whether humans survived because of their morality or despite it.

Although biology constrains the results, many of the processes that are responsible for norms and institutions are social rather than biological and are at most analogous to biological evolution. We may pose at least two kinds of questions concerning social evolution. First, given a set of individuals within a particular social setting, how does their compliance affect how they fare over time? The tools of game theory (discussed in Chapter 14) are crucial here. In repeated interactions, various cooperative strategies may have a high survival value and tend to predominate, but there may be benefits to free riding, too. Second, one can use the tools of game theory to study the contribution that norms make to the survival of groups. These two inquiries can also be combined: for at the same time that temptations to free ride may be undermining norms within groups, the cooperation that norms make possible may be helping those societies with strict norms and high compliance to supplant societies with weak norms.

6.4 Philosophical Implications

The evidence cited so far strengthens the case that normative commitments matter. In this respect these studies bolster the case for taking morality seriously. At the same time, these studies say little about *how* moral norms influence individuals' economic conduct. As Allan Gibbard (1990) has argued, the psychology of norm-guided behavior is not well understood. What exactly does it mean, for example, for a norm to be "internalized"? Gibbard distinguishes the psychological state of "accepting" a norm from the state of "being in the grip" of a norm. The former includes a deliberative judgment and a conscious endorsement of the norm, which typically result in behavior conforming to the norm. The latter is an involuntary state in which one finds oneself moved by the demand of a norm – without judging the demand as worthwhile and, indeed, perhaps in the face of a judgment that the demand is pointless or improper. Most people are in the grip of norms governing how close they stand to each other in conversation, whereas most people accept norms governing proper dress. It is not always clear whether people accept norms or are merely in their grip. Did the cash posters, for example, understand and endorse their gift exchange norms? From the point of view of their overt behavior at a given time, it may not matter much, but different factors explain why people accept norms than explain why they are in the grip of norms. From the standpoint of most moral philosophy, which links moral conduct to reasoned decisions, the cognitive element involved in accepting a norm in Gibbard's sense is an important feature.

Empirical studies by sociologists and economists of the consequences of norms and of the factors that cause people to accept or reject norms can be relevant to philosophical inquiries concerning the content of morality. Though it is obvious that widely accepted norms are not automatically morally defensible, the reasons why norms are widely accepted and the consequences of norms may shed light on the question of whether such norms are justified. For example, Arrow (1974) and Reder (1979) have shown how the virtues of honesty and trustworthiness may promote economic efficiency in circumstances of uncertainty. Arrow has also argued more specifically that medical codes of ethics may be an efficient and socially desirable response to physicians' opportunities to exploit their informational advantages over patients. Arguments of this kind may contribute to the justification of these norms by showing that they lead to good consequences, including consequences for economic well-being. Those who recognize these consequences have a reason (though not necessarily a self-interested reason) to accept the norm, and in this way the good consequences may also help explain why people embrace the norm in their behavior.

The good consequences of a moral norm are not, however, by themselves enough to explain its emergence and persistence. As Elster (1989a,b) has argued in his critique of functional explanations, to explain the existence or persistence of a norm also requires an account of the mechanisms by which the favorable consequences produce or sustain the practice (see also Cohen 1982). For example, suppose (as many Marxists have alleged) that racism has prevented the American working class from organizing to oppose the interests of capitalists. The fact that this supposed consequence of racism is beneficial to capitalists does not explain why so many Americans have been racists unless some account can be given of the mechanism by which this consequence of racism leads Americans to be racist. In biology, natural selection provides such a mechanism: an organism with a fitness-enhancing trait is more likely to have offspring, and so that trait will become common. There are some social mechanisms that can help explain the persistence of norms: recognizing the good consequences and seeking to bring them about is one simple one. But the mechanisms required by functional explanations in the social sciences are often missing, and functional explanations are often dubious.

Moral norms do not always have good economic consequences. The norms regarding work rules and wages discussed earlier in this chapter seem ambiguous in this regard. On one hand, these norms promote wage equality, high wages, and stable and productive working groups; these would widely be judged to be good consequences. Yet these norms may also contribute to unemployment. Note also that more established "in groups" may benefit disproportionately from these norms while minority and disadvantaged groups are excluded from the benefits.

What people take to be moral norms are sometimes morally indefensible. For example, Amartya Sen points out that in many societies the norms governing the distribution of resources and power within families discourage women from objecting to intrafamilial inequalities in the distribution of food between boys and girls. These inequalities apparently lead to literally millions of premature female deaths in poor countries (Sen 1990). What is relevant to whether moral norms are justified is not whether they are influential but whether they are beneficial. And even the fact that a norm is socially beneficial may not be sufficient to justify it if the benefits are obtained by means that are morally objectionable.

6.5 Morality and Utility Theory

We have argued that mainstream economists need to broaden their theory of economic behavior in order to consider when norms will form and what sort of influence they may have. Doing so requires recognizing that economic agents are not always self-interested and that self-interest is not always material. Economic agents do not care only about the commodities and leisure they may consume. Sometimes they also want to benefit or to harm other people, and even when self-interested they may care more about the esteem and affection of others than about the size of their houses or style of their clothing. For example, it would be just as foolish for those designing a health-care system to suppose that doctors and nurses don't care whether their patients live or die as to suppose that doctors and nurses don't care about their salaries (Hausman and Legrand 1999). Since there is nothing in the formal theory of rationality that bars altruistic and moral preferences, these data do not refute the theory. But even if utility maximization and morally motivated conduct are formally consistent, utility theory might not be very helpful in analyzing moral conduct. Sen (1977) has argued that, even though utility theory may be too demanding in some regards (such as in its insistence that preferences be complete), it may at the same time impose *too little* structure on problems of choice to allow productive analysis of moral conduct.

Some of the structure that needs to be added creates no formal difficulties. For example, one can distinguish self-regarding or egoistic from other-regarding or altruistic preferences by whether or not either the utility or consumption levels of other individuals enter into a person's utility function. Someone who is altruistic will prefer states of affairs in which others consume more or in which their preferences are better satisfied. Someone who is envious or malevolent will prefer states of affairs in which others do worse. Terminologically, it is useful to refer to the satisfaction of self-regarding preferences by some term such as "personal interests" or "standard of living," which typically differs from the overall extent to which a person's preferences are satisfied. But one must be careful not to forget that self-regarding preferences include preferences for the approval of others. There is no canonical usage here, and what is important is not the precise names but rather an awareness of distinctions among the concepts (see Sen et al. 1987; Hausman 2005).

Not all nonegoistic concerns are captured by altruistic preferences. For example, one may be concerned not only with the level of another person's welfare or consumption, but with how it is attained. In particular, one may wish to be an active contributor to another's well-being. We not only want our children to acquire good moral characters, we also want some of the responsibility for their upbringing. As Arrow has noted, the desire to contribute may have substantially different implications than mere altruism. If many rich people prefer a higher level of consumption for "the poor," then there may be a "free rider" problem: people want poverty eliminated, but they'd like others to pay for it. If instead each wishes to contribute to helping the poor, the "public goods" or free-rider problem goes away (Hochman and Rodgers 1969). More generally, economists should not ignore the desire to *do* certain things rather than simply to enjoy the consequences of their being done. But once again there is no formal difficulty about capturing such desires within utility theory. All one need do is include actions that involve helping others among the objects of preferences.

A more serious complication arises because morality is not the same thing as altruism. An altruistic concern for the members of one's family is consistent with bloodthirsty behavior toward the rest of the world, and many people regard prudence as a moral virtue even though it concerns only oneself. Even generalized altruism leaves out important features of morality. A commitment to property rights constrains not only those who would steal for themselves but also those who would steal for others. People see their moral commitments as sometimes conflicting with their preferences, whether these be self-interested or altruistic. Those who act on principle may in fact also benefit themselves, but if genuinely motivated by moral principle then they will "do their duty" even if it's not personally beneficial.

Mainstream economists regard preferences as ranking alternatives in terms of all relevant considerations, including moral considerations. So by definition there could not be a conflict between an agent's moral commitments and an agent's preferences. Moral commitments play their part in determining preferences, not in mediating between preference and choice. As we shall see in Chapter 8, most economists are also inclined to take an agent's well-being to be the extent to which that agent's preferences are satisfied. Taking moral commitments to be incorporated into preferences and welfare to be preference satisfaction equates morality and self-interest by dint of arbitrary definition.

If one takes an agent's moral commitments as influencing an agent's preferences rather than as competing with them, one cannot at the same time take the satisfaction of preferences as personal welfare unless one is prepared to argue, rather than merely stipulate, that moral commitments are self-interested. Provided that one does not take the further step of identifying welfare and preference satisfaction, it might seem sensible formally to model an agent with moral commitments as satisfying the axioms of rational choice theory and consequently as maximizing utility. But, as Sen (1977) argues, it may be more enlightening not to do so. For then one can begin to consider what role morality might play in *constraining* choices, and one can recognize the special problems of self-control and weakness of will to which moral duties give rise.

By not modeling moral action and deliberation as utility maximization, one might be able to capture more directly and simply the phenomena of reflecting on, assessing, and sometimes revising one's motives or reasons for acting. It may be useful to suppose that individuals have different preference rankings or utilities depending on whether they consider matters from a self-interested, an altruistic, or a moral perspective or to suppose that individuals have "metapreferences" – that is, preferences concerning what their preferences should be. Multiple preference systems very naturally model internal conflict concerning such personal choices as whether or not to smoke (Schelling 1984, pp. 57–112) and also more clearly moral choices such as whether to contribute to charity (Frankfurt 1971; Etzioni 1988, pp. 177–80). Some have argued that the choices of morally influenced persons are better represented in such models (Margolis 1982; Etzioni 1988). However, this approach also raises difficult questions, for it is not clear within the multiple-preference framework how to explain which ranking will prevail in which circumstances. If that determination is itself made by a consistent preference ranking at a higher level, then it may make sense to re-invoke standard utility theory.

If one takes seriously the reasons for regarding moral concerns as *competing with* preferences rather than as *determining* preferences, then there is reason to regard some of morality as "irrational" in terms of the notion of rationality defined by utility theory. But one might instead question whether utility theory defines rationality in an acceptable way. If one takes acting rationally to be acting on good reasons, then there seems to be no basis for a conflict between rationality and morality. Even though this last statement in defense of the rationality of morality is vague, it is important to recognize that the alleged "irrationality" of moral commitment is very different from what people generally take irrationality to be. Moral commitment is not like hallucinating or contradicting oneself, since it depends on reasons and responds to arguments.

6.6 Conclusion: On the Rationality of Morality

The idealized economic agent of contemporary economics is a rational agent. This fact embodies a commitment to a modeling strategy and to a theory of prudence. The theory of rationality is already a fragment of a theory of morality. But the view of rationality that most economists endorse – utility theory – may not provide a rich enough picture of individual choice to permit one to discuss the character, causes, and consequences of moral behavior. Economists need not, of course, aspire to provide a general theory of human action. Yet they should not shrug their shoulders at the difficulties in meshing moral behavior with economic rationality. For it seems that moral behavior may have important consequences for economic outcomes, and the propagation of utility theory has moral implications.

Suggestions for Further Reading

For the general contrast between “self-interest” and “present-aim” theories of rationality, see Parfit (1984, pt. 2). For more on the relations between rationality and self-interest in economics, see Sen (1977). On the interdependence of Adam Smith’s views on morality and on economics, see Fleischacker (2004).

Useful discussions of norms can be found in Elster (1989a,b); Coleman (1990); Pettit (1990); Goldfarb and Griffith (1991); Frank (1993); Ben-Ner and Putterman (1998); and Brennan and Pettit (2004). Efficiency wage theory is discussed in Akerlof (1984), Akerlof and Yellen (1986), and Solow (1990). An informed though polemical popular overview of evolutionary psychology can be found in Pinker (2002).

The possible influence of moral factors in determining the efficiency of alternative ways of organizing the workplace is discussed in Reich and Devine (1981); McPherson (1983a); Putterman (1984); Bowles (1985); and Bowles and Gintis (1993). A related but distinct point is the role of *trust* in governing worker–employer relations, which is discussed in Arrow (1974), McKean (1975), and Gambetta (1988).

General problems concerning the explanation of public goods provision and the role that moral norms may play can be found in Schelling (1978); Kelman (1981); Taylor (1987); Elster (1989a); and Mansbridge (1990). For discussions of the existence of altruism and its possible importance in economics see Arrow (1972); Boulding (1978); Collard (1978); Becker (1981); and Batson (1993).

Multiple-preference rankings are introduced and discussed in Frankfurt (1971); Sen (1977a); Schelling (1984); Etzioni (1988); and Brennan (1989). The distinction in Section 6.5 between moral commitments and altruistic preferences is influenced by Sen's widely cited distinction between "sympathy" and "commitment" (1977); but sympathy (as Sen defines it) is not altruism, and commitment need have nothing to do with moral principle. See Hausman (2005).

PART TWO

WELFARE AND CONSEQUENCES

Economic outcomes may be better or worse along several dimensions. Some may make people better-off. Other outcomes may show more respect for human dignity, and still others may permit greater freedom. To decide which dimensions are more important requires moral judgment.

As illustrated by Larry Summers's memorandum (discussed in Chapter 2), economists evaluate outcomes in terms of individual welfare. One outcome is better than another if and only if the first makes people better-off than the second. Since the evaluation of outcomes rests exclusively on their consequences for individual welfare, the theory of individual welfare is crucial to normative economics. Indeed, normative economics is often called "welfare economics." As we saw in Chapter 5, economists take welfare to be the satisfaction of preferences.

Economists evaluate institutions and policies as well as outcomes. In principle, institutions and policies might be evaluated in many ways; for example, markets might be valued because of the freedoms they involve regardless of whether they promote individual well-being. But in fact economists assess institutions and policies in terms of their welfare consequences. So, for instance, economists typically value freedoms for their contribution to welfare. Economists focus on only one of the many evaluative questions they could ask about economic institutions, policies, and outcomes: "How well do they satisfy preferences?"

Part II focuses on *welfarism* – the evaluation of outcomes, institutions, actions, and policies in terms of their effect on individual welfare. Chapter 7 discusses utilitarianism, which ties assessment to total or average well-being. In focusing entirely on consequences for individual well-being, utilitarianism resembles welfare economics; but unlike welfare economics, utilitarianism requires a notion of well-being that can be compared across persons. Otherwise, there would be no way to add up or average people's well-being.

Utilitarianism preceded and influenced the development of contemporary economics. It purports to be a complete moral theory – assessing individual actions and characters as well as economic policies, institutions, and outcomes – and it is not necessarily wedded to a preference satisfaction view of well-being, which we examine in Chapter 8. Chapter 9 concludes Part II with a discussion of standard welfare economics and the way it attempts to capture a part of utilitarianism while remaining faithful to a preference satisfaction view of welfare and rejecting interpersonal welfare comparisons.

SEVEN

Utilitarianism and Consequentialism

Consequentialism is the doctrine that one should judge things morally by their intrinsic value and the value of their consequences. It specifies a particular *structure* for ethics. In a consequentialist framework one must first decide what is intrinsically valuable. Questions of intrinsic value are not necessarily the most important moral questions, but they must be answered first because everything else depends on their answers. Then one assesses actions, policies, and institutions in terms of their "results" – that is, their own value and the value of their consequences. As we shall see, welfare economics presupposes a consequentialist moral theory.

In particular, a consequentialist takes an action, policy, or institution to be morally right or permissible if its results are no worse on the whole than the results of any alternative. If the results of a particular policy are better than those of any alternative, then the policy is morally *obligatory.* Whether a policy or action is right or wrong depends both on how things will be if it is implemented and how things will be if any alternative is implemented. Appraisal is always comparative. The right thing to do may have very bad consequences when the consequences of the alternatives are even worse, and even a policy with terrific results is impermissible if there is an alternative with better results.

A consequentialist thus says that one should do whatever maximizes the good. A utilitarian is a consequentialist who says that what is good is individual "welfare" or "well-being." As we shall see, there are different varieties of utilitarianism. Maximizing total or average welfare is a different thing if individual welfare is some mental state like happiness than if it is instead the satisfaction of actual preferences, the satisfaction of "rational" or "informed" preferences, or something else not tied to preferences or mental states.

The fundamental thesis of utilitarianism is that one should do whatever maximizes overall welfare. This formulation masks a disagreement between utilitarians who favor the maximization of total welfare and those who support maximizing average welfare. Adding up or averaging well-being requires measuring the well-being of different individuals on the same scale or (in other words) comparing the increases or decreases in one person's well-being to the increases or decreases in the well-being of others. Notice that utilitarianism is not a selfish doctrine, since the objective is not to maximize one's own welfare. Indeed, one objection that has been pressed against utilitarianism is that it may require that people make large sacrifices to benefit total strangers (Singer 1972; Unger 1996) or that people refrain from favoring their friends or family over strangers unless doing so increases total welfare.

In order to get clear on what utilitarianism is and to assess its plausibility, think about, for example, how a utilitarian would approach the question of whether capital punishment is justified. Since the entire focus of utilitarianism is on the consequences of policies (such as executing certain classes of criminals), questions such as whether (say) murderers deserve to die will be irrelevant. The relevant question is whether capital punishment results in more total or average welfare than the alternatives. So a utilitarian argument for or against capital punishment will depend on factual issues such as whether capital punishment deters people from committing crimes, whether criminals who are not executed might commit other crimes, and whether convicted criminals might come to contribute to total welfare. Everybody's welfare, including that of the criminal, counts in this assessment.

The application of utilitarianism can be a very subtle matter. For example, although utilitarianism looks forward to consequences to determine what system of punishment is morally permissible, it does not say that judges should look forward to the consequences of alternative punishments when they are determining sentences. Whether permitting judges this much discretion will increase or decrease total welfare is an empirical question, and it is plausible that a utilitarian would favor a judicial system that requires sentencing to be determined mainly by the criminal's past behavior rather than by estimates of future outcomes.

7.1 Clarifying Utilitarianism

The sketch of utilitarianism just provided leaves many questions unanswered. First, individual "welfare" can mean many different things. Happiness and the satisfaction of preferences are not the same thing even for those whose only interest lies in their own happiness, because satisfying an agent's preferences does not always make the agent happier. Since happiness is not the same thing as satisfying preferences, a utilitarianism that aims to maximize happiness is not the same as one that aims to maximize the satisfaction of preferences. The major nineteenth-century utilitarians (especially Bentham, Mill, and Sidgwick) took utility to be a mental state like happiness or pleasure (or, more precisely, to be that property of objects that *causes* such mental states; see Broome 1991a), but few contemporary moral philosophers agree. Most contemporary utilitarians take welfare to be the satisfaction of rational and self-interested preference, but Chapter 8 will show that it is difficult to come up with a plausible concept of welfare – let alone one that is also sufficiently measurable to allow calculating the effects of actions on the total amount of this good.

A second difficulty is that the consequences of actions, policies, and institutions are uncertain. Legislators may institute a policy believing mistakenly that its consequences will be better than any of the alternatives. Did they do the right thing because the expected consequences were better than the alternatives, or did they do the wrong thing because the consequences of some alternative would have been better? It is awkward and paradoxical to take the first alternative and say that the policy was "right from their perspective" but "wrong from a later perspective"; it seems better to say that the legislators made the wrong choice – no matter how rational, blameless, or even noble they were in making their mistake. To say this is not to deny that utilitarian appraisals of actions, policies, and institutions always depend on beliefs about what their consequences will be. It is instead to insist merely that appraisals should change as beliefs are corrected in light of later experience. Utilitarianism should be construed as an objective account of what actions or policies are morally right. What is right is whatever, in fact, maximizes happiness.

On this view, whereby the rightness of actions depends on how good their consequences turn out to be, one should judge the goodness of people by examining whether or not they did what appeared to be right on the basis of the knowledge they had, rather than by the rightness of their actions. For example, Milton and Rose Friedman argue that the Federal Reserve's failure to rescue failing banks in 1930 and 1931 and to prevent the money supply from contracting brought on the Great Depression (1980, ch. 3). If they are right, then clearly the Federal Reserve did the wrong thing. But our judgment of the wisdom and character of the decision makers depends on our understanding of their motivation and of what they should have known rather than on any such hindsight.

Third, one needs to be clear about *whose* welfare counts. Should one consider only the welfare of currently living human beings? What about the welfare of those not yet born and of those who might, as the result of one policy or another, not even be conceived? (See Parfit 1984, ch. 17.) Are the interests of future generations adequately represented via the concerns of those who are now living for those who will come after them? What is meant by "adequately"? It seems that issues of fairness, which cannot themselves be justified in utilitarian terms, intrude here. (By the way, these issues concerning the interests of future generations are as worrying for standard welfare economics as they are for utilitarianism.)

In addition to deciding which human interests count, there is also the question of whether one should consider the welfare – or at least the pains and pleasures – of other animals. Most people believe that there is something morally wrong about inflicting suffering on animals "needlessly." Commercial farming of animals for their meat and hides raises moral questions, particularly when animals are confined, frightened, and badly treated. If utilitarians count the pleasures and pains of all sentient beings then they can explain quite naturally why the treatment of animals matters morally, but they will face difficult problems in comparing human and nonhuman welfare (Singer 1975). (Cost–benefit analysis may be in even more trouble determining what a chicken would be willing to pay to roam freely!) Utilitarianism confronts greater difficulties in dealing with issues in environmental ethics concerning the preservation of species or natural beauty.

Fourth, should utilitarians be concerned with total or rather with average happiness? Since the average is the total divided by the size of the population, total and average utilitarianism will always agree when the population size is fixed. But if alternative policies have consequences for how many people there will be, then what is permissible for the total utilitarian can be very different from what is permissible for the average utilitarian. Average utilitarianism seems initially more appealing. A very large population of only middling average happiness might have a larger total utility than a smaller population that has a high average amount of happiness, yet the latter seems more attractive. On the other hand, imagine a state of affairs in which everybody is extremely well-off and the question arises of whether a couple should conceive a child who will have an excellent life and who will not lessen anybody's utility. If this new child's well-being is lower than average, then bringing this new child into existence will be morally impermissible because the child's life, though good, will pull down the average. But how can it be wrong to do something that adds to total well-being and harms no one? Moreover, whether this new child's well-being will raise or lower the average may depend on how well-off people were in the distant past, which seems irrelevant. Average utilitarianism also supports the repugnant thought that it would be a good thing if everyone whose utility is below average could somehow simply disappear. These theoretical issues bear directly on currently pressing problems concerning population control (Parfit 1984, pt. 4; Broome 2004).

Fifth, the utilitarian needs to make clear how utilitarianism is supposed to guide individual action and social policy. Actions and policies are morally obligatory if they maximize utility, but it certainly does not follow that the best method of individual moral decision making or social policy making involves calculating the welfare consequences of actions and policies. The thought that utilitarians might prohibit judges from basing sentences on their estimates of what would maximize utility is an instance of this complication. Attempting to maximize utility is not, in fact, a good method of individual moral decision making. Not only is calculating the consequences of one's actions on the welfare of the whole present (and future?) human (and animal?) population of the world fraught with uncertainties (and not much fun), it is also likely to introduce lots of bias because people tend to take a more favorable view of the overall consequences of actions that benefit them personally. Calculation is itself an action, and not one that is likely to maximize utility. Furthermore, uncoordinated actions by numerous individuals may lead to bad aggregate consequences that could be avoided by requiring people to follow simple rules. People are more likely to perform actions that actually maximize welfare if they do not calculate the welfare consequences of their actions and instead act on general rules, such as "Tell the truth," "Keep your promises," and so forth.

In policy making there is more scope for investigating the welfare consequences of alternatives, but in many circumstances policy makers are also better advised to adhere to rules, for it is difficult to tell what the consequences of policies will be and what impact they will have on total welfare. Consider, for example, the Bush administration policies of imprisoning "enemy combatants" indefinitely without any right to an attorney. Since it is difficult to calculate whether the consequences will be better than respecting international law and the civil liberties of those imprisoned, there is a case for treating such policies as off-limits and simply not to be considered – which is precisely the point of constitutional and international law.

At this point it might appear that utilitarianism is pulling some sort of disappearing trick. Now you see it, now you don't; and traditional moral principles are back in full force. Utilitarianism has a rather nebulous existence in a society where neither individuals nor governments regulate their actions by direct calculation of utility consequences (Williams 1973). But even though it is probably true that the best way for individuals to maximize utility is to stick to many of the traditional moral rules, utilitarianism does not automatically endorse all of traditional morality. It seems plausible that rules against lying maximize utility, but it is not obvious, for example, that moral taboos against paid adoptions ("baby selling") or euthanasia maximize welfare.

Although we are emphasizing that people almost always do better following moral rules and not attempting to estimate the aggregate welfare consequences of alternative actions, we are not endorsing "rule utilitarianism" (Hooker, Mason, and Miller 2000). According to rule utilitarianism, actions are right if and only if they conform to a set of rules such that the consequences of the general adoption of these rules for total or average utility are no worse than the consequences of the general adoption of any other set of rules. Hence rule utilitarianism maintains that people ought to adhere to the utility-maximizing rules even in the unusual case in which they are confident that breaking a rule would result in more total welfare. Since this is at odds with the whole point of morality, as conceived of by the utilitarian, so-called rule utilitarianism does not seem to us to be a version of utilitarianism. (For one possible response to this challenge see Hooker 2001, ch. 4.)

The sixth problem deserves a whole section to itself.

7.2 Interpersonal Comparisons of Well-Being

If policy 1 benefits Ira and harms Jill while policy 2 benefits Jill and harms Ira, then there is no way to say which has the greater welfare benefits unless one can compare how much Ira and Jill are benefited and harmed. Since the costs and burdens of alternative policies usually fall on different individuals, this case is typical. Interpersonal comparisons are necessary in order to make comparisons of total or average welfare, and even nonutilitarian ethical systems require interpersonal comparisons of well-being. To be rationally benevolent, one must be able to judge where one's efforts will do the most good. To treat the interests of different people equally, one must be able to compare the net effects of one's actions on the interests of each.

Interpersonal comparisons are required even though utilitarians are concerned only with total or average utility, not its distribution. Indeed, if some people's wants are modest or some people are not capable of much happiness, then utilitarianism would support giving them less than those who want more or can be made happier. But economists writing near the beginning of this century, such as A. C. Pigou, argued that overall welfare is in fact maximized by equalizing incomes as much as is consistent with retaining incentives to produce. Citing diminishing marginal utility of income, they maintained that an extra hundred dollars contributes less to the well-being of someone with an income of $50,000 than to the well-being of someone with an income of $5,000. Other things being equal, a more equal distribution of income increases total welfare. This argument assumes that one can make interpersonal comparisons of the amount that a hundred dollars contributes to the well-being of different people with different incomes. If interpersonal comparisons cannot be made, then this argument cannot be made either. Though the diminishing marginal utility of income provides a reason to favor more egalitarian distributions, there is nothing in utilitarianism itself that finds inequalities undesirable.

In modern economic theory as developed in the 1930s, economists put aside substantive conceptions of well-being, such as wealth or happiness. Because they found that the basic propositions of demand theory and consumer behavior could be accounted for simply by supposing that people had stable preference rankings with certain properties, most economists took well-being to be the satisfaction of preferences.

If well-being is the satisfaction of preferences, then interpersonal comparisons of well-being are comparisons of the extent to which preferences are satisfied. So-called unit comparisons of utility *differences* are needed to compare the benefits and harms policies might cause to different people. One need only compare how individual utilities change, not their absolute levels. Comparisons of utility levels are also needed if, for example, policy makers want to know who is worst-off. Unlike comparisons of well-being made on the basis of some substantive view of well-being, in this case we are not comparing amounts or intensities of the things or states individuals possess. An interpersonal comparison of preference satisfaction is not a comparison of Jill's and Ira's *feelings* of satisfaction or happiness. A comparison of how much pleasure Jill and Ira feel might be difficult to make because it is hard to measure mental states; but there is no conceptual problem in making the comparison. Interpersonal comparisons of preference satisfaction are not comparisons of mental or physical states but rather of the extent to which the world is as Jill and Ira prefer. Indeed, Ira and Jill may not even know whether some of their preferences are satisfied. An interpersonal utility comparison is a comparison of how well satisfied Ira's and Jill's preferences are, not of how much satisfaction Ira and Jill feel.

What sort of a comparison is this? The main way economists and philosophers have attempted to understand these comparisons is via what Kenneth Arrow (1978) calls "judgments of extended sympathy" (see also Kolm 1972). Suppose we ask people to express preferences not only among ordinary alternatives but also among "extended" alternatives, such as the alternative of being Ira with some option x and being Jill with y. We might then say that Jill-with-y is better-off than Ira-with-x if Jill-with-y is preferred to Ira-with-x.

This way of understanding interpersonal comparisons faces serious difficulties. For example, McPherson might think that Jill is better-off with y because y involves Jill accomplishing something that McPherson admires, while Hausman might think that Ira is better-off with x because Ira is more contented with x than Jill is with y. Whose extended preferences decide whether Ira or Jill is better-off? What is the connection between (on the one hand) whether Ira is better-off with x than Jill is with y and (on the other) McPherson's or Hausman's preferences between the extended alternatives of Ira-with-x and Jill-with-y? John Harsanyi (1977b) suggests that, rather than employing one's own preferences to compare Jill-with-y and Ira-with-x, one compares how well-off one would be with x if one had Ira's preferences to how well-off one would be with y if one had Jill's preferences. (Just how this is supposed to be easier than comparing how well-off Ira is with x and Jill is with y is not obvious.) In Harsanyi's view, there is a single impersonal extended preference ranking to which our empathic abilities are a useful although imperfect guide. Judgments resulting from putting yourself in someone else's shoes in this way can be used to construct interpersonal comparisons. Alfred MacKay (1986) calls this the "mental shoehorn" tactic (see also Griffin 1986, ch. 7).

What could this "universal" extended preference ranking be to which empathy supposedly provides access? It cannot be a ranking in terms of some theory of what makes human life good, because economists disavow any substantive theory in favor of the view that well-being is the satisfaction of preferences. Neither can the ranking be based on comparisons of Jill's and Ira's well-being, for there is nothing to measure except where they stand in their preference ranking. Furthermore, the interpersonal welfare comparison is supposed to be determined by the impartial extended preference and thus cannot itself be the basis for the preference. The ranking cannot be based on a "bare" preference for Jill's state over Ira's, for the mental shoehorn technique that people employ in thinking about what life is like for others depends on their *not* jettisoning all of their preferences and values. For once one has set aside all one's own preferences, upon what could a preference for Jill-with-y over Ira-with-x be based? "If this proposal for comparability uses (basic) preference purged of any particular point of view, it looks like using preference purged of what is needed to make sense of preference" (Griffin 1991, p. 54). What then can the ranking be based on?

In our view, extended preferences cannot provide the basis for interpersonal comparisons. Even if they were unanimous, which is often not the case, they answer the wrong question. A unanimous extended preference for Jill-with-y over Ira-with-x demonstrates that everyone would prefer to be Jill with y rather than Ira with x. But what is at issue is whether Jill's preferences are better satisfied with y than Ira's preferences are with x. The two issues are distinct. Everyone might prefer to be Jill with y – even though Ira's preferences are better satisfied with x – simply because they admire Jill. One might, for example, prefer to be Keats (even with his early death) rather than Queen Victoria yet nevertheless believe that Victoria's preferences were better satisfied than Keats's preferences. If well-being is the satisfaction of preferences, then interpersonal comparisons of well-being must be interpersonal comparisons of the extent to which preferences are satisfied. But (see Hausman 1995) can one make sense of such comparisons? How could they be made?

Most economists take the problem of interpersonal comparisons as a decisive reason to reject utilitarianism. But, as we have seen in the previous section, there are other problems and many different versions of utilitarianism. Yet utilitarianism remains an extremely powerful ethical view, particularly when one is concerned with issues of public policy, because it makes ethical questions in principle matters of straightforward calculation. Such calculations may be difficult to carry out, and in some cases utilitarianism will not give any definite advice owing to problems in learning the effects of actions or policies and in measuring their welfare consequences. But there will be clear cases, and the reasons for indecision and disagreement will be precisely known.

7.3 Justifying Utilitarianism

Utilitarianism is a tempting ethical theory. In many cases it matches common intuitions well; it neatly shows what morality is all about and why it matters; and it has historical connections with economics. In principle, it shows how moral questions may be decided: all it takes is a calculation of the consequences of actions or policies. Evaluations depend on the facts concerning the consequences of alternative policies. Utilitarians can cite the imperfections of human knowledge of consequences to explain why ethical questions are so hard to answer. If actual ethical systems are to a considerable extent implicitly utilitarian, then the utilitarian can offer a plausible explanation for why moral codes differ in different societies. Leaving one's grandparents out to die in the cold may have been morally permissible among those living in the harsh conditions of the Arctic; it may have maximized welfare in those conditions. The same policy is morally impermissible in contemporary New Zealand because it does not maximize welfare there. According to utilitarianism, what is morally right or wrong depends heavily and sensibly on what the facts of the case are. Utilitarianism is “absolutist” in one sense – whatever action maximizes welfare is *the* morally right action to perform – but it is not absolutist in the sense of supporting inflexible formulas for conduct. Whether an action is right or wrong depends on the consequences. Even killing an innocent person might in some circumstances produce more welfare than any alternative action and thus turn out to be morally obligatory.

Even so, as we pointed out in the previous section, utilitarianism gives rise to many complications, ambiguities, and doubts. In the face of such difficulties, can any compelling reasons be given in support of the thesis that the morally right thing to do is to act so as to maximize the sum of people’s utilities? Ultimate questions of justification are difficult for any moral view to answer. But utilitarians do have things to say. In fact, deeper justifications of two kinds can be given, and the different justifications point to different ways of resolving some of the ambiguities noted in the previous section.

The first kind of justification rests on the claim that “utility” is the only intrinsically good thing. If the basic ethical obligation is to promote good states of affairs, then utilitarianism will follow naturally; this is the justification for utilitarianism offered by Mill and Sidgwick. A second approach relies on the notion of equal respect (Griffin 1986, ch. 9). Interpreting equal respect as giving equal weight to everyone’s *interests* leads naturally to utilitarianism. The defense of utilitarianism in terms of equal respect can alternatively take the form of a contractual argument (see Chapter 12). In one version of a contractual approach, ethical principles arise from a willingness on the part of reasoning actors to consider their interests from an impartial point of view. Assume that people have an interest in promoting their own welfare. If they take up an impartial point of view, then they will endorse moral principles that serve individual interests impartially and will thus endorse utilitarianism. This line of thought can be recast as a formal “proof” of utilitarianism – although, of course, only relative to the axioms assumed. (For references and further discussion, see Chapter 13.)

Although these two justifications share a disposition to view one’s own interests from an impersonal point of view, they point toward differing interpretations of utilitarianism. The first argument – in which utilitarianism derives from the attraction of maximizing the good – leads naturally to maximizing the total utility of all sentient beings (whether human or not), whereas equal respect or contractualist arguments point toward maximizing the average utility only of those sentient beings who are rational. As noted in the preceding section, these two outlooks have sharply different implications with regard to population policy.

The two approaches to justifying utilitarianism also point toward different notions of what "utility" is. An "intrinsic good" justification of utilitarianism points away from preference satisfaction notions of utility. It is hard to see why it should be considered objectively good ("from the viewpoint of the universe," as Sidgwick put it) that a state of affairs a person happens to desire should come about whether or not it provides any benefits to that person or anybody else. It seems more comprehensible that one should regard mental states like happiness or the absence of pain as intrinsically good. "Contractualist" views and views that emphasize equal respect turn the spotlight instead on the satisfaction of interests. This emphasis makes it more natural to think of something like the satisfaction of rational preference as the notion of utility worth maximizing.

7.4 Contemporary Consequentialism

During the decades that preceded the 1980s, utilitarianism appeared to be almost dead. Although it continued to influence policy makers, most philosophers did not take it seriously as a moral philosophy. Even economists abandoned utilitarianism in the face of the difficulties posed by interpersonal utility comparisons. And to pound the final nails in the coffin, John Rawls, in *A Theory of Justice* (1971), offered systematic criticism as well as an alternative theory that was suitable for guiding policy. The resurgence in practical moral philosophizing that was so prominent in the 1970s usually took for granted some sort of rights perspective rather than any sort of consequentialism.

Yet by the end of the 1980s utilitarianism and consequentialism were alive again and highly influential in both theoretical and applied moral philosophy. Though the pendulum now seems to be swinging again the other way, consequentialism was resurrected and transformed in many works, mainly in the 1980s (see listing at end of chapter). These more recent consequentialists defend very different ethical theories, with Richard Brandt, for example, developing a sophisticated mental-state variant of utilitarianism and Amartya Sen developing a nonutilitarian view of consequential evaluation in which rights, capabilities, and functionings play a more central part than happiness. For instance, Sen (1979) suggests that consequentialists can regard rights violations as themselves bad consequences in addition to any welfare losses that may accompany them. Despite its diversity, this work shares many of the features of modern economic theory and should be attractive to economists for several reasons.

First, most contemporary consequentialists link ethics to the theory of rationality. John Harsanyi, for example, writes: "Ethics … is a theory of rational behaviour in the service of the common interests of society as a whole" (1977a, p. 43). Most theorists do not go this far; but, as Samuel Scheffler especially has stressed, refusing to make trade-offs among different objectives and to maximize some objective opens one to charges of irrationality.

This link between rationality and consequentialism carries some weight, and the charms of maximizing theories – to which economists have already succumbed – have been seducing philosophers as well. Much of this ethical theorizing is thus more congenial to economists and more easily integrated into normative economics than are, for example, rights-based ethical views (see Chapter 10). The consequentialist theories defended by contemporary philosophers are, to be sure, not easily operationalized, but the difficulties of putting them to work are not insuperable, either. Some of this work, such as Harsanyi's, would in some cases justify returning to old-fashioned utilitarian policy analysis.

Second, almost all contemporary utilitarians join economists in taking utility as an index of preference satisfaction rather than as an object of preference. But philosophers do not join economists in regarding utility as an index of the extent to which *actual* preferences are satisfied. This refusal to identify welfare with the satisfaction of actual preferences sets moral theorists apart from welfare economists: Focusing on rational or informed preferences leads away from a person's actual wants and toward what is, in some substantive sense, good for the individual. Yet these theories remain closer to normative economics than theories that emphasize rights or liberty, and they provide more evident opportunities for building bridges. It might, for example, be possible to operationalize some of the conditions that moral theorists have placed on "rational" and "well-informed" preferences and thereby to strengthen the moral basis for the conclusions of welfare economics.

Third, some of the work during the past generation on utilitarianism and consequentialism has been influenced by developments in economics and game theory, and more of it arguably should be. In Russell Hardin's utilitarianism, for example, human ignorance of consequences and the difficulties of measuring and comparing utilities hold center stage, and concepts from game theory are put to work generating the outlines of a utilitarian theory of property rights and their limits (Hardin 1988). Hardin's game theory is elementary and his derivations are informal, and there exist further opportunities for collaboration between economists and philosophers in developing institutional implications of utilitarianism in circumstances of scarcity, uncertainty, opportunism, and so forth. We shall have more to say about these opportunities in Chapter 14.

Of particular interest has been the development of consequentialist moral theories that are not utilitarian – theories for which the good consequences to be maximized are things *other* than either happiness or the satisfaction of preferences (see especially Parfit 1984). It is difficult to find empirical proxies for some of these good consequences, but the more objective the goods to be maximized, the more opportunity (other things being equal) for empirical application. If philosophers can specify a well-defined and clearly measurable good to be achieved, then the welfare economist can step in and discuss how best to achieve it.

One important objective good in many of these theories is the satisfaction of needs; even utilitarians emphasize it, though not as a fundamental and intrinsic good. For example, although James Griffin is sympathetic to informed preference utilitarianism, he argues that policy should focus on needs because a government can more easily tell what people need than what will satisfy their informed preferences (1986, ch. 3). Griffin's emphasis on the empirical tractability of needs is ironic, given how averse economists have been to distinguishing needs from mere preferences. Yet this aversion has arisen not from empirical difficulties but rather from theoretical objections to drawing the distinction. Here is a case where moral philosophers may be more practical than economists! In political discussions of economic policy, concern about human needs is already ubiquitous, and if philosophers can provide both a rationale for taking needs seriously in social decision making and a principled way of drawing the distinction between needs and wants (Braybrooke 1987; Thomson 1987), then economists can put their modeling tools to work to help devise policies that will satisfy needs. A good example may be in the area of medical research, where difficult choices must always be made about how much to invest in searching for treatments for various diseases. Rational allocation of scarce resources across different fields of research requires assessing the costs of research, the probabilities of success, and the relative urgency of progress in different fields. It is plausible that philosophers can contribute to making such choices well by developing defensible judgments about the value of health and the relative importance of the needs that would be met by successful treatment of different diseases.

The deep problems of utilitarianism – in particular, those concerning interpersonal comparisons of well-being – do not preclude operationalizing utilitarianism by providing a specification of the utility function to be maximized. One could, for example, stipulate a single utility function that roughly represents everyone's preferences. Preferences represented by such a common utility function would most plausibly be defined not over marketed goods and services but instead over more fundamental goods that the agent extracts or constructs from marketed goods and services (Lancaster 1966; Becker 1976). Behind the large differences in people's manifest preferences there might be agreement in preferences among the underlying goods. For example, Jill may prefer to eat at home while Jack prefers restaurants. The differences in their preferences may result more from differences in the opportunity costs they face in producing tasty and nourishing meals at home than from any differences in underlying preferences. It is not absurd to postulate (as an approximation) a common utility function, and in terms of that function there might then be little problem determining the total utility of alternative policies – apart from the general difficulties of predicting their consequences. The hard part might not be making the operationalization of utilitarianism workable, but justifying it. What basis could one have for regarding the single utility function as representing, however crudely, nearly everybody's preferences and as capturing their relative strength? Can the idea of a shared "deep" utility function underlying apparent differences in preferences be rendered both plausible and practically workable?

7.5 Is Utilitarianism Plausible?

The most powerful objection to utilitarianism is that in some cases it clashes sharply with our moral intuitions. Suppose, for example, that false testimony by one of the witnesses to the Rodney King beating in Los Angeles could have ensured a conviction at the first trial of the officers accused of police brutality. Given the grievous consequences of the acquittal – the massive Los Angeles riots of 1992 – it seems that a utilitarian would judge that the witness ought to have committed perjury. Yet most people believe that perjury is morally impermissible.

There are three possible responses to this apparent conflict between utilitarianism and what most people believe is right and wrong. First, utilitarians can argue that the conflict is only apparent. John Rawls, in his celebrated essay entitled "Two Concepts of Rules" (1955 – written before he developed his own theory of justice, which is discussed in Chapter 12), distinguished questions about the design of institutions from questions about enforcement of their norms. Even if perjury would maximize utility, perjury is forbidden by law and custom. New laws and customs – permitting perjury whenever witnesses conscientiously believe that perjuring themselves would maximize utility – would not be favored by a utilitarian or anybody else. So perjury should be forbidden by law and custom, even in a case such as this one where committing perjury might in fact maximize utility. Rawls is not defending rule utilitarianism here. He establishes only the weaker conclusion that perjury should be illegal and socially proscribed, and his argument cannot rule out the possibility that it might still be morally obligatory for witnesses (secretively) to break the law.

The utilitarian has other grounds to question whether in such cases utilitarianism does recommend morally repugnant actions such as perjury. Hardin (1988) stresses that hypothetical examples such as this usually presuppose knowledge that is unattainable. The witness cannot be sure that the perjured testimony will lead to a conviction or that the perjury will not be detected. So the best thing to do is to adhere to the rule against perjury. When one takes into account the unavoidable uncertainties, it turns out that utilitarianism does not in fact recommend perjury. In response, many people would say that whether perjury is morally right or wrong in this case does not depend on whether it would be detected or on whether it would ensure a conviction. The intuitive objection has not been completely answered.

A second response to the apparent conflict between utilitarianism and moral intuition is to challenge the authority of "intuition." A utilitarian would want people to be educated to have strong moral intuitions that promote desirable conduct in typical situations. It will be good on the whole if people are strongly moved by such feelings, but their intuitions have no independent evidential force that can help in resolving hard cases (see Hare 1981). Even if the morality most people were taught were devised by a utilitarian, well-brought-up people should feel a strong repugnance toward perjury and should have the moral intuition that perjury is wrong.

In reply, the critic of utilitarianism can question what basis for morality there could be *apart* from intuition. If people can't take their intuitions seriously (though not uncritically), then they have no foundation upon which to argue for or against moral principles. So one arrives at a third response to the apparent conflict between intuition and utilitarianism, which is to reject utilitarianism. Utilitarianism remains unpalatable to many people, but its difficulties are not (in our view) obviously greater than those of alternative ethical theories.

7.6 Consequentialism and Deontology

It might be thought that the general structure of evaluation which we have called consequentialism is not subject to similar intuitive objections. Because consequentialism per se does not specify what counts as good, one might think that all ethical theories could be regarded as consequentialist. For example, those moralities that stress freedom rather than welfare might be regarded in consequentialist terms as seeking to maximize freedom, while those that stress duties might mandate maximizing conformity with duty. However, consequentialism cannot in this way encompass all moral theories unless there is some "agent neutral" way to rank states of affairs with respect to goodness. For example, every consequentialist view that counts intentional killings as bad will imply that, other things being equal, it is obligatory to kill one person if one can thereby prevent two intentional killings. Such a conclusion is drastically at variance with most people's moral views. Everyday morality incorporates "agent relative" reasons or values. The fifth commandment does not say, "Minimize killings." It says, "Thou shalt not kill."

Moral systems like the Ten Commandments are called "deontological." In Scheffler's terminology, deontological (nonconsequentialist) ethical theories employ both agent-centered prerogatives (they sometimes *permit* agents to act in a way that does not maximize the good) and agent-centered constraints (they sometimes *prohibit* agents from acting so as to maximize the good). Agent-centered prerogatives and constraints are puzzling. How can it be morally permissible, let alone morally obligatory, to make intentionally a choice that results in less good? Deontological theories not only conflict with consequentialism, they appear to conflict with rationality itself.

One reaction is to conclude that deontological moral systems are irrational. If one is serious about regarding killing as wrong and if one is convinced that murdering one innocent person will prevent the murder of two others, then only irrational squeamishness should prevent a rational individual from committing the single murder. But refusing to murder an innocent person does not appear to be mere squeamishness. It seems to be the right thing to do. The consequentialist might argue that appearances are misleading because our intuitions are geared to common, real-world circumstances, not to philosophers' hypothetical examples. Cases in which one can certainly prevent two murders by committing one almost never occur, and it is a good thing that we're unwilling to commit murder even when it seems we could bring about better results by doing so.

This claim on behalf of the consequentialist is questionable. Conflicts between deontological constraints and rationally minimizing harms do in fact arise with some regularity in everyday life. Consider, for example, the case of Baby Jessica, a toddler who in 1987 fell into a well in Texas. After a massive human effort costing hundreds of thousands of dollars, she was rescued. The resources that were used to rescue her could have prevented the death and injury of hundreds of other American children if they had been devoted instead to better prenatal care. They could have saved the lives of thousands of malnourished infants in less developed countries. The distribution of health-care resources is shot through with similar examples. Consequentialism and rational benevolence imply that resources should go where they do the most good. The consequentialist would maintain that acting otherwise is irrational.

It is possible to bring about a formal reconciliation between the theory of rationality and deontological views forbidding murder or requiring we do whatever possible to rescue those in danger. If one distinguishes between *x being murdered* and *my murdering x*, then there is no inconsistency in preferring *x* being murdered to two others being murdered and at the same time preferring two others being murdered to my murdering *x*. Similarly, there is no inconsistency in preferring the death of a single child, Baby Jessica, to the death of hundreds yet preferring the death of hundreds of children to *my* standing by and letting Baby Jessica die. In order to make such reconciliations more than ad hoc gimmicks, one must say something about why it is "reasonable" not to be indifferent between *x* being murdered (by somebody or other) and my murdering *x* or between Baby Jessica dying and my not saving her. Here again we can see that utility theory presupposes a substantive background theory of "rational requirements of indifference" (cf. Section 4.3). But there is a new wrinkle. The background theory, which here determines when it is rational to distinguish among actions and outcomes, depends on substantive moral principles.

The proposed reconciliation allows evaluation of the goodness or badness of a state of affairs (or set of consequences) to vary with the perspective of the person doing the evaluating (Sen 1982b). An onlooker might readily conclude that it is morally better for one person to be murdered rather than for two to be murdered or for one child to perish rather than hundreds. But from the standpoint of the decision maker, what is at stake is whether *I* commit a murder or whether *I* stand by and let a child die. One may act on one's own moral commitments and judge that, despite the good consequences of performing the murder or letting the child die, the act is wrong. Everyone might agree both that two murders are worse than one and that, from the point of view of an agent facing the prospect of committing a murder, carrying out the murder is morally worse than failing to prevent two murders from happening. Although what is right or wrong on such an account depends on the consequences, the account is not "consequentialist" and is indeed ruled out by consequentialism, which presupposes an agent-neutral evaluation of states of affairs.

It is worth underscoring that deontological constraints need not be absolute. Whereas it might be morally wrong to kill one person to save two lives, it might be morally obligatory to shoot down a hijacked airliner heading for a nuclear power plant. (Conversely, it might be hard to find someone so thoroughly consequentialist as to approve of killing a person in order to save one life and to relieve one headache.)

It might seem at first glance that evaluator relativity is less likely to be a consideration in public policy than in personal morality, because policy should be made from an impersonal point of view. But as the Baby Jessica case and other examples from health-care policy show, it is not obvious that policy should attempt simply to override moral concerns to help or not to harm. If Americans were consequentialists and did not rescue the likes of Baby Jessica – but also did not look the other way when hundreds of thousands of other children needlessly died – then the United States would in important ways be a better society. But if Americans could ignore Baby Jessica's plight with no more than a self-congratulatory recognition of how rational they were, who knows how inhumane they might become?

7.7 Conclusion: Should Economists Embrace Utilitarianism?

Welfare economists ought to find utilitarianism attractive because it is so similar to standard welfare economics. Just define the rough-and-ready utility functions that will represent the preferences of the individuals affected and stipulate a way of making the interpersonal comparisons, and the way to utilitarian policy analysis is open. Welfare economists would not have to give up their focus on outcomes or their view of welfare. Of course, it is a bold and controversial step to stipulate what the utility functions should be and how the utilities of different people should be compared, but this step is no bolder than the identification of welfare with the satisfaction of preferences. Taking this further step would enrich normative economics.

Yet many of those concerned with policy would take issue with the suggestion that normative economists should become more explicitly utilitarian, precisely because they would like to see economists make a more radical break with current practice. Before considering this critique, let us turn in the next two chapters to a more detailed treatment of current practice.

Suggestions for Further Reading

Classic writings on utilitarianism include Bentham (1789), Mill (1863), and Sidgwick (1901). The collection of essays by Plamenatz (1967) provides a number of helpful philosophical treatments of problems of utilitarianism.

An illuminating exchange on utilitarianism by two contemporary philosophers is Smart and Williams (1973). A helpful volume with emphasis on economic aspects is Sen and Williams (1982).

The most important consequentialist works are those of Brandt (1979); Broome (1991b, 2004); Griffin (1986, 1996); Hardin (1988); Hare (1981); Harsanyi (1977a); Kagan (1989, 1997); Parfit (1984); Railton (1984); Sen (1982b); and Singer (1979); see also Smart's essay in Smart and Williams (1973). Brandt, Hare, Harsanyi, Singer, and Smart are clearly utilitarians, while the others reject theories of human well-being that are characteristic of utilitarianism. For discussions of rule utilitarianism see Hooker et al. (2000) and Hooker (2001).

Applications of consequentialist moral philosophy to practical issues abound. See particularly Glover (1990) and Singer (1986, 2002). For discussion of the rationality of deontological moral principles that permit options not to do what maximizes the good or forbid maximizing the good, see Nagel (1986, ch. 9); Scheffler (1982, 1988); Kagan (1989); and Kamm (1992).

EIGHT

Welfare

When people in modern Western cultures think about morality, they first think about what is morally permissible or impermissible, right or wrong. But there are other matters of moral concern. Of special importance among these are questions about what things are good or bad and, more specifically, about what things are good or bad for people.

Exactly what is good for a particular agent, Murphy, will depend on Murphy's character, ability, and circumstances, and what is good for Murphy may be very different from what is good for Marlow. But most of the differences between what is good for Murphy and Marlow concern *instrumental* goods – things that are good because they are means to something else. If one focuses on *intrinsic* goods – things that are good in themselves – without regard to their consequences, then there may be much less variation from individual to individual. Size-7 shoes are good for Murphy while size-12 shoes are good for Marlow, but both pairs serve the same end. Instrumental goods, like shoes of the right size, are good only if the ends they serve are good. So if nothing were good in itself, then nothing could be good as a means to some other end. There must be intrinsic goods in order for there to be instrumental goods. A central problem of moral philosophy has been to determine what things are intrinsically good for human beings. Thus Aristotle, for example, held that happiness was the sole intrinsic good. We disagree. We believe that many things are intrinsically good for people.

All plausible moral views assign an important place to conceptions of individual good, utility, welfare, or well-being. This is obviously true of utilitarianism, but even nonutilitarian views that emphasize notions of rights, fairness, and justice need a conception of human well-being. Not only do these views recognize the virtue of benevolence, which requires some notion of human good, but even their core notions often make reference to well-being. For example, justice or fairness is understood in terms of treating the interests of different persons properly, and acting rightly will often involve avoiding harm to other individuals. Notions of harm and interest are plainly connected to notions of well-being.

8.1 Theories of Well-Being

In chatting with one's neighbors, as in studying moral philosophy, one finds many different theories of well-being. In some religious views, the ultimate good lies in a relationship with God, while in others such a relationship with God is good because of the eternal happiness it brings. Many people believe that only mental states are intrinsically good, but there is less agreement here than it seems because there are so many different views concerning which mental states are intrinsically good. Jeremy Bentham holds that the good is pleasure, while John Stuart Mill holds that it is a diverse set of mental states he calls "happiness." Mystics find the good in contemplative states of mind. Henry Sidgwick argued for the hybrid view that the good is any mental state that is intrinsically desirable.

There are also many philosophers who deny that well-being is a matter of mental states. Friedrich Nietzsche regarded great achievements as the ultimate goods. Others endorse as intrinsic goods a whole potpourri ranging from health and intimate personal relationships to achievements such as those that Nietzsche admired. The theory of well-being is a messy area of philosophy. It is difficult even to categorize the various theories, and they all face serious difficulties. All of this is enough to send economists running back to their graphs. But they cannot avoid addressing these philosophical problems if they want to be able to judge when welfare increases or decreases.

Theories of well-being or welfare can be classified as either "formal" or "substantive." A substantive theory of well-being says what things are intrinsically good for people. "Hedonism" is an example of a substantive theory of well-being; it says that well-being is happiness or pleasure. "Formal" theories of well-being specify how one finds out what things are intrinsically good for people, but they do not say what those things are. To maintain that welfare is the satisfaction of preferences is to offer a formal theory of well-being. This theory does not say what things are good for individuals. Instead it says how to find out: by seeing what people prefer. Formal theories may be compatible with substantive theories. For example, if happiness is the ultimate object of preference, then it could be true both that well-being is the satisfaction of preference and that well-being is happiness.

Most economists do not directly address questions concerning welfare, and we have no systematic evidence concerning their views. Among those economists who do talk about welfare, most are attracted to a formal theory, no doubt because formal theories appear to involve less controversial philosophical commitments than substantive theories about what is good or bad for people. By leaving the substantive question of what is good for an individual up to the individual, normative economists aim to show their philosophical modesty. The preference satisfaction view of well-being also appeals to the antipaternalist values of many economists. But, as we shall argue in this chapter, formal theories of the good (such as the preference satisfaction theory) do not avoid philosophical controversy.

As discussed in Chapter 5, given mainstream economists' commitments to utility theory in explaining human choices, it is natural that they would look to levels of utility – that is, preference satisfaction – as the fundamental measure of human well-being for evaluative purposes as well. If individuals are exclusively self-interested, then they will prefer x to y if and only if they believe that x is better for them than y is. If they are perfectly well-informed then their beliefs will be true, and x is better for them than y if and only if they prefer x to y. Hence it is tempting to take well-being to be the satisfaction of preferences. In applied work, economists often rely on more objective measures of "real income" rather than utility measures, but most economists view this as a compromise with data limitations. They regard real income as an imperfect proxy for preference satisfaction.

However, there are serious objections to a preference satisfaction view of well-being. One difficulty is that, unlike the purely self-interested creatures who populate much economic theory, people care about more than their own well-being. They are sometimes altruistic and sometimes malevolent. People sometimes sacrifice their own well-being in order to benefit others or to do harm to those they hate. Almost no one's welfare is affected by whether the continuum hypothesis turns out to be true, but it is easy to see how people could have a preference about that. Even more clearly, many people prefer that their unborn great great grandchildren have good lives, even though that will have no bearing on their own well-being.

A second objection arises from the fact that people are also ignorant of many things. Thus people may prefer something that is bad for them because they mistakenly believe it is beneficial. It is not true that x is better for A than y if and only if A prefers x to y. Why then do most normative economists take welfare to be the satisfaction of preferences?

8.2 Is the Standard View of Welfare Plausible?

These objections are not hard to see. Giving a powerful motorcycle to an irresponsible teenage boy does not necessarily make him better-off, no matter how desperately he wants it. Yet many economists continue to espouse the preference satisfaction theory of welfare. Why?

We explored one reason in Chapter 5: If one accepts the standard view of rationality and the standard idealizations of positive economics, then what one prefers must coincide with what is good for one. Within the theoretical world depicted in standard economic models, welfare is preference satisfaction. Economists recognize that this world is not the real world, and the fact that welfare is preference satisfaction in standard models does not imply that welfare is preference satisfaction in real life, but economists often regard the differences between theory and reality as mere noise. "If one looks past the complications of actual life to the central realities captured in standard economic models, one can see that welfare is in essence the satisfaction of preferences." This line of thought will not convince anyone who does not see the world the way mainstream economists do, but it helps explain why these economists are content to identify well-being with the satisfaction of preferences.

Second, even though what satisfies Smith's preferences does not necessarily make Smith better-off, Smith's preferences may be the best *guide* to what is beneficial to her. What better way is there to benefit people than to satisfy their preferences? Legislators know less of Smith's circumstances than she knows and they have a less tender concern for her well-being than she does. The judgments of legislators about how to make Smith better-off are likely to be worse than her own judgments. It is also safer to rely on people's preferences. The mistakes individuals make about their own good will be unsystematic, and – whether mistaken or not – if the choice is theirs then legislators cannot arrogate to themselves the power to substitute their judgment for persons' own judgments. This defense does not attempt to show that welfare *is* the satisfaction of preferences. Rather, it denies that normative economists need any philosophical theory of welfare. Regardless of what human well-being truly is, the best measure of well-being is the extent to which preferences are satisfied. However, there is no way to defend the claim that preference satisfaction is the best measure of welfare until one has said what welfare is.

We guess that many economists do not take the standard definition of welfare literally. They believe instead that welfare is a desirable mental state, such as happiness. They find it plausible to believe that the best measure of well-being is preference satisfaction, because they also believe that the best way to implement a social policy that aims to make people happy is to satisfy preferences. We cannot prove our conjecture that economists are often inclined to take welfare to be a mental state like happiness, but their use of language in talking about welfare is quite suggestive. Normative economists often slide from talking about utility to talking about happiness. They often talk about individuals "seeking" utility, which makes no sense if utility is just a measure of the extent to which preferences are satisfied. And it is easy to equivocate on the word "satisfaction." To satisfy a preference is for a preference to come true. When one knows that one's preferences have been satisfied, one may feel satisfied. But there is only this contingent connection between the satisfaction of a preference and the satisfaction of a person. The satisfaction of preferences is like the satisfaction of degree requirements. It has no necessary connection to any *feelings* of satisfaction. Ellen's grandmother's preference that her granddaughter become a doctor is satisfied if Ellen becomes a doctor, even if Ellen's grandmother never lives to see that day and cannot feel any satisfaction at the event. Welfare economists nevertheless often speak of satisfaction as a feeling.

It is questionable whether well-being consists of mental states. Suppose there were an "experience machine" that could give people the highest quality experiences possible (Nozick 1974, p. 41). These terrific experiences might be intense physical pleasures or they might be experiences of climbing Everest or composing a symphony that surpasses Beethoven's best efforts. (Once attached to the machine, people no longer know that they are attached; the machine takes over their consciousness and memories completely.) Suppose people attached to this machine experience whatever states of consciousness that mental-state theorists of well-being claim are ultimately and intrinsically good. The mental-state theorist would then have to say that people would be better-off permanently hooked up to a reliable experience machine than living their own lives and experiencing the decidedly mixed mental states that come with them. If one believes that those who are hooked up to the experience machine are missing out on some of the intrinsically good things in life – even though they are not, by assumption, missing out on the best mental states – then one cannot accept a mental-state view of well-being.

8.3 Equating Well-Being and Preference Satisfaction

Even if the theoretical and practical defenses of measuring welfare by the satisfaction of preferences do not succeed, they permit one to see why sensible normative economists so readily link welfare and the satisfaction of preferences. Regardless of its philosophical difficulties, preference satisfaction views of well-being dominate orthodox welfare economics. So let us now look more carefully at the consequences of a preference satisfaction view of welfare for economic evaluation.

Measurements of well-being have many purposes: to establish benchmarks for equality, to assess progress over time, to compare living standards across communities, to weigh the comparative claims that different persons may justifiably make on social resources, and so on. Different conceptions may suit different purposes. We shall argue in this section that the view of welfare as the satisfaction of preference is unsuitable for most of these purposes.

8.3.1 Changing and Conflicting Preferences and Preferences Based on False Beliefs

To regard welfare as the satisfaction of preferences leads to complications when preferences change. As Brandt (1979, ch. 13) and Parfit (1984, ch. 8) point out, if an individual's preference ranking changes then it is unclear whether the individual is made better-off by satisfying or frustrating the original preferences. Perhaps one faces within an agent's life something akin to the problem of interpersonal utility comparisons (Gibbard 1986).

Should one care about satisfying preferences that a person no longer has? Is there any reason now to satisfy Hausman's childhood desire to be a garbage collector? At this stage in his career, Hausman thinks not. Why not? Why should his current desires take priority over his past desires? One answer might be that his current desires not to be a garbage collector are stronger. That answer can't be right, because his current desires are probably not stronger. As a 6-year-old he longed intensely to drive a garbage truck. On further reflection, it seems irrelevant which desire is stronger. A second and more plausible explanation for why we give no weight to satisfying preferences that people no longer have is that giving people what they no longer want brings them no "satisfaction." But in giving this reason one has shifted from a preference satisfaction theory to a mental-state theory of welfare. Can someone who holds a preference satisfaction theory of well-being explain why the satisfaction of preferences that people no longer hold is of no moral importance?

This theoretical problem of preference change is linked to a practical problem. Suppose, as seems plausible, that policies and institutions have systematic effects on preference rankings. (It is hard to believe that the billions spent on propaganda and advertising have *no* effect on preferences.) Assessments of policy must then depend in part on one's views concerning just which preferences to promote or concerning which institutions provide a suitable framework for developing desirable preferences (McPherson 1982, 1983b). Should one be concerned about satisfying current preferences, when one judges that they are likely to change? Should one aim to modify preferences so that they will be easier to satisfy? How should one choose between either satisfying existing preferences or modifying preferences first and then satisfying the modified preferences? A preference satisfaction theory of well-being makes some of these questions difficult to answer and suggests implausible answers to others. In particular, a preference satisfaction view of well-being apparently implies that (other things being equal) we should educate people to have easily satisfied preferences.

Problems also arise when agents have false beliefs. In the simple folk psychology of choice sketched at the beginning of Chapter 4, actions depend on preferences and beliefs. But preferences and beliefs are not independent of one another. Preferences combine with beliefs to generate further preferences. Adapting some terminology from Sen (1970a, ch. 5), let us distinguish "basic" preferences, which are independent of beliefs, from nonbasic preferences, which depend on beliefs. Few preferences are basic. Someone's preference for avoiding pain might be an example. Most preferences are nonbasic, and if the beliefs upon which they are based are false then nonbasic preferences may radically misrepresent the underlying concerns of those who hold them.

For example, if citizens of New York State prefer to locate a nuclear power plant in the middle of New York City because they seek security and falsely believe that long-distance power transmission poses greater risks than do nuclear power plant accidents, then frustrating their preferences would be more likely to serve their underlying concern with security and to make them better-off than satisfying their preferences. Whether or not policy makers should respect the underlying values of the population, they should not necessarily be governed by their beliefs. This leaves the policy maker in a bind, because preferences virtually always depend upon beliefs. The only consistent ways out of this impasse are either to follow people's preferences, even if they depend upon beliefs that are false and unreasonable, or to employ some substantive theory of welfare (Hausman and McPherson 1994).

Conflicts within an agent's preferences create problems, too. The standard model of rationality rules out conflicts among preferences, but one cannot rule realities out of the real world, and such conflicts obviously exist. Some students have been known to want to ace an examination and also to party the night before, even though they know that the partying probably rules out the A. In addition, there are conflicts between first-order preferences, such as a preference for smoking, and second-order preferences (or "metapreferences"), such as a preference that one not prefer to smoke. Such conflicts can give rise to internal struggle, feelings of regret and deprivation, and apparently inconsistent behavior, such as purchasing cigarettes but then locking them away or flushing them down a toilet – and then running out to purchase more. A preference satisfaction view of well-being provides few resources to deal with such conflicts. If Jones does not have a consistent preference ranking, then there is no way to say whether anything is better for Jones than anything else. If Jones has two or more preference rankings that conflict with one another, then there is no way to say whether x is better for Jones than y unless x happens to be above y in all of his rankings. Without privileging one of the rankings, one cannot say which ranking would be best for Jones to keep and which would be best to drop. Indeed, one cannot even make sense of the notion of how good it is for Jones to have one preference ranking rather than another without invoking some preference ranking.

Although problems due to preference changes and conflicts among preferences may be relatively isolated, the dependence of preferences on beliefs looms large in many policy problems. All three of these difficulties are both theoretical and practical, and a preference satisfaction view of well-being cannot cope with them very well.

8.3.2 Assessing Preferences

On the preference satisfaction view, one makes Smith better-off by satisfying her preferences, regardless of how idiosyncratic or obnoxious they are and regardless of how they were formed. But some of the uses of the theory of welfare apparently demand that one discriminate among preferences.

For example, as Thomas Scanlon has pointed out, policy makers and public discourse rely on a relatively objective standard of "urgency" when weighing the strength of claims for social provision. The urgency of people's claims does not correspond to the strength of their preferences. For example, even if members of a destitute group prefer subsidies to build religious monuments over receiving food and shelter, their beneficent fellow human beings (whether fellow citizens or foreign donors) might acknowledge a moral obligation only to provide food and shelter (Scanlon 1975; see also Sagoff 1986). Without pretending to settle all questions about what makes people better-off or worse-off, and without taking any stand concerning what things are of ultimate value, there is settled agreement on some of what beneficence demands. Concern for one another demands that we address starvation and homelessness. The fact that building a religious monument matters more to some people is neither here nor there. *Their* preferences do not dictate *our* obligations. Here one can see, by the way, one reason for preferring "in-kind" provision to transfer payments. If donors provide cash benefits, destitute members of this group will not use them to alleviate the needs that gave rise to our obligation to assist them. Those who identify welfare and preference satisfaction might claim that our resistance to honoring idiosyncratic preferences is purely practical: we would open ourselves to all kinds of manipulation and misrepresentation if we let people's subjective reports of their needs govern public distribution of benefits. On the other hand, one could object that it is paternalistic not to respect the worshipper's preferences. But paternalism and antipaternalism are views about what it is permissible to *do*; they are not views about what individual welfare *is*.

Should public benevolence or justice be sensitive to what people prefer? What moral pull should satisfying Hausman's preferences have on McPherson? Thomas Nagel has argued that if something is valuable to people only because they want it, then their getting it has no direct moral importance for others (1986, ch. 9). Others have no reason to satisfy Ellen's preferences unless they can make sense of why what she wants is worth wanting, or why her life will be better in some substantive way if she gets what she wants. For example, many Boston residents desperately wanted the Red Sox to win the World Series in 2004. Their happiness when the Red Sox won gave others, even Yankees fans, some reason to want their preferences to be satisfied. But Nagel maintains – very plausibly, we believe – that the mere preferences of Red Sox fans (as opposed to their happiness or unhappiness) should be of no moral importance to others. This line of thought implies that social policy should not be concerned with satisfying preferences except insofar as doing so coincides with other objectives.

Consider those who have expensive tastes. An even-handed concern to satisfy their preferences appears to be unfair. If welfare is preference satisfaction, then a person who has cultivated a taste for "prephylloxera claret and plover's eggs" (Arrow 1973, p. 254) without an income that makes them affordable is worse-off than someone with a similar income who wants only affordable beans and franks (see also Dworkin 1981a). But should social policy be responsive to expensive preferences? Defenders of a preference satisfaction view may argue that what is at issue is fairness rather than well-being: the person with expensive preferences actually is worse-off without the plover's eggs, but policy makers don't have to care. But do the unsatisfied preferences of those with expensive tastes count at all? Perhaps their unhappiness counts – though that is controversial, too – but is their inability to procure the expensive things they want itself a matter of social concern? Defenders of more objective views of well-being can maintain that, except insofar as they are unhappy, the fancy eaters are not at all worse-off; there is no need to explain away the policy maker's indifference to their "plight." If one refuses to get all choked up about failures to satisfy expensive preferences, then either one's benevolence is limited or one does not accept a preference satisfaction view of welfare.

Racist, sadistic, and other antisocial preferences raise related problems. One reason why it is bad to satisfy them is that doing so frustrates other (and characteristically stronger) preferences. But should this be balanced against the "benefits" of satisfying them? Should antisocial preferences count at all (Harsanyi 1977a, p. 56)? Given either a hedonistic or a preference satisfaction view of well-being, it seems that antisocial preferences ought to count and that a benevolent person should, other things being equal, strive to satisfy them. More objective views of well-being, on the other hand, could permit one to deny that individuals are made better-off when their antisocial preferences are satisfied.

Defenders of a preference satisfaction view of welfare have some ways of deflecting the claims of those with antisocial and expensive preferences without questioning the identification of well-being and the satisfaction of preferences. First, they can claim that some moral consideration distinct from benevolence, such as justice, is involved. Second, they can note that preferences are malleable and that expressions of preferences respond to incentives. Frustrating expensive and antisocial preferences discourages their expression and might contribute to changing them. Thus, frustrating such preferences would in the long run enable more preferences to be satisfied. These points are well taken, and they provide additional reasons not to satisfy expensive and antisocial preferences regardless of whether one identifies welfare with the satisfaction of preferences. But they do little to defend preference satisfaction as a *measure* of well-being.

A final difficulty concerns preference formation. People may have particular preferences because of previous coercion or manipulation, and they may come to prefer things as the result of problematic psychological mechanisms. Some people, for example, want things precisely because they cannot have them ("the grass is always greener on the other side of the fence") while others spurn what is beyond their reach, like the fox who judged the unobtainable grapes to be sour (Elster 1983; Sen 1987b, 1990). For example, women who are systematically denied roles in public life or equal shares of consumption goods may learn not to want these things. Women who have been systematically oppressed may not have strong preferences for individual liberties, for the same wages that men earn, or even for protection from domestic violence. But liberties, high wages, and protection from domestic violence may make them better-off than giving them what they prefer. It seems that those who are benevolent need to consider not only preferences but also the origins or the justifiability of preferences. Satisfying preferences that result from coercion, manipulation, or "perverse" preference formation mechanisms may not make people better-off. These suspect preferences sometimes reflect false beliefs, and some of the problems here are versions of problems we have already discussed. But coercion and manipulation may also distort the psychological mechanisms that lead people to value some things more than others. When the preferences of oppressed people derive from their oppression, one cannot measure their welfare by considering how well their preferences are satisfied.

8.4 Modifying the Preference Satisfaction View

As we have seen, the preference satisfaction view has many awkward implications. It suggests that expensive and antisocial preferences count. It suggests that one should not look behind preferences that result from manipulation or perverse psychological mechanisms. It flies in the face of shared views of the urgency of objective needs. And the preference satisfaction view leads to serious puzzles in cases of preferences based on false beliefs, changes in preferences, and conflicts between preferences.

One reaction to these difficulties is to reject the preference satisfaction view and endorse some other theory of welfare. In the next section, we will consider two plausible candidates. But there is another alternative. Rather than maintaining that well-being is the satisfaction of *actual* preferences, one can maintain that well-being is the satisfaction of self-interested preferences that are "rational" or "well-informed" (Gauthier 1986, ch. 2; Goodin 1986; Griffin 1986). This modified theory of well-being is still a formal theory. It does not say what a rational or well-informed person prefers; it says that what is good is whatever people rationally prefer for themselves. But views like these provide a foot in the door for arguments concerning substantive goods, goods that it is rational to prefer. To take well-being as the satisfaction of informed rather than actual preferences shifts the emphasis from what people prefer to what is rational to prefer. If it could be shown, for example, that it is rational for everyone to prefer happiness to unhappiness or virtue to vice, then a view of well-being as the satisfaction of rational preferences would lead to a substantive view of well-being as involving happiness or virtue.

It is more plausible to maintain that well-being is the satisfaction of suitably "laundered" self-interested preferences than to maintain that it is the satisfaction of actual preferences. This view grants that people may prefer something that is bad for them because of ignorance or because of a concern to benefit or harm others. Such preferences would not be rational or self-interested, and so satisfying them would not make the person better-off. Furthermore, satisfying preferences that result from coercion and manipulation no longer counts as benefiting a person if, but for ignorance or irrationality, the person would not have had such preferences. When actual preferences change, there need be no puzzle about how to make people better-off if their rational preferences have not changed. One might also maintain (although this is more controversial) that racist, sadistic, or other antisocial preferences are not rational and so those who are concerned to benefit people can discount such preferences.

A rational self-interested preference satisfaction view of well-being thus resolves most of the difficulties facing the actual preference satisfaction view, but it does not resolve all of them. It makes an individual's state of well-being even less measurable than does the actual preference satisfaction view. Not only does the welfare economist face the problem of determining what people prefer, the normative economist must now determine which of these preferences are "rational," "self-interested," and "well-informed." It is difficult to flesh out these terms so as to make this a well-defined theory of well-being, and the result is unlikely to be the sort of theory that economists would find congenial. And yet, as mentioned before, well-being as the satisfaction of rational and self-interested preferences might coincide with some substantive account of well-being.

8.5 Alternative Theories of Welfare

Substantive theories of welfare purport to say which things are intrinsically good. Traditional hedonistic mental state views – to which, as we suggested, many economists still pay secret allegiance – are substantive theories, as are "perfectionist" views (Griffin 1986, ch. 4; Raz 1986, ch. 12) and what Parfit calls "objective list" views. Substantive views are objective in the sense that what is good for people is not determined by whether people believe it is good for them. Even as they take well-being to be the satisfaction of preference, many economists are also committed to a substantive view of well-being as material self-interest. These views are compatible if and only if people's preferences reflect their material self-interest.

The objectivity of substantive views of well-being does not imply that subjective states do not matter, and most substantive theories of the good count subjective mental states such as pleasures as goods and mental states such as pains as bads. Objective views are problematic, because it seems that what is intrinsically valuable varies from person to person. Objective views should nevertheless be tempting to economists when they make well-being more readily measurable.

In addition to material self-interest, one objective view of what matters for economic policy – which may be relevant to normative economists – is John Rawls's notion of "primary goods." This is not an alternative theory of well-being. In *A Theory of Justice* (1971) Rawls takes well-being to be the satisfaction of rational preference, but he does not think that justice should focus on well-being. The reason is that people's well-being depends in part on their own efforts. Social policy should attend to that component or aspect of welfare for which society is responsible. Rawls suggests that the relevant aspect of well-being be measured by an index of "primary social goods." Such goods (like income) are all-purpose means or, as Rawls put it: things that you want, whatever else you want (1971, p. 92; see also Chapter 12 in this volume). Primary social goods are not proxies for utility levels, which in any case are not (in Rawls's view) at issue in discussions of justice. Primary social goods offer an alternative basis for a more settled social agreement on which matters of social responsibility are important to well-being (1982). Rawls's approach avoids the expensive tastes and antisocial preferences problems and, as he argues, provides a more impartial perspective for comparing what society contributes to the well-being of different individuals than a preference standard does.

Amartya Sen defends a theory of well-being mainly in terms of a person's capabilities and the "functionings" an individual attains. A person's functionings are those things that the person does and experiences. Walking, playing the piano, being well-nourished, loving one's friends, understanding Chinese, and appreciating cubism are all functionings. But, like Rawls, Sen does not think that social policy should concern itself directly with the functionings that people achieve. Social policy, insofar as it is motivated by a concern for welfare, should instead focus on "capabilities" (Sen 1987c, 1992a). A capability is the ability to achieve a certain sort of "functioning." For example, literacy is a capability while reading is a functioning. People may value capabilities for their own sake as well as for the functionings they permit – you're glad to know you can walk around even if you are inclined to stay put.

In Sen's view, social policy should focus on capabilities rather than on functionings or on resources. Policy should not be concerned with functionings – with what people make of their capabilities – because functionings are to a considerable extent determined by individual choice. For example, a shortfall in functioning such as malnutrition might stem from an individual's decision to embark on a religious fast rather than from any deprivation. Policy should not focus on resources, either – and it is for this reason that Sen criticizes the primary-goods approach – because internal features of individuals also affect the range of functionings from which they choose. Owing to a digestive disorder, someone may be malnourished despite having a normal diet. The focus of policy should be on capabilities and hence on freedom, broadly understood.

At first independently of Sen and then in collaboration with him, Martha Nussbaum has also developed a conceptualization of human flourishing in terms of capabilities (2000, 2001). Her view of capabilities differs from Sen's in three ways. First, its roots are in Aristotle's moral philosophy rather than in the inadequacies of conventional welfare economics. Second, Nussbaum has formulated a specific list of central human capabilities (see Figure 8.5.1). She recognizes that some items on the list might be realized differently in different societies and that some items are more firmly fixed than others (with the centrality of bodily integrity being clearer, for example, than the role played by concern for other species and the world of nature). Nonetheless, she argues that the list has survived considerable cross-cultural scrutiny and defines a conception of well-being and freedom that can serve as the starting point for political deliberations in particular societies. While Sen has also been concerned to make the notion of capabilities measurable and practical, he has been more cautious about specifying the most important human capabilities.

The third difference between the way in which Nussbaum and Sen develop the capabilities approach is linked to the weighting of capabilities. Specifying a measure of well-being requires somehow weighting the many different capabilities and functionings that each person possesses. This issue of weighting might seem to bring preferences in through the back door. Ultimately, it might be objected, the importance of different functionings depends on people's preferences. But a reasoned public discussion over these weights need not wind up concluding that individuals' subjective preferences about them are their only or most important determinant. Nussbaum maintains that cross-cultural agreement can be reached on the centrality of the capabilities on her list. She goes on to argue that this list can then serve as "the moral basis of central constitutional guarantees, by people who otherwise have very different views of what a complete good life for a human being would be" (2000, p. 74).

Even if the problems of weighting and measuring the components of well-being prove to be greater for more objective approaches than are the problems of measurement in standard utility theory, it can still be argued,

1. *Life.* Being able to live to the end of a human life of normal length; ....
2. *Bodily Health.* Being able to have good health, including reproductive health; to be adequately nourished; to have adequate shelter.
3. *Bodily Integrity.* Being able to move freely from place to place; to be secure against violent assault, including sexual assault and domestic violence; having opportunities for sexual satisfaction and for choice in matters of reproduction.
4. *Senses, Imagination, and Thought.* Being able to use the senses, to imagine, think, and reason – and to do these things in a "truly human" way, a way informed and cultivated by an adequate education, ... Being able to use one's mind in ways protected by guarantees of freedom of expression with respect to both political and artistic speech, and freedom of religious exercise ....
5. *Emotions.* Being able to have attachments to things and people outside ourselves; to love those who love and care for us, to grieve at their absence; ....
6. *Practical Reason.* Being able to form a conception of the good and to engage in critical reflection about the planning of one's life. (This entails protection for the liberty of conscience and religious observance.)
7. *Affiliation.*
   A. Being able to live with and toward others, to recognize and show concern for other human beings, to engage in various forms of social interaction; ....
   B. Having the social bases of self-respect and non-humiliation; .... This entails provisions of nondiscrimination on the basis of race, sex, sexual orientation, ethnicity, caste, religion, national origin.
8. *Other Species.* Being able to live with concern for and in relation to animals, plants, and the world of nature.
9. *Play.* Being able to laugh, to play, to enjoy recreational activities.
10. *Control over One's Environment.*
    A. *Political.* Being able to participate effectively in political choices that govern one's life; having the right of political participation, protections of free speech and association.
    B. *Material.* Being able to hold property (both land and movable goods), and having property rights on an equal basis with others .... In work, being able to work as a human being, exercising practical reason and entering into meaningful relationships of mutual recognition with other workers.

Source: Abridged from Nussbaum (2001, pp. 416–18).

Figure 8.5.1. The Central Human Capabilities

following Sen, that "it is better to be vaguely right than precisely wrong." In particular, these objective approaches to well-being lead research on the topic in directions that link up more naturally to the normative terms characteristic of policy debate. For example, if one wants some measure of the extent of deprivation in less developed countries, utility theory is of little use. An objective index is needed. The one that is currently most used, the human development index, owes a great deal to Sen's views (and to Sen himself).

8.6 Conclusions

The identification of well-being with the satisfaction of preferences is questionable. It mistakenly suggests that social policy should attend to all preferences – even if they are expensive, antisocial, or the results of false beliefs, manipulation, or problematic psychological processes. The focus on preference satisfaction fails to link up with the normative terms of policy debate, and it leads to difficulties and implausible implications when preferences change and conflict. Moreover, the standard notion of utility is not easily measurable; and even if it were, normative economists cannot defend a measure of individual well-being as a good operational proxy without some conception of just what it is supposed to be a proxy for. The only real advantage of identifying well-being with the satisfaction of preferences is that it apparently ties welfare problems immediately to fundamental economic theory. But the link is of little value unless the theoretical "welfare" problems are truly problems of human welfare or well-being.

Suggestions for Further Reading

For an extended defense of an informed preference satisfaction view of individual well-being, see Gauthier (1986, ch. 2), Goodin (1986), and Griffin (1986). The introduction to *Utilitarianism and Beyond* (Sen and Williams 1982) contains a critical summary of arguments against preference-based approaches to measuring well-being. Arneson (1990) has provided the most powerful general response in defense of preference satisfaction views of welfare. See also Mill (1863); Rawls (1982); Sen et al. (1987); and Scanlon (1991).

The problems that arise in respecting preferences when there are uncertainties are discussed in Harris and Olewiler (1979); Hammond (1983); Machina (1987); Broome (1991b, ch. 10); and Hausman and McPherson (1994).

Sen's views on capabilities and functionings are developed mainly in two of his essays (1985a and 1987c) and are discussed at length in Nussbaum and Sen (1993). Nussbaum (2000, 2001) attempts to recast political theory in terms of fundamental capabilities that should be guaranteed to everyone (rather than in terms of natural rights).

NINE

Efficiency

Mainstream economists evaluate economic institutions, policies, and outcomes by asking whether they make people better-off. Given their theory of welfare, they are asking how well institutions, policies, and outcomes satisfy preferences. Yet few economists are utilitarians. Most normative economists deny that evaluations should be based on comparisons of how well satisfied are the preferences of different individuals. Following Robbins (1935, ch. 6), mainstream economists typically regard interpersonal comparisons as untestable subjective value judgments that are scientifically illegitimate. Though we have qualms about these arguments, in Chapter 7 we endorsed the conclusion that interpersonal comparisons of welfare are impossible when welfare is interpreted as preference satisfaction. We take this result as a further reason to question the identification of welfare with the satisfaction of preferences. For people do, in fact, make interpersonal welfare comparisons – and they need to do so. In our view, policy makers need to make interpersonal comparisons, too. They need to be able to compare the consequences of policies for winners and losers in a nonarbitrary way, which is also impossible without some sort of interpersonal comparison.

Welfare economists thus face a serious problem: How can they even make sense of the notion that one economic outcome satisfies preferences better than another if they cannot compare how well satisfied are the preferences of different individuals and cannot add up utilities? Indeed, denying the possibility of making interpersonal welfare comparisons – which is virtually inevitable once one identifies welfare with the satisfaction of preferences and adopts the framework for representing preferences that is accepted in positive economics – largely determines the character of mainstream normative economics. It undercuts any hope of developing a general ethical theory (like utilitarianism) that can encompass such other dimensions of moral appraisal as freedom, rights, equality, and justice, because all of these notions are concerned with the differing weights of claims of different persons, which cannot be addressed within a purely welfarist framework unless one can make interpersonal comparisons. So some strategy is inevitable whereby economists appraise policies, outcomes, and institutions "other things being equal" or "along one of several moral dimensions."

This strategy encourages the idea that there is a specifically economic dimension of evaluation – a realm wherein Summers can consider the economic "logic" of transferring pollution to poorer countries. This strategy makes it possible to envision a unified *theory* of economic evaluation as opposed to a mere set of normatively motivated inquiries into consequences and properties of economic policies and institutions. Welfare economics depends not only on a specific view of welfare but also on the view that inquiries into welfare can be separated from inquiries into freedom, rights, equality, and justice. In one way, this separation limits economists. They appraise policies only along one dimension or in one regard. But it also frees them from having to be concerned with anything but welfare. Having passed the buck with respect to everything except preference satisfaction, welfare economists have only too often (as in Summers's memorandum) felt themselves free to ignore all other moral questions and to exaggerate the significance of their own partial mode of evaluation.

9.1 "Efficiency" as Pareto Optimality

If welfare is taken to be preference satisfaction and if the extent to which the preferences of different individuals are satisfied cannot be compared, then how can one say anything substantial about economic welfare? The welfare economist's answer lies in the notions of a "Pareto improvement" and of "Pareto efficiency" or "Pareto optimality," which we defined in Section 5.2. A state of affairs X is a Pareto improvement over another state Y if and only if nobody prefers Y to X and at least one person prefers X to Y, and a state of affairs is Pareto efficient if there are no Pareto improvements that can be made. These are just definitions; they do not express any attitude toward Pareto improvements or Pareto efficient states of affairs. To maintain, as many economists do, that Pareto improvements are good things requires endorsing an interpretation of the so-called Pareto principle, which will be discussed further in Chapter 13. The Pareto principle states that if X is a Pareto improvement over Y then X is morally better than Y. It says nothing about how states of affairs X and Y compare when at least one person prefers X to Y and somebody else has the opposite preferences. As discussed in what follows, this means that the Pareto principle does not permit one to say much on behalf of Pareto efficient (or optimal) states of affairs.

The preference satisfaction theory of welfare (which says that any person is better-off if and only if his preferences are better satisfied) plus the Pareto principle implies the principle (which John Broome calls "the principle of the personal good") that "If two prospects are equally good for everyone, they are equally good, and if one of two prospects is better for someone than the other and at least as good for everyone, then it is better" (1989, p. 11). Conversely, the principle of the personal good and the preference satisfaction theory of welfare also entail the Pareto principle. Broome's principle of the personal good also implies the principle of minimal benevolence defined in Section 5.2, which states that, other things being equal, it is a morally good thing if people are better-off. But the minimal benevolence principle is weaker than the principle of the personal good, because when other things are not equal an outcome that makes everyone better-off may nevertheless be worse because of other morally relevant features, such as an increase in inequality.†

There are problems with endorsing all Pareto improvements (as the Pareto principle does) and further problems with endorsing only Pareto improvements (which goes beyond what the Pareto principle advocates). First, supporting a Pareto improvement may be unappealing when it involves honoring people's "nosy" preferences regarding others' conduct (see the discussion of Sen's liberty paradox in Chapter 13). Second, there are technical problems. It has been proven that one cannot consistently endorse Pareto improvements in circumstances of uncertainty if one also holds that individual and social preferences satisfy the axioms of expected utility theory (see Hammond 1983; Seidenfeld, Kadane, and Schervish 1989; Mongin 1995). Unanimity in individual preference rankings may rest on offsetting disagreements both in subjective probability judgments and in preferences among options involving no uncertainty. For example, suppose the bourgeoisie and proletariat both prefer an armed workers' revolt, although they have opposing aims and form different estimates of the likely consequences. The preference and belief systems of the two classes satisfy the axioms of

† Let PP be the Pareto Principle, PST the preference satisfaction theory of welfare, PPG the principle of the personal good, and MB the principle of minimal benevolence. Then one can summarize these logical relations as follows (where → denotes "implies"):

PP & PST → PPG;
PPG & PST → PP;
PPG → MB;
PP & PST → MB.

expected utility, but there is no coherent set of social preferences and expectations; and the two classes' agreement with respect to this one preference does not establish that an armed revolt would be socially better (see also Levi 1990).

Third, one must be able to make comparisons between states of affairs when neither is Pareto superior to the other. Pareto improvements are rare. Economic changes usually involve winners and losers, and it is not a matter of indifference who wins and who loses. Suppose that there is a single consumption good (bread, say) in some fixed quantity and that everybody prefers more rather than less of it. Then every distribution of bread that exhausts the bread supply is Pareto efficient, and none of these distributions is a Pareto improvement over any other. Moreover, *R* may be Pareto optimal and *S* may be suboptimal without *R* being a Pareto improvement over *S*. Suppose there are only two people, *A* and *B*, and 10 loaves of bread to distribute. A Pareto efficient allocation that gives 7 loaves to *A* and 3 to *B* is not a Pareto improvement over the (suboptimal) allocation that wastes 2 loaves and gives 4 to both *A* and *B*, and the Pareto principle says nothing about which is better.

Very few economic states can be ranked in terms of the relationship of a Pareto improvement. If *R* is a Pareto improvement over *S*, then *nobody* can prefer *S*. If everybody in *R* except for Donald had triple the income in *R* that they would have in *S* and Donald had one penny less, then (assuming that Donald for reason of that penny prefers *S*) *R* is not a Pareto improvement over *S*. Notice that with just a tiny bit of redistribution *R* could be transformed into another state R' that would be a clear Pareto improvement over *S*. This point, which is crucial to cost–benefit analysis, will be discussed later in the chapter.

The Pareto principle has some real ethical appeal because satisfying preferences surely has something to do with promoting well-being. For example, consider the following potential exchange situation between individuals *A* and *B*: *A* has a Toyota and *B* has a Ford. Both *A* and *B* value their own vehicle but would prefer to exchange. Exchanging is the "cooperative" solution, and it is a Pareto improvement over the initial allocation. Yet unless both parties can count on being protected from non–Pareto improving outcomes, they may hesitate to interact at all. For if *A* could have both the Toyota and the Ford then that would be best for *A* and worst for *B*, and it would be worst for *A* if *B* could have both. Property rights prevent either from forcibly appropriating what belongs to the other, thus enabling the cooperative solution to arise voluntarily (Hardin 1988, ch. 3). The Pareto notions help one to understand the roles that property rights and exchange play, and they permit one to see why some systems of property rights are superior to others.

Perfectly competitive markets are in this regard "good" institutions because – as discussed in Chapter 5 – the outcomes of interactions on them are Pareto efficient. But real market outcomes may be suboptimal, particularly (though not only) in the case of *externalities.* As we saw in the introductory discussion of the Summers memorandum, sometimes the costs and benefits of an agent's actions do not fully register as costs or benefits for that agent. Those who pollute the air or deplete the fish population in a lake often need not take the costs imposed on others into account in their private calculations of economic costs and benefits. Nor can someone who builds a lighthouse collect from all those who benefit from it. One solution to the problems posed by externalities is a more refined assignment of property rights so that either (a) polluters will have to compensate those with a right to clean air or (b) those who want clean air will have to pay polluters not to do so. As Coase (1960) proved under some fairly restrictive assumptions, the same "optimal" amount of pollution will result with either assignment of rights. Since the "transaction costs" – the costs of finding the parties one needs to bargain with and striking and enforcing these bargains – are often prohibitive, precise rights assignments do not solve all the problems. Government provision of collective goods (such as lighthouses), government restrictions (such as hunting and fishing limits or limits on pollution), and government taxes or subsidies can mitigate suboptimal outcomes.

Most economists regard taxes and subsidies as preferable to restrictions or mandates, since taxes and subsidies are arguably Pareto superior to less flexible requirements and also permit a greater range of individual choice. But recall the discussion in Chapter 5. Taxation expresses a different attitude toward behavior than does legal prohibition. Whether or not it were more efficient, no one would propose taxing robberies rather than prohibiting them. To tax rather than to fine pollution is to treat pollution as socially and morally acceptable. Some pollution is socially acceptable. For example, so long as we need power generated by fossil fuels, those who own and operate the plants cannot avoid discharging carbon dioxide (the major culprit in global warming) into the air. Indeed, all of us who have not given up breathing join them in this pollution. Prohibition is out of the question, and taxes on greenhouse emissions may be the most sensible way to reduce them. On the other hand, some pollution is not socially acceptable and indeed some pollution seems criminal. Dumping known poisons into people's drinking water should be prohibited rather than merely taxed (Kelman 1981). Efficiency in the sense of Pareto superiority should not trump all other moral considerations.

The Pareto notions obviously have little bearing on questions of fairness. Economists often suggest that questions of economic welfare be factored into questions of efficiency, to which the Pareto concepts are by definition pertinent, and questions of equity, upon which theoretical welfare economics has little to say (Okun 1975). Such factoring is implicit in the distinction in the Summers memorandum between the "economic logic" of transporting pollution to LDCs and the "moral reasons" or "social concerns" for which Summers has little patience. As normative economists are themselves well aware (see e.g. Laffont 1988), this factoring is dubious because questions of efficiency and questions of distribution cannot be separated. Distribution affects well-being and hence "efficiency" through its effects on social solidarity, crime, social mobility, health, and even envy (Hirsch 1976, pp. 131–2). An unequal distribution fuels potentially wasteful competition for the top positions in a "winner take all" society (Frank and Cook 1995). The size of the social pie can depend on how it is cut.

9.2 How Welfare Economics Narrows Normative Questions

An example may clarify how the Pareto notions may be put to work and also show how a standard economic treatment of policy questions narrows the morally relevant issues. Consider the question of whether government should provide health care "in kind" or provide cash that the needy can use to purchase health care. Many economists would offer a simple argument for why the government should supply cash: Consider the graph of Figure 9.2.1.

The vertical axis represents quantity of health care (h) while the horizontal axis represents the quantity of some composite consumption commodity – that is, of everything else consumed (ee). Points in the plane represent possible individual consumption bundles, which contain quantities of health care q_h and the "everything else" commodity, q_{ee}. As one moves northeast, the amount consumed increases, and if one assumes that people prefer having more than less of both health care and "everything else," then northeast is also the direction of preference. The curving lines I_1, I_2, and I_3 are "indifference curves" for some individual whom we call Alice; they are called indifference curves because they are drawn through points (consumption

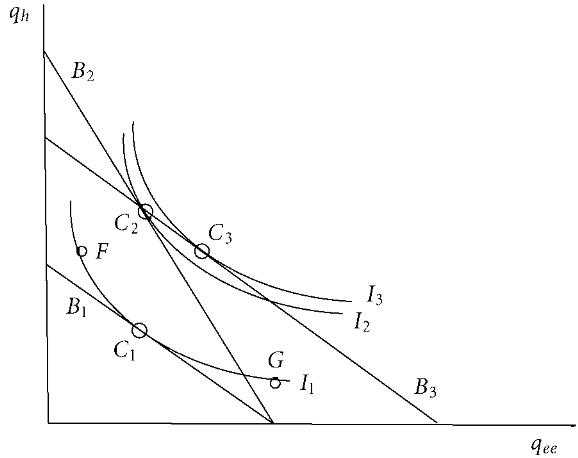

Figure 9.2.1. Cash versus In-Kind Benefits

bundles) among which Alice is indifferent. For example, Alice does not care whether she consumes bundle F that contains lots of health care and comparatively little of everything else or bundle G that contains little health care and lots of everything else. Fundamental principles of the standard theory of consumer choice imply that indifference curves have the shape shown here and that they do not cross. Since some points on I_2 are northeast of points on I_1, it follows that every bundle on I_2 is preferred to every bundle on I_1. If Alice is able to "move" from I_1 to I_2, then her preferences are better satisfied and she is better-off.

The straight lines sloping downward to the right represent three different budget constraints on Alice. Given an amount of money to spend equal to B_1, Alice can afford only those consumption bundles on or below the line B_1. Since points above B_1 are unaffordable and since, for every point below B_1, Alice prefers some point on B_1, Alice will consume a bundle on her budget line – and indeed (apart from some technical complications) a bundle on an indifference curve that is tangent to her budget line.

Suppose that Alice begins with the budget constraint B_1 and so her consumption is C_1, which is on the indifference curve I_1. Suppose then that the government decides to subsidize health care. This does not change the amount of money Alice has available to spend, but it does make health care cheaper. Her budget constraint will now look something like B_2, and her consumption will be C_2 on indifference curve I_2. Since I_2 is northeast of I_1 it follows that Alice is better-off at C_2, where she consumes more health care and less of everything else than she did at C_1. (Although here the health-care subsidy led Alice to consume more health care and less of everything else, this result depends on the shape of the indifference curves and does not hold in general.)

Now suppose that instead of subsidizing health care the government gave Alice enough cash that she could, if she wished, purchase C_2. Since a cash allowance to Alice does not change the relative prices of health care and everything else, it will shift her budget constraint to B_3, which is parallel to her original budget line B_1 and passes through C_2. But given her preferences, Alice will not consume C_2; she will instead consume C_3 on indifference curve I_3. The cash payment must satisfy Alice's preferences at least as well as the in-kind health-care benefit, and it will typically satisfy her preferences better. The cash payment is apparently Pareto superior to the in-kind benefit, because no one will prefer the in-kind provision and many will prefer the cash payment.

The advantages of the cash payment do not stop there. It is also apparently cheaper to give people cash than to give them health-care benefits, because one needs no bureaucracy to process claim forms or to investigate medical frauds. Since cash benefits are cheaper to administer, one can satisfy all recipients' preferences at least as well with cash benefits as with in-kind provision and at a lower cost to taxpayers, or one can provide even more benefits.

Some criticisms of this argument for cash payments can be formulated in terms that are consistent with taking welfare as the satisfaction of preferences. Thus, one could argue that taxpayers may themselves have preferences for certain kinds of consumption patterns among the recipients of assistance and hence the welfare of taxpayers may be increased by resorting to in-kind provision. An obvious response is to object (on grounds of liberty) to constraining the consumption decisions of the needy in order to please taxpayers. However, this line of argument departs from the framework of welfare economics by appealing to the value of freedom. A second criticism of cash provision that can be cast within the standard welfare economics framework appeals to the difficulty of monitoring claims for assistance. Cash benefits create incentives to feign illness, whereas the luxuries of waiting in line at a free medical clinic are not very enticing to the healthy.

The central arguments in favor of in-kind provision cannot, however, even be expressed explicitly in a framework that focuses exclusively on welfare as the satisfaction of preferences. One argument, which we mentioned in Section 8.3, distances the considerations that govern social policy from subjective measures of well-being. Health care possesses an "urgency" that may be lacking in a person's other objects of desire.

Other arguments are openly paternalistic. Regardless of whether the poor actually want more health care, "we" know that more health care would be good for them. Paternalism is the view that it is sometimes right to coerce individuals for their own good. The paternalist case for in-kind benefits is not a frivolous one to be dismissed by supposing that what people most prefer is automatically what is best for them. Whether or not one is inclined to support some paternalistic policies (such as mandatory seat-belt laws or prohibitions on recreational drugs), it is important to recognize the morally significant differences between *judging* that someone is acting against her interest, *persuading* her to act differently, and *coercing* her. To deny that individuals always prefer what is good for them is merely good sense, not paternalism. Indeed, in his classic critique of paternalism in *On Liberty* (1859), John Stuart Mill notes explicitly that people may be wrong about what is good for them (see also Dworkin 1971, 1988; Feinberg 1986). If you think that a friend would be better-off not smoking, even though your friend thinks otherwise, you are not being paternalistic. (In fact, you're probably right!) Nor are you being paternalistic if you attempt to persuade your friend to give up smoking. But if you throw away your friend's cigarettes or threaten to beat him up if you find him smoking, then you are behaving paternalistically.

It may be good social policy to give people cash with no strings attached. But whether there should be any paternalistic strings attached to benefits needs to be considered rather than assumed away by supposing that whatever people prefer is automatically best for them. Furthermore, the most forceful objections to paternalism are not that it reduces welfare but that it improperly interferes with freedom. These aspects of the discussion are very poorly handled when the analysis is confined to Paretian welfare concerns.

Although the value of freedom lurks within the standard argument for cash benefits, the argument remains within the terms set by orthodox economic theory. There is no mention of needs, of the presuppositions of individual dignity, of opportunity, of rights, or of fairness. There is no concern with the moral reasons that make individuals willing to pay taxes to provide such benefits. Are people motivated by a general concern to satisfy the preferences of others, or do they instead see themselves as obligated to help others in need? Might they regard people as having rights to food or medical care which justify taxing others? These questions are hard even to pose within the framework that mainstream economists typically employ.

9.3 Cost–Benefit Analysis

We have criticized economists for relying so heavily on the Pareto concepts on the grounds that welfare is not the satisfaction of preference and that focusing only on welfare unreasonably narrows the terms in which institutions and policies can be evaluated. Yet what has bothered normative economists is instead that the scope of the Pareto concepts is so limited. There is almost always some person who prefers that the government not intervene to build a dam, save an endangered species, or clean up a river. So even though externalities often lead to unfortunate outcomes, the feasible remedies are rarely true Pareto improvements.

Theorists have struggled to increase the range and force of welfare economics in three ways. One way is to develop formal theories of social choice. These will be discussed in Chapter 13. A second way, which we discussed in Chapter 5, relies on the central theorems of welfare economics. These purportedly show that perfect competition is ethically attractive. But as we argued there, the fact that perfect competition guarantees Pareto efficiency is of little moral importance. One reason for this is that Pareto efficiency is not a big deal, morally speaking. Inefficiency just means that some improvements could in principle be made; efficiency means that no further such improvements are possible. Another reason not to be overly impressed with the welfare theorems is that perfect competition is unattainable. Furthermore, the efficiency of perfect competition does not imply that eliminating particular impediments to actual competition increases efficiency.

The third way to extend welfare economics is to find some way of comparing policies when none is Pareto superior to the others. Hicks (1939) and Kaldor (1939) approached the problem as follows. Consider two economic outcomes or states of affairs *X* and *Y*. There are many different moral comparisons people might make of them. One morally significant difference between *X* and *Y* may be *distributional,* as in the case of the 10 units of bread. Another way that economic states of affairs may differ is in the *quantity* of economic benefits to be distributed – that is, in their capacity to satisfy preferences. Suppose that in *X*, the status quo, 4 units of bread go to *A* and 6 units go to *B*. A new policy is considered that would increase bread supply and result in *A* getting 7 units of bread and *B* getting 5. This alternative, *Y*, is not a Pareto improvement over *X* because *B* receives fewer units of bread, but there is, Hicks and Kaldor argue, an unambiguous increase in economic benefits and economic efficiency. With the new policy, the capacity of the economy to satisfy preferences has increased; the "pie" has grown larger. That increase does not show up as a Pareto improvement because there is also a change in distribution to *B*'s disadvantage. Owing to the way the pie has been cut, *B*'s portion diminishes. In the view of Hicks and Kaldor, welfare economists are in no position to pass moral judgments on economic distribution – but they do not have to. The increase in efficiency, the purely economic benefit, is independent of distribution. Economists should be concerned to enlarge the pie and should leave its division to politicians and moralists. Hicks and Kaldor envision a separate dimension of purely economic evaluation.

According to Hicks and Kaldor, in order to determine whether *X* is more economically efficient than *Y* – that is, whether *X* has a greater capacity to satisfy preferences than *Y* – one need only determine whether *X* is a "potential Pareto improvement" over *Y*. Policy *X* is a potential Pareto improvement over *Y* if there is some way of redistributing the goods available in *X* that makes *X* an actual Pareto improvement over *Y*. So, in the simple bread example, the distribution of 7 units to *A* and 5 to *B* is a potential Pareto improvement over the distribution of 4 units to *A* and 6 to *B*, because it would be possible to redistribute the 12 units so as to achieve an actual Pareto improvement. (For example, both *A* and *B* could receive 6 units of bread.)

One can also describe a potential Pareto improvement in terms of the possibility of "compensation": if *X* is a potential Pareto improvement over *Y*, then it is possible for the winners in a change from *Y* to *X* to compensate the losers. Whether the winners could indeed compensate the losers is then operationalized. The simplest way (and the only way we shall discuss here) is in terms of willingness to pay. If the amount that winners would be willing to pay to bring about a policy is larger than the amount that losers would need to be compensated to accept the policy, then the policy is a potential Pareto improvement over the status quo and thus purportedly brings about a more efficient state of affairs in which there is a "net benefit" – a greater capacity to satisfy preferences.

The notion of a potential Pareto improvement thereby provides one possible justification for cost–benefit analysis (Mishan 1971, 1981). In theory, one asks the winners in each policy how much they would pay to institute the policy and also asks the losers how much compensation they would require not to oppose implementing the policy. The policy with the largest net benefit is best, other things being equal. Dumping toxic wastes in LDCs without paying compensation would thus apparently be a potential Pareto improvement. All things considered, potential Pareto improvements might be bad things as a result of their distributional consequences. But according to this view, the distributional questions are not questions concerning which economists have any particular expertise. Furthermore, if the problems are distributional, then so are the solutions. The judgment concerning economic efficiency stands:

> whether or not compensation should be given in any particular case is a question of distribution, upon which there cannot be any generally acceptable principle .... If measures making for efficiency are to have a fair chance, it is extremely desirable that they should be freed from distributive complications as much as possible. (Hicks 1939, pp. 711–12; see also Kaldor 1939, p. 550)

Economists have also defended cost–benefit analysis as a way to implement a utilitarian policy of maximizing total welfare. Here is how the basic ideas of cost–benefit analysis are introduced and defended in one text.

> The only basic principle is that we should be willing to assign numerical values to costs and benefits, and arrive at decisions by adding them up and accepting those projects whose benefits exceed their costs.
>
> But how are such values to be arrived at? If we assume that only people matter, the analysis naturally involves two steps. First, we must find out how the decision would affect the welfare of each individual concerned. To judge this effect we must ultimately rely on the individual's own evaluation of his mental state. So the broad principle is that "we measure a person's change in welfare as he or she would value it". That is, we ask what he or she would be willing to pay to acquire the benefits or to avoid the costs ....
>
> The second step is to deduce the change in social welfare implied by all the changes in individual welfare. Unless there are no losers, this means somehow valuing each person's £1. If income were optimally distributed, £1 would be equally valuable regardless of whose it was; that is what "optimally distributed" means .... And if income is not optimally distributed [then] most economists would argue that it should be redistributed by cash transfers rather than through the choice of projects. But what if we think that cash will not be distributed, even if it should be? Then we may need to value the poor person's extra £1 more highly than the rich person's. (Layard and Glaister 1994, pp. 1–2)

This quotation reveals a great deal about a familiar economic approach to problems of policy evaluation. Layard and Glaister claim that the "only basic principle" is that the considerations bearing on policy evaluation must always be comparable and quantifiable. This is a basic principle – and indeed a controversial one – but it is hardly the only one, and Layard and Glaister immediately assume another: that only people matter. This second principle is controversial, too, and has been explicitly challenged by those concerned with the consequences of human actions for animals (Singer 1975) and the nonsentient environment (Stone 1996). Layard and Glaister then jump from the premise that only individual people matter to the conclusion that only the *welfare* of individual people matters. They do not say why one should believe that the only thing that matters is welfare. What about, for example, people's rights and liberty? The next sentence asserts that we should accept the individual's own evaluation of his welfare (which Layard and Glaister equate with a mental state, despite the official identification of welfare with preference satisfaction). From comments they make later, it seems that they would argue for this assertion on the basis of principles of antipaternalism and democracy. Their reliance on such principles suggests that Layard and Glaister do not, in fact, believe that only welfare matters. The final paragraph then equates the net benefits of projects with their aggregate welfare consequences. Of course, these cannot be determined without making interpersonal welfare comparisons. If one assumes that willingness to pay can serve as an interpersonally comparable measure of individual welfare, then projects whose benefits exceed their costs will increase total welfare. As Layard and Glaister themselves point out, this assumption is implausible, and we shall have more to say about it.

Regardless of the justification for basing social policy on willingness to pay, it is in practice costly to ask people how much they would pay or how much compensation they require, and their answers may be neither truthful nor accurate. Welfare economists have accordingly devised cunning methods of gleaning such information from data on prices and quantities traded and sometimes from responses to surveys. Much of cost–benefit analysis is devoted to devising, criticizing, and improving methods of imputing costs and benefits.

9.4 Objections to Cost–Benefit Analysis

As mentioned before, the progenitors of cost–benefit analysis equated potential Pareto improvements with improvements in productivity, efficiency, or real income, and they hoped to separate the study of these uncontroversial economic benefits from controversial questions concerning distribution. But as already argued and by now generally recognized, there is no way to separate questions of efficiency in satisfying preferences from questions concerning distribution. Indeed, as Paul Samuelson pointed out more than fifty years ago, even a preference for policies that yield actual Pareto improvements is not, in general, free of distributional commitments.

So one objection to cost–benefit analysis is that the existence of a net benefit does not establish that there has been a real increase in the capacity to satisfy preferences. A net benefit does not imply a bigger "pie." To see why, consider Figure 9.4.1, which we have borrowed with some modifications from Samuelson (1950). The utilities of two representative individuals

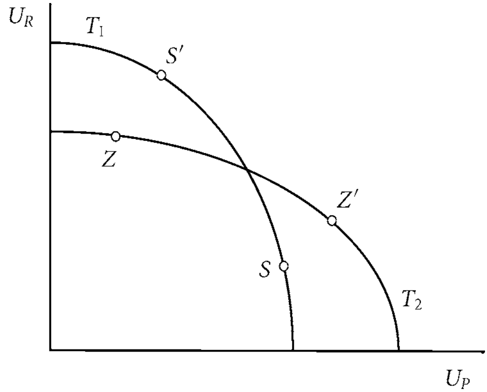

Figure 9.4.1. Interdependence of Equity and Efficiency

Rachel (R) and Peter (P) are measured along the two axes, and the two curves represent the possible utility combinations depending on whether technology T_1 or technology T_2 is employed. Technology 1, which is the status quo (perhaps in 1920), involves a rail transport system with few goods or people carried on highways. Technology 2 involves an extensive road system like the one currently in use in the United States. The utility status quo in 1920 is, let us suppose, point S, and the result of changing to an automobile technology will (other things being equal) result in the utilities for Rachel and Peter represented by point Z. Note that Z is not a Pareto improvement over S since Peter is worse-off, but Z is a potential Pareto improvement over S because Rachel can pay compensation to Peter and the economy can move along the T_2 curve to Z'. However, one cannot conclude, as Hicks and Kaldor hoped, that T_2 is more efficient than T_1 – or that it involves a greater capacity to satisfy preferences, a larger real income, a bigger "pie" – because S here is also a potential Pareto improvement over Z: one can move from S along the T_1 curve to S', which is an actual Pareto improvement over Z. Furthermore, even if the choice were between S and Z' rather than between S and Z, so that what was at stake were an actual Pareto improvement rather than merely a potential Pareto improvement, one would still not be justified in concluding that T_2 is more efficient than T_1.

In just the same way, the efficiency of relocating polluting industries in LDCs that Summers points to depends in large part on the lopsided distribution of wealth. Indeed, one can interpret R and P as standing for "rich" and "poor" and take T_2 to be the current technology (with polluting industries located in rich countries), Z the current level of preference satisfaction for representative members of rich and poor countries, and S' the Pareto improvement that relocating the polluting industries will make possible.

Though S' is a Pareto improvement over Z, it cannot be compared to Z', which could be achieved by redistribution without shifting polluting industries. Only when one utility possibility curve is inside of the other can one make a "pure" efficiency comparison that does not take a particular distribution for granted.

A second problem with cost–benefit analysis is that it compares the "willingness to pay" – rather than the welfare gains and losses – of different people. Though willingness to pay obviously has something to do with welfare and preference, it also depends on expectations concerning what is appropriate to purchase and for what price. One would get very misleading information concerning the relative value of a car and a spouse by investigating how much someone would be willing to pay for each! Willingness to pay, like the amount of money one would require in order to consent to an unwanted change, obviously also depends on wealth. Since preferences in cost–benefit analysis are weighted with dollars and since the poor have fewer of these, their preferences count for less (Baker 1975). If the benefits of a policy accrue to the poor and the costs to the rich, what the poor would be willing to pay for the new policy might not be enough to compensate the rich. The policy might nevertheless have had large net benefits if there had been a more even distribution of wealth, and someone willing to make interpersonal comparisons might be convinced that the welfare of the poor would increase more than the welfare of the rich would decrease. See Figure 9.4.2.

Third, cost–benefit analysis – like other methods of evaluation that employ the Pareto criteria – ignores questions of justice; moreover, unlike the endorsement of Pareto improvements, cost–benefit analysis endorses policies that make some people worse-off. The compensation considered is only hypothetical. Some people win and some lose. Questions of fairness are obviously pressing in such circumstances. If each policy had different winners and losers – so that, in the long run, every individual won as often as she lost – then the unfairness of individual policies taken separately might wash out. But the bias built into cost–benefit analysis against the preferences of the poor suggests that the unfairness will not wash out. Exactly those people whom policy makers should be most concerned to protect are those who are most likely to be harmed. Proponents have consequently explored ways of modifying cost–benefit analysis to compensate for possible injustices (Little 1957; Harberger 1978), but in practice wealth adjustments are seldom made.

A fourth objection to cost–benefit analysis is that social policy should not be based on the unreflective and unargued preferences that analysts deduce from people's economic choices. Some preferences, such as preferences for

J. Bradford DeLong shows that a policy of maximizing net benefits coincides with a utilitarian policy if and only if each individual's utility is weighted inversely proportionally to each individual's marginal utility of wealth. Since the marginal utility of wealth (the extent to which an increment of wealth contributes to the satisfaction of an individual's preferences) decreases as one gets richer, utilitarianism supports maximizing net benefits if and only if one places increasing weight on satisfying the preferences of the rich. Here is a sketch of the argument: Suppose that maximizing net benefits coincides with maximizing some weighted sum of individual utility. Overall social welfare equals $w_1U_1 + \cdots + w_nU_n$, where U_j is the utility of the jth member of the population and w_j is the weight placed on that individual's utility. If social welfare is maximized, then transferring a small amount of some specific commodity c from individual i to individual j should keep the total near the maximum (in the limit, as the amount transferred is smaller and smaller, one remains at the maximum). This means that the amount of utility that i loses multiplied by the weight given to i's preferences should equal the amount of utility that j gains multiplied by its weight. So $w_i\Delta U_i = w_j\Delta U_j$. Since each individual is using his or her wealth to maximize overall preference satisfaction, it follows that $\Delta U_i/p_c$ equals the marginal utility of wealth for i and that $\Delta U_j/p_c$ is the marginal utility of wealth for j. Hence the weights given to i's and j's utilities are inversely proportional to i's and j's marginal utility of wealth. See ⟨http://www.j-bradford-delong.net/movable_type/occasional/The_Market's_SWF.pdf⟩.

Figure 9.4.2. Utilitarianism and Maximizing Net Benefit

communities free of urban sprawl, are hard to signal when one buys groceries, cars, or even homes. Furthermore, people's preferences for public goods of all sorts respond to arguments and may be different after public debate than they were before. Substituting cost–benefit analysis for public deliberation means that people's preferences are never subjected to such challenges.

In addition, it is questionable whether decisions about roads or schools or tanks should be equally responsive to all preferences, whether or not they are weighted by wealth. For example, Americans may prefer a larger military budget for many reasons: They may believe that the United States should aim for world domination in order to secure international peace; they may believe that preemptive strikes against rogue states will be increasingly necessary to protect the United States from terrorist threats; they may believe that the Soviet empire is regrouping; they may believe that lower military budgets will cause unemployment. Are not the reasons relevant? Should a decision about the size of the military budget respond only to the strength of people's preferences (or to their willingness to pay), or should it rather look to the reasons for the preferences?

Finally, uncertainty – coupled with the fact that preferences and willingness to pay typically depend on beliefs – creates serious problems. When people have mistaken beliefs about the constitution of the exhaust from the factory down the road, their willingness to pay to avoid breathing it will be an unreliable indicator of their true preferences, let alone of the exhaust's welfare consequences (see Chapter 8). People often do not know the consequences of alternatives and hence cannot identify which alternative they would prefer if they did know the consequences. The problems of uncertainty are usually finessed by supposing that individuals possess subjective probability distributions over all the possible outcomes. But to suppose this involves extreme idealization, and there is little justification for respecting preferences based on largely fictitious probability distributions.

9.5 Cost–Benefit Analysis as a Social Practice

Cost–benefit analysis does not enable welfare economists to settle questions of social policy "scientifically," without employing value judgments. It cannot illuminate the "economic" dimensions of a problem while leaving the other dimensions to be decided elsewhere: it is not generally possible to decompose "distributive" and "efficiency" effects of policies. Nevertheless, it is clear that cost–benefit analysis has played and will continue to play a significant role in policy making, and indeed the U.S. government requires that cost–benefit analyses be conducted when it issues regulations governing product safety or environment protection. If cost–benefit analysis lacks the firm analytical and ethical foundations that its originators thought it had, what should we think about it as a social practice?

One option might be to push for some more refined version of cost–benefit analysis that employs distributive weights or explicitly incorporates assumptions about the social welfare function. But refined forms of cost–benefit analysis still lack firm ethical foundations, and they are much harder to apply. Another possibility is to regard cost–benefit analysis as in some cases a convenient and reasonable rule of thumb to help guide social decisions but not as a way to find the "right" answer in some more basic sense. (Majority voting can be seen as a decision rule with some similar features; see Chapter 13 for further discussion of social decision-making rules.) Such rules of thumb make sense if they are reasonably unbiased and if the costs of more deliberative decision-making processes are excessive. For example, a hospital emergency room may sensibly apply a rule of "first come, first served" for most patients, but ought to invoke other criteria when a woman is brought in while undergoing cardiac arrest.

In light of the objections to cost–benefit analysis, it would be unreasonable to apply it indiscriminately even as a rule of thumb. Cost–benefit analysis is often biased against those with low ability to pay, and its results should be regarded with suspicion when people are likely to be influenced by mistaken beliefs. A cost–benefit rule may be less worrying in cases where the winners and losers are close to equal in ability to pay, where uncertainty about consequences is limited, and where the kinds of reasons recommending different policies are widely understood. A cost–benefit analysis might thus provide a reasonable basis for deciding whether to put a baseball field or a soccer field in a public park. It would be harder to defend using such analysis to decide whether a public park or a free clinic for destitute people was a better use for a particular piece of land.

Although we have not surveyed the many uses of cost–benefit analysis in public decision making in the United States, we suspect that the criteria that would make its use reasonable are frequently not satisfied. Moreover, any temptation to drape the results of such analysis in the robes of science is to be resisted. The results of a cost–benefit analysis can be informative, but they can never show that a particular policy is "scientifically" the right thing to do. It is clear that, on issues of sufficient import to warrant serious public deliberation, the findings of a cost–benefit analysis should not trump other considerations that citizens wish to bring to bear.

9.6 Conclusion: Welfare Economics in Limbo

If one evaluates policies exclusively in terms of their consequences for individual welfare, and if one takes welfare to be the satisfaction of preferences, and if one denies that it is possible to compare how well-off different people are, then it seems there is little that can be said about which policies and institutions are better. Orthodox welfare economists are in precisely this position. Straining against these self-imposed shackles, they have managed to salvage a practical method of evaluation – cost–benefit analysis – but it rests on shaky foundations precisely because the fundamental evaluative apparatus of contemporary normative economics is so narrow. A central problem, which is widely acknowledged in principle but often ignored in practice, is that the separation that Hicks and Kaldor envisioned between questions concerning efficiency and distribution – between the size of the pie and the way it is sliced – cannot in general be made. Because distributions can influence the willingness to pay by which cost–benefit analysis weighs preferences, there is no way to separate questions concerning the capacity to satisfy preferences from distributional questions. One cannot reasonably evaluate policies, institutions, or states of affairs exclusively in terms of their success at satisfying the interpersonally noncomparable preferences of individuals. For welfare is not preference satisfaction, and values such as freedom, equality, and justice also matter.

Cost–benefit analysis is not a scientifically valid or value-neutral procedure for social decision making. Even if one accepts its ethical presuppositions (minimal benevolence and the identification of welfare with the satisfaction of preferences), cost–benefit analysis is still a potentially misleading source of information concerning the real benefits and drawbacks of alternative policies. Furthermore, it is illusory to suppose that one could acquire information concerning "the real benefits and drawbacks" of policies in isolation from broader moral questions about fairness. Willingness to pay or to accept compensation is at best an imperfect indication of preference satisfaction – one that privileges the status quo distribution – and preference satisfaction is at best an imperfect indicator of welfare. That the winners from a policy change could (but do not) compensate the losers does not show that the policy increases the capacity to satisfy preferences, let alone that it increases welfare. And information concerning the net benefits of a policy is of course nowhere near showing that the policy is right.

But, one might object, what other measure of the good versus the bad consequences of policies do we have? The enthusiasm of political supporters or the complaints of political opponents of a policy are hardly reliable indicators of the policy's net benefits. Furthermore, in the right circumstances, it seems that a cost–benefit analysis can contribute useful information to a social decision problem. Consider the example just mentioned of a town council trying to decide between putting a soccer field or a baseball field in a public park. Suppose a poll indicates that the majority want a baseball field while a cost–benefit analysis reveals that the residents would be willing to pay considerably more for soccer use than for baseball use. Further investigation reveals no difference between the wealth of those who favor soccer compared to those who favor baseball, and the policy choice will not affect the distribution of income. Nor are the results skewed by a few fanatics. So willingness to pay is in this case a reasonable indication of intensity of preference, and it is plausible with respect to matters such as athletic fields to take preference satisfaction as a measure of well-being. So even though the majority of residents prefer the baseball field, the town council might rely on information provided by a cost–benefit analysis to conclude that the soccer field should be built instead.

Cost–benefit analysis may have its most plausible uses in relatively mundane cases like this. In circumstances where those who favor a policy are richer than those who oppose it, where the consequences will be lasting, serious, and uncertain, or where the policy choice raises other moral questions (such as questions of equity), the application of cost–benefit analysis – even as merely a source of information – will be much more questionable. Unfortunately, it is in cases with precisely these features – cases such as determining what neighborhood should be destroyed for a new highway, or whether a dam that threatens an endangered species should be built – that policy makers often seek the illusory guidance provided by cost–benefit analysis.

In making limited use of the findings of a cost–benefit analysis, policy makers need to ask to what extent the net benefits depend on the existing distribution of income and wealth and whether greater benefits might be obtained through redistribution. To examine a single alternative to the status quo (which, like the status quo, takes for granted the existing distribution of income) is one of the easiest ways to abuse cost–benefit analysis. The fact that the alternative provides net benefits says nothing about whether it should be instituted. Recall our discussion of the Summers memorandum: It is plausible that redistributing income would increase overall well-being much more than redistributing pollution. And once one begins thinking of overall well-being, one has moved beyond cost–benefit analysis and the futile hope that efficiency questions can be sharply separated from distributional questions.

Provided that one takes account of distributional presuppositions and consequences and also recognizes that cost–benefit analysis is a highly fallible source of data, such analysis can be of use. And given the absence of alternative measures of benefits and harms that are accurate and unbiased, it will be tempting to rely on cost–benefit analysis heavily and frequently. Though we see no good alternative, we conclude on a note of warning. Policy makers need numbers, and cost–benefit analysis provides numbers. Though the numbers that cost–benefit analysis provides are sometimes informative, they are often misleading. There is no substitute for rational deliberation and no uniformly reliable source of information concerning the welfare possibilities or consequences of alternatives.

Suggestions for Further Reading

There is an extensive literature concerning interpersonal comparisons. Elster and Roemer (1991) is an excellent recent collection. Other notable contributions include Robbins (1935, ch. 6); Harsanyi (1977b, ch. 4); Davidson (1986); Gibbard (1986); MacKay (1986); and Hausman (1995).

The inconsistency between rationality and the Pareto principle in circumstances of uncertainty is proven and discussed in Hammond (1983); Broome (1989); Seidenfeld et al. (1989); Levi (1990); and Mongin (1995). The most extensive discussion of the ethical appeal of Pareto optimality has been concerned with Sen's paradox (see Section 13.4 and the suggested readings for Chapter 13).

For critical discussion of the argument in favor of cash over in-kind provision of benefits, see Hochman and Rogers (1969); Okun (1975); Thurow (1977); Kelman (1986); and Blackorby and Donaldson (1988).

For general discussions of cost–benefit analysis see Mishan (1971, 1981); Sugden and Williams (1978); Ng (1983); and Layard and Glaister (1994). For more sophisticated treatments of compensation and net benefit, see for example Drèze and Stern (1990). Especially via the work of Richard Posner (1979), the compensation criterion has had a major impact on legal theory – see also Coase (1960); Baker (1975); and Coleman (1984). Coase's work has spawned a line of thinking that proposes to contribute to economic welfare not by government provision but rather by clearer rights assignments in order to make feasible a wider range of private enterprise. See Coase (1960); Demsetz (1964, 1967); and Cowen (1988).

PART THREE

LIBERTY, RIGHTS, EQUALITY, AND JUSTICE

Though well-being is obviously important, people also care deeply about freedom, rights, equality, and justice – indeed, people risk their lives pursuing them. Economists would be wrong to assume that welfare is all that matters. Even if economists themselves had no interest in anything except welfare, they would still need to understand these other values in order to understand the goals of policy makers and to help devise policies that achieve these goals.

But it is patently not the case that economists are indifferent to these other moral concerns. On the contrary, economists care as deeply about freedom and justice as do other people. These other values may be difficult to understand, but welfare is not a simple notion either. In assessing economic outcomes, processes, institutions, and policies, economists need to be sensitive to the whole range of moral concerns.

In Part III we shall discuss many important moral notions under the headings of liberty and rights (in Chapter 10), equality (Chapter 11), and justice (Chapter 12). Our way of classifying the issues is an expository convenience, and readers should not take it too seriously. Nor should readers take our discussion of these important moral notions as exhaustive. For example, the deepest moral commitments of many of those concerned to protect wilderness areas are not captured in these chapters. What is important to keep in mind is that there are many important moral notions that are distinct from the notion of well-being.

At the same time, we shall also discuss three kinds of moral *theories*: libertarianism (in Chapter 10), egalitarianism (Chapter 11), and contractualism (Chapter 12). It is natural to link libertarianism with liberty and rights, because liberty and rights are the central concepts in libertarian theories. But liberty and rights are also important to nonlibertarians. It makes sense to discuss egalitarianism while discussing concepts of equality, but concepts of equality are important in theories of justice that are not egalitarian. It is natural to place the discussions of justice and contractualism in the same chapter, because contractualist theories focus on justice. But justice is important in all moral theories. We want particularly to emphasize that the importance of the concepts introduced in this part goes beyond the importance of the particular theories that characteristically invoke them.

TEN

Liberty, Rights, and Libertarianism

Many people have a passionate commitment to individual liberty. They may object to paternalistic laws (such as those laws requiring the use of seat belts) as infringements on freedom, regardless of whether these laws increase welfare. Many would favor protecting the freedom of those with unusual lifestyles or unpopular religious convictions even if it means diminishing welfare. Social deliberation, unlike welfare economics, often treats protecting freedom and enhancing welfare as independent goals.

It is ironic that normative economics focuses on welfare, because economists value individual freedom very highly. Indeed we would conjecture that economists value freedom more than do most noneconomists. When economists have criticized socialism, for example, they have not only questioned whether it can "deliver the goods," they have also argued that economic and political power must be kept separate in order to protect individual liberty (Friedman 1962, ch. 1). Economists also value the prosaic liberties that are part of market life, such as the freedom to change jobs, to start a business, or to move from place to place. Thus, for example, many economists favor effluent taxes or exchangeable emission rights over direct state regulation of pollution. They do so not only because the taxes or exchangeable emission rights are purportedly Pareto superior to regulation but also because state regulation limits freedom directly and, by increasing the reach of government, threatens freedom indirectly. Similarly, economists often favor cash over in-kind transfers because cash transfers leave the recipients freer.

The efficiency case has been considered a part of "scientific" welfare economics because it has appeared to rely only on uncontroversial moral premises, while the argument in terms of freedom has been reserved for "unscientific" essays because its moral premises have appeared to be of a more philosophical nature. Contentment with the efficiency standard has doubtless been fostered also by a belief that the Pareto standard itself promotes liberty because it values outcomes that best accommodate the voluntary choices of individuals. But Pareto optima do not always respect rights or maximize liberty. Consider a problem such as homelessness. Though those with no place to live are almost always badly off, one will not begin to appreciate what is at stake if one does not consider what homelessness implies for individual liberty or the questions it raises about justice.

The concerns about liberty that underpin economists' normative judgments should be incorporated systematically into their methods of evaluation. Doing so raises problems of *definition,* of *moral justification,* and of *weighting.* In other words: (1) What is freedom? (2) How can claims about the moral importance of freedom be defended? (3) How can liberties be incorporated along with other values into a coherent scheme of ethical evaluation?

10.1 Freedom

We find it useful to follow Gerald MacCallum's (1967) suggestion and regard "freedom" as a relation among three things: (1) an agent, (2) obstacles or costs of some kind, and (3) objectives or possible outcomes. An agent is free *from* some obstacles or costs *to* doing or becoming something. An agent is free in some particular regard, when there are no obstacles of a specific kind preventing or discouraging the agent from doing something. For example, though *X* may not be legally free to purchase beer, *X* may be legally free to purchase bread. Yet *X* may not be economically free to purchase either, because *X* lacks money; or, if *X* is stranded on a deserted island, *X* may not be physically free to purchase anything. Rather than thinking of different meanings of the term "freedom," it is helpful to think about different kinds of obstacles to different sorts of states or actions.

People are free and unfree in various ways. Some freedoms are important while others are not, and there are substantive disagreements about which freedoms are more important. In his influential essay entitled "Two Concepts of Liberty" (1969), Isaiah Berlin argued that social policy should aim to protect "negative" liberties rather than to promote "positive" liberty. One protects negative liberties by respecting a protected sphere around individuals and placing few obstacles in the way of their efforts to satisfy their preferences. Negative liberty has nothing to do with natural or internal impediments to action. Promoting positive liberty, in contrast, involves attempting to realize some idealized conception of the self as rationally self-determining.

Libertarians often hold that the only limitations on freedom that are of social concern are those that violate pre-existing civil and property rights. Consider the difference between parents paying a ransom to a kidnapper of their child and parents paying the same amount to purchase a life-saving operation for their dying child. In neither case are the parents free to keep their money and to have their child live. Yet libertarians would maintain that only the first case involves a limitation of freedom, because the "obstacle" in the second case does not arise from a violation of pre-existing rights. Similarly, libertarians maintain that the obstacles faced by those who are starving are not infringements on liberty unless they are the result of rights violations. Libertarians who see competitive markets as always promoting freedom do not deny that market outcomes sometimes place barriers in people's way. They deny instead that these barriers, as the unintended consequences of voluntary interactions, are morally objectionable.

As we will discuss further, explaining what rights exist and what makes them legitimate is a large challenge for libertarians. For now, though, we will note that many people, including many scholars of liberty, do not have such a narrow view of which freedoms matter. Berlin himself argued that the barriers unintentionally placed in one's way by others may be morally significant limitations on freedom. Most would regard *any* obstacle to agents' pursuit of their objectives as morally significant if it results from the actions of others or if it could be removed by others.

The extent of an individual's freedom depends on how many possibilities are open to her, how easy they are to attain, how important in her plan of life these possibilities are, how far they are influenced by the actions of others, and how the society at large values these possibilities. Assessing the extent of liberty in a society requires, moreover, that one aggregate the liberties of different persons (Berlin 1969, p. 130). This is a mind-boggling amalgam of disparate elements. Some philosophers have suggested ways of cutting down the complexity. Rawls, for example, distinguishes a relatively short list of "basic liberties," including for example freedom of conscience and freedom of expression. Then, within this short list, he further distinguishes between "liberty" as a formal notion and "the worth of liberty," which depends on having the means to exercise the freedom. Billionaires and beggars are both free to vote and to speak freely on behalf of candidates and issues that matter to them, but the worth of their liberties differs drastically. Others have similarly distinguished "formal" from "effective" liberty. Sen's "capabilities" (see Section 8.5) also correspond to morally significant freedoms. For example, if one is literate then one is free from intellectual obstacles to reading. This use of the term "free" leads to an important political point: a regime that countenances illiteracy has much in common – from the standpoint of liberty – with one that bans books.

The concept of autonomy or self-determination is important in thinking about freedom. Autonomy can be used to refer to a *capacity* to govern oneself, a *condition* of self-governing, an *ideal*, or a matter of moral *authority* (Feinberg 1986, p. 28). The most important of these notions are the capacity to be autonomous and the condition of autonomy (Raz 1986, pp. 204–5). It is easier to identify factors that may impair the capacity to be autonomous than to list the factors that enhance the capacity. External factors such as poor or deliberately misleading information, insufficient schooling, inadequate opportunity for reflection, and a limited range of alternatives among which to choose threaten the capacity to be autonomous. Internal threats to autonomy include limited abilities, such as illiteracy, and psychological defects, such as an inability to cope with uncertainty.

The condition of autonomy is just as complex. It involves satisfying one's own preferences rather than having them satisfied by others. It involves actively accomplishing things rather than merely enjoying well-being. Autonomy also involves aspects such as self-possession, individuality, authenticity, integrity, self-control, initiative, and responsibility (Feinberg 1986, pp. 32, 40–4). The condition of autonomy may be valued simply as a matter of fact: most people want to possess the traits just mentioned. Alternatively, autonomy can be defended as a partial ideal of the person – as something people should care about even if they don't. This latter view has the potential to pose a threat to negative liberty and political freedom.

Autonomy as an ideal of the person is often connected to some notion of the "real interests" or "true needs" of the person, which governments should promote. In famous passages from *The Social Contract*, Rousseau wrote:

> anyone who refuses to obey the general will shall be forced to do so by the whole body; which means nothing more or less than that he will be forced to be free .... [A]long with the civil state, man acquires moral liberty, which alone makes him truly master of himself; for the impulse of mere appetite is slavery, and obedience to self-imposed law is liberty. (1968, p. 64)

The defender of autonomy in this sense thus sees the agent's significant objectives not as whatever the agent prefers (whether or not it is "the impulse of mere appetite") but what is truly in the agent's interests. Erecting barriers may actually enhance autonomy by removing subtler obstacles that stand in the way of doing what is truly in one's interests, just as a law against jaywalking may help pedestrians reach their destinations (MacCallum 1967, p. 330). Promoting autonomy (like promoting welfare) may thus, as Berlin warned, conflict with promoting negative liberties and a wide range of choice.

An analysis of morally significant individual freedoms must take a stand concerning how important various objectives are. It might allow the individual's preferences to determine the importance of actions and states of affairs, but this is as controversial and substantive a moral decision as is any specification of the "true interests" of individuals. In addition to determining what matters and how much, an analysis of morally significant individual freedoms must specify what kinds of obstacles are of moral concern. The freedom that matters is an absence of obstacles of the specified kind to the pursuit of significant objectives.

10.2 What Are Rights?

Though it is difficult to give a precise definition of a "right," either in law or in morality, one might roughly characterize a right as protecting or promoting a way of acting or a way of being treated. A right specifies that a range of actions or conditions open to an individual – such as driving, voting, or being sheltered – should be protected or promoted (see Martin 1993). Such normative direction may be embodied in law or convention, or it may have the status of a moral demand or recommendation.

Rights are often linked to liberties – and for good reason, since one kind of liberty (e.g., the freedom to emigrate) simply *is* a right and since notions of freedom are important in analyzing rights and in making sense of their rationale. Although the details need not concern us, legal analysts following Hohfeld (1923) have recognized that rights involve complex clusters of permissions and constraints regarding the actions of individuals. Rights typically involve both "privileges" for the right-holder and correlative duties for others. Rights that are the flip side of the duties of others are called "claims." For example, the right of workers to strike after their contract expires implies both a privilege to strike and a claim-right – that is, a duty on the part of others not to interfere in certain ways. Rights also involve "powers" and "immunities." For instance, occupational safety and health laws provide workers with immunities from efforts by firms to assign them certain duties, thereby limiting the powers of firms to determine aspects of working conditions. Rights are clusters of claims, privileges, powers, and immunities, though different elements may be central in different rights. Claims are central in what are called "rights," and privileges are central in what are called "liberties." But rights are not just claims, and liberties are not just privileges.

Legal rights are probably the most familiar kind of right. For example, the legal right to strike consists roughly in there being laws permitting workers to strike and prohibiting others from preventing them. Other sorts of *conventional* rights may be less familiar, but they are equally straightforward. For example, company policy may give workers the right to seventeen days of paid vacation. Parents have (or used to have) a conventional right to their children's respect.

Moral rights are more puzzling. It is tempting but misleading to think of them as analogous to legal and conventional rights: Just as legal rights are determined by legislation, so moral rights are determined by "moral law." Unlike legal rights and conventional rights, some moral rights might not be acknowledged or respected. In what sense can unacknowledged moral rights be said to exist? The reason for speaking of moral rights is to assess legal and conventional rights – for example, to criticize the former legal rights of the white minority in South Africa or to urge that rights to health care ought to be recognized by laws or conventions in the United States. To say that an agent has a moral right is, we maintain, to say that the agent ought to have a legal or conventional right (Mill 1863, ch. 5; Sumner 1987).

10.3 The Importance of Rights

Rights serve the dual roles of safeguarding people's interests and of providing them with control over their choices. Very often these two purposes will coincide, since people usually try to choose what is in their interest and regard themselves as having an interest in controlling their own choices. So, for example, giving an artist a property right in a painting she has created serves her interest in being able to profit from the painting and at the same time gives her control over its disposition.

Sometimes these two goals – of safeguarding interests and of providing control – will differ. For example, infants and nonhuman animals have interests but cannot make choices, and the point of assigning rights to such beings is apparently to protect their interests. One may, however, want rights in order to recognize the value people attach to controlling their own choices, rather than simply to safeguard their interests. Suppose for example that someone makes use of a vacant private condominium owned by Britney Spears without her permission. When should such a trespasser be said to have violated her property rights? So-called *liability* rules would say that paying Britney Spears a standard rental fee fully respects her property rights. Liability rules protect her interests, but they do not protect her right to control. One may therefore prefer *property* rules, which hold that her rights have been violated unless one obtains her consent. Most theorists argue that either an interest-based ("benefit") or a control-based ("choice") conception of rights is the valid one (Calabresi and Melamed

1972). However, we are attracted to Coleman and Kraus's (1986) view that, depending on facts and circumstances, a particular right may protect interests or choices or both. Among the factors that determine which conception best applies to a particular case are the value for the owner of preserving her choice and the costs to the "invader" of gaining consent. For example, a desperate traveler breaking into a remote cabin on a frigid night might be required to compensate the owner but might be judged not guilty of trespass.

Rights are important in economics for at least three reasons. First, clear definitions of rights, especially of property rights, promote economic efficiency. A familiar result in transaction-costs economics, well illustrated by current problems in the former Soviet Union, is that the absence of clear property rights results in socially wasteful efforts to defend ambiguous claims and discourages investment, because investors cannot be sure that they will reap the rewards of their investment.

Second, rights (and again especially property rights) are often taken as starting points in economic analysis. For example, analyses of the distributional properties of economic arrangements often begin from a given allocation of property rights. In adopting such a starting point, economists may find themselves assuming uncritically that this starting point is morally justified – that the rights recognized in the existing property regime really are the "true" rights in some normative sense. A crucial controversy among rights theorists that is of great importance to economists concerns whether there are "welfare rights" – that is, rights to things such as food, housing, and health care. We return to this question shortly.

Finally, rights may be invoked to limit the pursuit of economic goals. Most people today would recognize that the institution of slavery violates human rights and should therefore be forbidden, regardless of whether it might contribute to economic growth. More controversially, one can criticize socialism on the grounds that it violates people's rights by outlawing "capitalist acts between consenting adults" (Nozick 1974, p. 163) or criticize laissez-faire capitalism on the grounds that it violates people's rights to a job or their rights to minimum subsistence. The legitimacy of welfare rights is a primary issue regarding this third role of rights. Should the public provision of health care to those unable to pay for it be viewed as a matter of a *right* legitimately claimed by the needy, or is it instead best viewed as a matter of *charity* based on and limited by the goodwill of the affluent? Answers to this question will depend on the kind of justification offered for rights. Even if there are no welfare rights, governmental provision of charity might still be justified on the grounds that people prefer that welfare benefits be provided but need state action to eliminate the temptation to "free ride" on the charitable efforts of others (Friedman 1962; Hochman and Rodgers 1969).

10.4 The Justification of Rights

Why should there be moral rights, and what rights should there be? How can moral rules, which are independent of actual law or custom, have prescriptive force? One traditional answer is that human beings can "perceive" what the correct moral rules are and recognize that they must be followed. John Locke put it this way:

> The State of Nature has a Law of Nature to govern it, which obliges every one: And Reason, which is that Law, teaches all Mankind, who will but consult it, that being all equal and independent, no one ought to harm another in his Life, Health, Liberty, or Possessions. (1690, sec. 6)

Similarly, the authors of the American Declaration of Independence regarded their claims concerning rights to be self-evident. The problem with such a view is that it leaves one with nothing to say to those who fail to acknowledge this "plain and intelligible" law of nature except that they must be "biased by their Interest, as well as ignorant for want of study" (Locke 1690, sec. 124). Immanuel Kant gives a second answer: equal respect for rational agents generates moral rules and rights. This view is notoriously difficult to interpret and is tied up with Kant's metaphysics, which we cannot pursue here. A third answer, which is related to Kant's, is that rational individuals who are concerned to promote their interests would choose to adhere to some set of moral rules. We will explore this contractualist answer in Chapter 12. A fourth answer is that of the utilitarian: specifying and protecting certain individual rights maximizes total welfare. There are other possibilities, too, but these are the main alternatives in modern Western moral philosophy.

Each of these ways of justifying rights provides guidance concerning what set of conventional and legal rights agents ought to be granted. The four are unanimous in defending some rights, such as a right not to be killed or a right to enter into contracts. But these moral theories do not always agree. A full theory of moral rights cannot be developed without taking sides on fundamental questions in moral philosophy. From the perspective of those who would hope to employ the moral vocabulary of rights and liberties when evaluating economic institutions and outcomes, the indeterminacies within each of these theories concerning exactly what rights agents have are even more challenging.

10.5 Weighing Rights, Liberties, and Welfare

The remaining problem concerning rights and liberty is incorporating them systematically into schemes of moral evaluation. This problem takes different forms given different accounts of why liberty or rights are valuable. According to instrumental or consequentialist views such as utilitarianism, the problem of rights articulation can be formulated as one of selecting rules that maximize good consequences. This is straightforward in principle, although difficult problems of strategic coordination may arise in devising effective rules.

What about cases where rights protections are viewed as intrinsically valuable? One option here is simply to include the protection or violation of rights and liberties among the consequences, weighting them according to some scheme that reflects their moral importance. For example, other things being equal, a murder would be worse than a natural death, because the murder would involve a rights violation in addition to the welfare losses attached to a death. When a job applicant loses out because of employer discrimination, the bad consequences include not only the welfare loss entailed by the worker's unemployment but also the violation of the applicant's right to equal treatment. Yet this approach seems to capture only part of the value of rights protection.

When we discussed the rationality of duty or rights-based moral systems in Chapter 7, we touched on some of the difficulties associated with incorporating such factors into an overall moral evaluation. Recall the example of a witness who considers committing perjury in order to ensure a conviction in the Rodney King beating case. In this situation, committing one serious wrong will most likely not only increase total welfare but also prevent others from perpetrating more serious wrongs. If the police officers accused of beating Rodney King at his first trial had been convicted, then Reginald Denny would not have been dragged from his truck and severely beaten, and hundreds of store owners would not have suffered theft and arson. But many would say that perjury violates people's right to a fair trial and that it is simply wrong, regardless of how many rights violations it might prevent. Even those who would not go so far as to rule out perjury regardless of the consequences would maintain that the moral pressure on an individual not to "do evil" goes beyond an impersonal weighing of the consequences. The consequentialist view does not capture the sense that the moral imperative on me is that *I* should not commit perjury.

One way to preserve a more thoroughly nonconsequentialist version of rights protection is to adopt Robert Nozick's view of moral rights as absolute "side constraints" on action (1974, pp. 28–35). On this view, an individual's obligation is not to strive for a minimum of rights violations or for a maximum of any social welfare function. Rather, agents may do as they wish within the constraints imposed by rights. If everyone acts morally then there will be no rights violations. It is not, however, permissible for anyone to violate rights in order to prevent other rights violations. Rights are simply not to be violated. This absolute prohibition on trade-offs has been found disturbing by many philosophers (e.g. Nagel 1983), and we suspect most economists would agree. Indeed, Nozick himself seems to blanch before the full implications of his own view and suggests that, in the unlikely event that the consequences of respecting rights were catastrophic (as they might be in case of a famine), it might be permissible to violate them after all (1974, p. 30).

As we discussed before, it is possible to permit rights to have an important role without accepting the extreme side-constraint view that Nozick defends. One can allow the evaluation to vary with the perspective of the person doing the evaluating. Hence individuals can justifiably demur when offered the prospect of violating the rights of others, even if doing so would minimize overall rights violations. Furthermore, quite rigid constraints may be justifiable even from an explicitly consequentialist perspective. Although utilitarians and Kantians offer markedly different stories about *why* rights are necessary, they join in defending a set of rights that sharply constrain individual action and social policy.

10.6 Libertarianism

Libertarians defend political liberty, property rights, and economic freedoms. Some libertarians do so because they believe that defending rights and liberties most enhances welfare. However, we shall mainly discuss the work of "philosophical libertarians," who maintain that freedom, which depends on the protection of rights, is the overriding moral consideration (Lomasky 1987). As Narveson puts it, "the only relevant consideration in political matters is individual liberty" (1988, p. 7; see also Machan 1982, p. vii). In libertarian writings, liberty – the fundamental value – is linked to a rights-based view of justice, and the liberties with which libertarians are concerned should be regarded as rights. Philosophical libertarians maintain that acts and policies are just if and only if they do not violate anyone's rights.

What distinguishes philosophical libertarianism is this identification of justice with respecting rights as well as a distinctive view of the content of rights and of the duties that rights entail. For example, most libertarians hold that redistributive inheritance taxes violate individual property rights and are morally impermissible. But this presupposes both that rights must be respected and that property rights include the right to bequeath one's property without any encumbrances. If property rights did not include the right to bequeath one's property then there would be no rights violation, just as the "entailments" on landed estates requiring that they be willed entirely to the eldest son did not violate the legally recognized property rights of the English gentry. Libertarians defend strong and encompassing property rights. It might appear that property rights could conflict with such welfare rights as subsistence or health care and that to judge whether redistributive taxation is morally impermissible would require adjudicating among these conflicting rights. But libertarians deny that such conflicts among rights will arise because, they maintain, the duties that rights imply are almost exclusively "negative" duties not to interfere.

Libertarians typically deny that there are any welfare rights. For example, libertarians hold that right to life is a right not to be killed, not a right to be given subsistence. In propounding such a view, libertarians are not necessarily endorsing selfishness – although some, such as Ayn Rand's "objectivist" followers, do espouse egoism (Rand 1964). Most libertarians, like other people, value beneficence and charity as virtues. They only insist that individuals must not be forced to be beneficent or charitable. The libertarian need not admire Mr. Bumble and the other officers of the Parish who (in Dickens's *Oliver Twist*) allow paupers to starve to death. Like other readers, libertarians will find Mr. Bumble a hard-hearted hypocrite. But libertarians who accepted the legitimacy of existing property rights would see no injustice in the wretched treatment of the indigent. The indigent have no welfare rights to food or shelter. On the contrary, insofar as the workhouse relies on taxation to provide its miserable relief, it is violating the rights of taxpayers and unjustly taking their property. Justice requires that one not interfere with the pursuits of individuals unless those pursuits violate rights.

Although generally supportive of small government and low taxes (which are not synonymous and need not coincide), consistent libertarians would not readily endorse the slogans of those who criticize taxation on the mistaken grounds that people are entitled to all of their pretax income and wealth. There are three reasons. First, while a strong case can be made that private property is an essential element in any scheme that maximizes liberty, the legitimacy of the current distribution of property rights is dubious because it has been influenced by property rights violations in the course of history. Second, since libertarians support taxation for purposes such as defense, police, and courts, they cannot maintain that individuals are entitled to all their pretax earnings. Third, the actual income and wealth of individuals depend heavily on social and political institutions. Without a framework of law, a regulated money supply, and a variety of public goods from vaccinations to highways, there would be little wealth or income to protect from the "encroachments" of government. Even if many of these services were supplied privately rather than by the government, people must still pay for them; and to suppose that one's gross income measures one's own contribution and entitlement finds no support within libertarianism.

Even if one grants that existing property rights are legitimate, how can the libertarian defend the claim that protecting them serves freedom better than violating them in order to prevent starvation? The answer lies in distinguishing coercion from disutility. Taxing people to feed the starving involves coercion, while starvation (unless intentionally inflicted on someone) does not. A crucial step involves adopting a "moralized" definition of coercion, which states that *A* is coerced only when *unjustified* force is used against *A*. This means that the enforcement of justified property rights through legal authority – even when the result is starvation – does not involve coercion. If one thinks of negative liberty as the absence of coercion of this kind and if one takes existing property rights to be legitimate, then a system that enforces existing property rights maximizes negative liberty.

This argument is questionable because it is doubtful whether existing property rights are justified and because the moralized definition of coercion is dubious. To claim that coercion exists only in the context of unjustified force would imply, uncomfortably, that putting a convicted criminal in prison involves no coercion. More generally, linking coercion and liberty conceptually to the definition of rights makes it impossible to conceive of rights as maximizing liberty or minimizing coercion, for to do that one needs conceptions of liberty and coercion that are independent of rights conceptions (see Martin 1993).

Some libertarians, such as Loren Lomasky and Hillel Steiner, defend some rights to positive assistance, such as a right to be given a fair trial or an infant's right to nurture (Lomasky 1987, pp. 164, 260). But even libertarians who grant the existence of such rights want to limit them narrowly. Libertarians are concerned to ensure (1) that the duties to provide these particular benefits, like the duties not to interfere, be easy to comply with and not compromise the character of an individual's projects, and (2) that the benefits be crucial to the possibility of pursuing any projects at all. But some rights to positive assistance, such as children's right to nurture (which is certainly crucial), impose serious burdens on others. Libertarians are uncomfortable about imposing such burdens, because they insist that value lies in the separate endeavors of individuals and that justice lies in not interfering with those endeavors unless they threaten the rights of others.

Taking freedom to be the fundamental value does not automatically commit one to a rights-based view of morality, to a particular view of the content of rights, to the view that rights rarely obligate others to give positive assistance, or to the view that enforcing existing property rights necessarily maximizes liberty. Joseph Raz, for example, argues at length that valuing autonomy commits one to an extensive role for government in providing the public goods that facilitate the achievement of autonomy (1986, ch. 8). Raz's argument depends unsurprisingly upon an interpretation of liberty that most libertarians would reject.

In Nozick's well-known version of libertarianism, natural rights – rights whose justification does not depend on consequences – secure individual autonomy. Justice is respecting rights. According to Nozick's entitlement theory of justice, an outcome is just if it arises from just acquisition of what was unowned or by voluntary transfer of what was justly owned. Just acquisition is acquisition of unowned resources that makes no one worse-off and violates no rights. Transfers are, in the relevant sense, voluntary if and only if none of the limitations on individual choices arise from rights violations. Only remedying or preventing injustices can justify redistribution or other interferences with voluntary action. A view such as Nozick's clearly places heavy demands on a theory of rights, which Nozick unfortunately leaves undeveloped (see Nagel 1983). Justice is entitlement and depends on the actual history, not on the resultant pattern. Since the past is, as a matter of fact, full of injustices, Nozick's libertarianism in practice requires extensive redistribution, though how to bring about the state of affairs that would have obtained if there had been no past injustices is anybody's guess. Whether Nozick's entitlement theory rules out redistributive taxation as unjust depends on what particular view of the content of rights one adopts. If just acquisition gave one the right to transfer only 80% of one's property, then a 20% transfer tax would violate no rights. Even if Nozick is right that justice is determined by entitlement – that is, by respecting rights – and not by "pattern," patterns might re-enter the story at a higher level when one is deciding what the content of rights should be (Pogge 1989, ch. 1).

Some of Nozick's most striking arguments in favor of this "entitlement view" show that even minimal efforts to regulate the pattern of distributive outcomes involve extensive interference in people's lives. In a much-discussed example, Nozick argues that taxing the income of a star basketball player like Wilt Chamberlain interferes with the choices of Wilt's fans, who would happily pay a premium to see him perform (1974, pp. 161–3). Since the point of rights is to secure liberty and to permit individuals to pursue their own projects, considerations of welfare never justify interference with individual rights. Since, in addition, Nozick takes rights to be absolute side constraints, even efforts to protect rights cannot justify infringements of rights. Thus Nozick argues that only an extremely minimal state can be justified. Notice that accepting the priority of liberty and rights over welfare does not by itself force one to accept the libertarian political program, since a more-than-minimal state might serve liberty better than a minimal state.

The philosophical libertarian's commitment to liberty is in principle independent of any welfare consequences. But libertarians would like it to be the case that protecting freedom also makes people better-off, and libertarians need to show that some of their more extreme proposals, such as privatizing all streets and highways (Block 1982), would not lead to disastrous consequences. Jan Narveson puts the point vividly: "those who toil in the libertarian fields devote the lion's share of their efforts to persuading us that the alleged benefits of the State are illusory" (1988, p. 183). Moreover, many who support libertarian policies are motivated by welfare concerns. Friedrich von Hayek, for example, argues that economic prosperity, social innovation, and political democracy are best advanced by keeping government to a minimum (von Hayek 1967, 1976). Most libertarian economists are less influenced by distinctive philosophical commitments than by specific reasons for doubting the efficacy of government intervention to improve efficiency or redistribute income. For example, the literature on property rights and transaction costs that derives from Ronald Coase's arguments in "The Problem of Social Cost" (1960) usually reaches libertarian policy conclusions, but philosophical libertarians would reject Coase's consequentialist view of rights as welfare-enhancing devices to reduce transaction costs and to alleviate the suboptimalities caused by externalities.

Rights and liberties are central elements in the European and American moral and political heritage. They are complex and controversial, but so are conceptions of welfare. Normative economics should undertake more explicitly the responsibility of investigating how well economic arrangements serve liberty and to what extent they secure rights. Welfare economics should be only one part of normative economics, not the whole of it.

Suggestions for Further Reading

Isaiah Berlin's celebrated essay, "Two Concepts of Liberty" (1969), has provoked much discussion. A valuable collection of essays inspired by Berlin's work is Ryan (1979). MacCallum (1967) is a valuable commentary on Berlin. For discussions of autonomy, see especially Feinberg (1986, pp. 31–2), Raz (1986), Dworkin (1988, ch. 1), and Christman (1989).

On the definition of rights: Hohfeld (1923, p. 38) regards claims as "the limited and proper meaning" of "right." Wellman (1985), Sumner (1987), and Thomson (1990) follow Hohfeld terminologically, though none of them is committed to construing pure claims as the paradigm cases of rights. See also Feinberg (1973, chs. 4–6) and Waldron (1984). For a defense of a "choice" construal of rights see Hart (1955); Raz (1986) defends a benefit construal. For general discussions of rights see Dworkin (1977), Sumner (1987), Thomson (1990), and Martin (1993); and for a sustained discussion of the difficulites of defending libertarian property rights see Waldron (1990).

Amartya Sen (1982b) discusses the option of including rights violations among the consequences that a consequentialist should take into account. Thomas Nagel, on the other hand, has argued powerfully for preserving a nonconsequentialist component in rights evaluation (1986, ch. 9).

Machan (1982), Lomasky (1987), Narveson (1988), and Steiner (1994) – as well as Nozick (1974) – provide systematic accounts of a libertarian viewpoint. Friedrich von Hayek's treatise (1960) argues that a variety of consequentialist considerations support the case for limited government and wide protections for negative liberty. Schmidtz (1994) makes an eloquent case for the importance of property rights to human flourishing, which draws on the work of Coase (1960) and Demsetz (1967); see also Cowen (1988). For a critique of the "vulgar libertarianism" that holds that people are entitled to their pretax holdings, see Murphy and Nagel (2004). Von Mises (1941) is the classic libertarian discussion of economics.

ELEVEN

Equality and Egalitarianism

The contemporary world is full of inequalities that most people find profoundly disturbing. For example, the median wealth of American families of African ancestry is about one tenth that of families of European ancestry. In 2000, life expectancy at birth in Japan was about 81, compared to 38 in Angola. In 1970, the top hundred American executives earned on average 38 times what the average full-time worker earned; in 2000, the ratio was more than 1,000 to 1 (*Detroit News,* 19 December 2002). One percent of Americans own at least a third of all America's wealth, while the bottom 40% of the population has only 1%. In 1998, the assets of the world's 358 billionaires were greater than the combined incomes of countries with 45% of the world's population – more than 2.5 billion people (⟨http://www.globalpolicy.org/socecon/inequal/⟩). Inequalities in wealth also correlate with other inequalities. Unsurprisingly, those who are richer tend to have higher incomes. Those with higher incomes also tend to be healthier, better educated, of higher status, taller, and to have more political influence.

What's wrong with inequalities? Why do most people find these inequalities so troubling? Notice first that it is not just the size of the inequalities that makes them disturbing. The disparity of life expectancy between infants born with Tay-Sachs disease (who usually die before age 5) and those without the disease dwarfs the inequalities in life expectancy between Japan and Angola, but it does not raise the same moral problems. Why?

Second, notice that the inequalities mentioned in the first paragraph fall into two groups and are morally objectionable for different reasons. First, there are inequalities in the possession of particular goods across *individuals,* like the global inequalities in telephone ownership. Inequalities of particular goods seem to be of moral concern only when those goods are by themselves of central importance – like health, income, wealth, or social status – or when they correlate with enough other goods that they have a major impact on people's lives. The inequalities across individuals in the ownership of particular goods or of correlated goods that are of moral concern must be pervasive. Isolated inequalities, such as inequalities in boat ownership, are in contrast of little moral interest. Second, there are inequalities of particular goods across *groups,* like inequalities in life expectancy between Japan and Angola or inequalities in wealth between Americans of European and African ancestry. Inequalities across groups are of concern both for their consequences and for their causes. The huge inequalities in life expectancy reveal a pervasive inequality between the prospects of people born in different countries, and they strongly suggest that avoidable social factors cause people to die young in countries such as Angola. The enormous wealth disparity between Americans of African and European ancestry provides evidence concerning the magnitude of both the legacy of past injustice and of the hurdles that contemporary African Americans face.

Though inequalities seem to be of enormous moral importance, it is not easy to conceptualize what (if anything) is wrong with inequality itself. At the same time – as Ronald Dworkin, James Griffin, Will Kymlicka, and Amartya Sen have all argued – the concept of equality lies at the heart of moral theories. Sen argues that all moral theories prescribe that people be treated equally in some regard (1992a, p. 3). Utilitarianism requires that everyone's welfare count equally. Even libertarian theories require equality of rights. What makes moral theories so different is that they would equalize different things. Dworkin, Griffin, and Kymlicka argue that the notion of "equal respect" defines what morality is all about. In this view, conflicts between different political philosophies can best be understood as conflicts over the interpretation of equal respect (Dworkin 1977; Kymlicka 1991).

Appeals to equality play a no less crucial role in discussions of economic policy. One reason for objecting to the massive tax cuts secured by the administration of G. W. Bush is that they contribute to inequalities in wealth and income. One reason to support the same administration's education policies is that they stand a chance of diminishing educational inequalities. Traditional welfare-state programs have attempted to lessen inequalities in incomes and status, and concerns about inequalities have been among the most important justifications for interfering with market outcomes. To judge whether policies to promote equality are just, one needs to understand how the causes, characteristics, and effects of the way in which goods (such as wealth, income, status, health, and education) are distributed link up with the underlying concerns of egalitarians.

An example may reveal some of the complexities in the notion of equality. One consequence of inequalities in wealth, income, and social services in the United States is that there are many thousands of homeless people. The homeless population can be roughly classified into three groups. The first group consists of those who have lost their foothold in the economy through disruptions such as firms relocating to other states. In the second group of homeless are people with mental, physical, or behavioral disabilities. A shortfall in residential treatment, outpatient care, or occupational rehabilitation has left these people homeless. The third group consists of people without any identifiable disability who have had only a casual connection to the labor force.

Does the value of equality urge that members of these groups be treated similarly or differently? Do the similarities between those who have lost their jobs "through no fault of their own" and those who cannot get jobs because of handicaps require that the two groups be treated equally, or do the differences demand that they be treated unequally? Should an egalitarian support treating those with congenital disabilities differently from those who may have suffered damage from drug addiction? Are those in the third group, the casual workers, cursed with an inability to muster the effort to sustain regular employment, or are they just lazy – and how does the answer affect what equality demands? The public debate about homelessness in the United States takes for granted that policy concerning the homeless should depend on what causes homelessness. Why? How are the causes of inequalities in wealth and income that result in homelessness relevant to their ethical evaluation? One needs to understand what's wrong with inequalities and what sort of equality one seeks.

As this example suggests, it is hard to judge which economic differences between people are morally justified. Should equal wages be a goal when workers may have to provide for families of different sizes? Is the value of equality served by transferring income to indigent surfers from those who work hard and are consequently relatively affluent? Notions of equality of opportunity, of circumstance, and of treatment are just as problematic. Do laws that prevent convicted felons from voting involve equal or unequal treatment? Does providing a more expensive education to those from deprived backgrounds render opportunities more or less equal? Are affirmative action programs an implementation or a violation of equal opportunity?

To determine what to equalize, one needs to understand what one wants equality *for.* What's wrong with inequalities? Determining what to equalize and determining why the pursuit of equality is morally justified are inseparable inquiries. There are also serious problems when the pursuit of equality comes into conflict with the constraints imposed by rights or with other objectives such as welfare or liberty. So, as with the discussion of liberty and rights in the previous chapter, here too we face problems of definition, justification, and weighting. What should be equalized? Why should it be equalized? How important is equality? We shall begin with the second of these questions.

11.1 Why Equalize?

In order to decide *what* to equalize, it helps to know *why* one wants to equalize anything at all. The staggering differences in life prospects depending on whether one is born on the Upper East Side of Manhattan or a couple of miles north in East Harlem seem unjust. But other differences, which may be just as large, are unproblematic. For example, people are not equally tall or equally articulate. Though such differences may reflect differences in nutrition or education, which raise moral qualms, differences in stature or articulateness do not themselves appear to be unjust; and one might well maintain that a world without such differences would be worse rather than better. Egalitarians do not want to lessen or eliminate all differences among people. Some differences are morally objectionable, while others are acceptable and still others are desirable – but which ones, and why? Furthermore, judgments about how much to equalize and about how much weight equality should have plainly depend on the role of equality in overall schemes of moral evaluation.

A commitment to moral and political equality is deeply embedded in American political traditions and in Kantian moral philosophy, but the basis for this commitment and its implications for economic equality are controversial. Why should equality matter morally? David Miller poses this question as follows.

> Why should *equality* be thought desirable? Equality after all means a leveling of differences; it means a smoothing down of irregularities or idiosyncrasies. Although I may from an aesthetic motive decide to trim my rose bushes to an equal height or polish my wine glasses to an equal shine, to treat people in such a way would be at best perverse and at worst immoral. The pursuit of equality seems to be impaled on a fork: either the ultimate end of the pursuit is not equality at all but some other value or values which have become confused in the popular mind with equality, or our societies are aiming at a goal that cursory inspection reveals to be quite monstrous. (1982, p. 73)

We shall comment on three responses to this puzzle.

11.1.1 Equality Is Intrinsically Good

The first and most direct response has been given by philosophers such as Larry Temkin, Derek Parfit and G. A. Cohen. In Cohen's view, the purpose of egalitarianism "is to eliminate *involuntary disadvantage*" (1989, p. 916). Parfit writes: "The Principle of Equality ... claims that it is bad if, through no fault of theirs, some people are worse off than others" (1984, p. 26). In Temkin's view, for some to be worse-off than others undeservedly – that is, through no fault or choice of their own – is "comparatively unfair" (2003, p. 767). So, for example, Temkin maintains that it is unfair if John is crippled because a tree limb falls on him while he is out for a walk, since millions of others go for walks without tree limbs falling on them (pp. 772–3). In Temkin's view, the badness of inequality is a fundamental moral intuition.

We doubt that "undeserved inequalities are bad" is a fundamental moral principle, and indeed we doubt that it is always true. It seems appealing, but this appeal may arise from conflating one's attitudes toward inequalities with one's reactions to the suffering and bleak life prospects that in fact accompany the inequalities we observe in the world around us. We reject Temkin's intuition in part because we reject its implication that natural occurrences are subject to moral appraisal. The tree and the wind do not treat John unfairly when they drop a limb only on him. Only those things that are caused by or could be remedied by individual or social action are subject to moral judgment. Furthermore, even if John's injury could be remedied, we doubt that the *inequality* – the fact that he is injured while others are not – is itself intrinsically bad. Picture a paradise in which everybody's life is wonderful yet those who have sunnier dispositions are better-off. If resources could be reallocated to lessen this inequality, would it be unjust to leave things alone? Since inequalities are in fact so closely linked to suffering on the part of those who have less, we doubt that people have clear intuitions about circumstances in which this link is broken.

In our view (unlike Temkin's), inequalities in material circumstances do not link up directly with morality or justice. When inequalities in life prospects, happiness, resources, or wealth are unjust, as they often are, this is not because there is something good about people doing just as well or just as badly as one another. As already suggested, the inequality may be bad because the prospects of those who are worse-off are bad. Furthermore, and more fundamentally, inequalities in prospects and outcomes may conflict with the equal respect toward one another that morality demands. In our view, the real connection between equality and morality lies not in the comparison between our goods but in our moral standing – in our authority "to make claims and demands of one another as equal free and rational agents" (Darwall 2004, p. 43).

11.1.2 Equality and Priority for the Worst-Off

Rather than taking equality to be intrinsically good, one might argue that the value of equality is instrumental to or inherent in other values. In the contemporary world, inequalities usually involve misery on the part of those who are worst-off. That misery provides a strong moral reason to object to the inequality, though not on the grounds that there is anything intrinsically bad about the inequality itself. What is shocking about the disparity in life expectancy between Japan and Angola is not that the difference is 43 years but rather that life expectancy in Angola is only 38. If life expectancy in Luxembourg were 124, the 43-year difference between Japan and Luxembourg would not be nearly as troubling. Though Mozart died at 37, a life expectancy of 38 is grim; and the other disadvantages faced by people in Angola – such as poverty, malnutrition, disease, lack of education, and insecurity – are (being remediable) morally objectionable.

Though it has no necessary connection to egalitarianism, a benevolent humanitarian concern for those whose life prospects are poor is surely one of the main moral grounds supporting egalitarian policies. In Joseph Raz's words,

> what makes us care about various inequalities is ... the hunger of the hungry, the need of the needy, the suffering of the ill, and so on. The fact that they are worse-off in the relevant respect than their neighbours is relevant. But it is relevant not as an independent evil of inequality. Its relevance is in showing that their hunger is greater, their need more pressing, their suffering more hurtful, and therefore our concern for the hungry, the needy, the suffering, and not our concern for equality makes us give them the priority. (1984, p. 240)

Egalitarianism can thus reflect simple benevolence or even a utilitarian commitment to maximizing well-being. But, as Derek Parfit argues, it might instead reflect the view that those who are worse-off should have a certain priority over those who are better-off. Like those who are benevolent, those who hold such a view – whom Parfit (1991) calls "prioritarians" and Temkin calls "extended humanitarians" (1993, ch. 9) – care about those who are badly off, not about distribution itself, though they will in practice seek to lessen inequalities. It can be difficult to describe examples where the policies endorsed by egalitarians and prioritarians diverge, and Fleurbaey (2006) argues convincingly that prioritarians are a species of egalitarians who differ from other egalitarians with respect to the justification of egalitarianism and the measurement of inequality. From the point of view of policy, prioritarians and egalitarians differ mainly in their views concerning how to measure inequality.

The debate over "equality versus priority" is posed by Parfit as a dispute concerning the distribution of well-being, but the debate could be reframed in terms of other egalitarian "currencies" (as we shall discuss) such as resources, opportunity for well-being, or capabilities. In contrast to fundamentalist egalitarians who (like Temkin) maintain that undeserved inequality is in itself a bad thing, prioritarians are concerned about those who are worse-off. What matters to Parfit's prioritarian is an individual's "absolute" level of well-being, not the relative position in a distribution – with how well-off an individual is compared to how well-off he could be. So, for example, a prioritarian would place more moral weight on benefiting a middle-income Bolivian than the poorest Austrian if the former were at a lower level of well-being than the latter. Distribution is of no intrinsic importance to the prioritarian, but – in weighting the good of those who are worse-off more heavily than the good of those who are better-off – prioritarians, like egalitarians, will fight against inequalities.

As a justification for egalitarian policies, prioritarianism seems to us as arbitrary as Temkin's fundamentalist egalitarianism. When those who are worst-off are badly off, as in the comparison between Japan and Angola, there is a ready justification for helping the worse-off and thereby lessening the inequalities. But when those who are worst-off are not badly off – that is, when inequality obtains among people who are all very fortunate – why should one place less moral weight on the interests of those who are better-off? The problem in our view is that neither fundamentalist egalitarianism nor prioritarianism is deep enough to provide any satisfactory account of what is disturbing about inequalities. To explain what is wrong with inequalities by saying that equality is intrinsically valuable or that one should give special weight to the interests of the worse-off invites virtually the same question that one was supposed to answer: Why should inequalities matter? Why should improving the well-being of the worse-off matter more than improving the well-being of the better-off?

The misery, shame, helplessness, degradation, and enforced servility of those who are poor give one good reason to fight against inequalities in wealth. The limitation of freedom and the stunting of intellect and sensibility caused by poor education give one reason to fight against inequalities in schooling. The suffering of those who are sick and the shrinking of their lives give one reason to fight against inequalities in health. But what does any of this have to do with egalitarianism as a philosophical position or with the prioritarian view that the moral weight placed on people's interests should be a diminishing function of how well-off they are? A humanitarian concern for the plight of those who are badly off leads to a radical critique of existing inequalities, but it has no necessary connection with either egalitarianism or prioritarianism.

The fact that some goods are "positional" (Hirsch 1976; Frank 1987) might be misunderstood as lending support to prioritarianism or to the view that inequalities are intrinsically bad. Goods can be "positional" in two different ways. First, some things are positional goods because the benefits they provide depend on how much of them others have. An elite education is in part a positional good in this sense. Regardless of the excellence of a Yale education, part of its benefits depends on the fact that only a small number of people have a Yale education or its equivalent. Other things are positional goods in a second sense: how much of them one has depends on one's possession of positional goods in the first sense. Recent research has shown that health is a positional good in this second sense: how healthy one is depends on one's position in the socioeconomic hierarchy. Income is a positional good in both senses. How much of it one has depends on one's possession of positional goods such as education, and one's relative income (which depends of course on how much income others have) in turn affects how much of other goods one has.

Inequalities with respect to positional goods are worse for those near the bottom than are inequalities with respect to goods that are not positional. Since many important goods are positional goods, inequalities are worse for those near the bottom than they might appear to be. But this fact lends no support to prioritarianism or to the view that inequalities are intrinsically unfair. On the contrary, recognizing the importance of positional goods permits one to recognize a nondistributional cost of inequalities to those with fewer positional goods.

11.1.3 Intrinsic Connections between Equality and Other Ends

In answering his own question about equality, Miller emphasizes the intimate connection between economic equality and explicitly moral objectives. In his view, economic equality has an intrinsic connection to some of the ends it serves, just as negative liberty has an intrinsic connection to autonomy. Miller lists four different ends that equality (including economic equality) may serve and to which equality has an intrinsic connection. Our discussion also draws on a paper by Thomas Scanlon, "The Diversity of Objections to Inequality" (2003), that complements and expands Miller's account.

First, Miller maintains that *equality is sometimes required in order to be fair.* If there are benefits or burdens to distribute then, other things being equal, it is unfair to distribute them unequally. Other things are of course often not equal: fairness and justice involve more than equality. It seems unfair, for example, for all students in a course to get the same grade regardless of their performance.

Scanlon articulates some of the relations between economic inequality and fairness by pointing out that economic equality may be necessary to protect the integrity and impartiality of social institutions and practices. Large inequalities in wealth lead to disparities in political power and in the enforcement of laws. These disparities can call into question the legitimacy of courts and elections. Similarly, a large gap between rich and poor may result in the rich opting out of public education or state-supported health care, thereby degrading those institutions. Fair treatment and the protection of rights can also rule out economic inequalities. For example, it may be impossible for government to make good on its obligation to protect all residents from violence and theft if economic inequalities are permitted to grow large.

Second, Miller argues that equality is a good thing because *some measure of equality is necessary for self-respect.* Except when they have behaved stupidly or immorally, individuals should be able to say, "I am as good as anybody else; I may not be as clever or hard working as you are, but I am as good as you are" (Davies 1963, p. 45; quoted in Benn 1967, p. 69). The great inequalities that characterize the United States today make it difficult for individuals at the bottom to maintain their self-respect. These inequalities cause what Scanlon calls "stigmatizing differences in status." The homeless are not only impoverished and uprooted, they are also often regarded with suspicion, fear, and contempt by the more fortunate. In these circumstances it is difficult to preserve one's self-respect. Notice that this argument might not work if everyone were very fortunate (though not equally so).

Third, Miller points out that equality is a good thing because *equal treatment implements the duty to show equal respect.* "While [people] differ profoundly as individuals in capacity and character, they are equally entitled as human beings to consideration and respect" (Tawney 1931, p. 34). "[Man] knows himself as a man only in society with other men. He needs, as other animals do not, self-respect and the respect of others; he needs to be treated by others as if his needs and purposes mattered to them" (Plamenatz 1967,

p. 93). Showing equal respect to people is a different issue than whether those people can maintain their self-respect. Showing equal respect implies recognizing that, apart from young children and those with extreme mental disabilities, all people have capacities to deliberate for themselves, to engage in relationships and activities that are intrinsically valuable, and to develop skills and traits that are admirable (Miller 1982, p. 81; see also Lukes 1977). Inequalities like those that separate destitute peasants in less developed countries from yuppies in their million-dollar apartments are inconsistent with the acknowledgment that everyone has such capacities and worth, and they violate the duty to show equal respect.

Finally, Miller links equality to fraternity: *Economic inequalities undermine social solidarity* and create barriers to friendship, community, and love. "What is repulsive ... is that some classes should be excluded from the heritage of civilization which others enjoy, and that the fact of human fellowship, which is ultimate and profound, should be obscured by economic contrasts, which are trivial and superficial" (Tawney 1931, p. 139). Such barriers are related to the stigmatizing differences in status to which Scanlon objects. Here again Miller's argument may implicitly depend on the idea that there is some real deprivation and not only inequality.

To Miller's list, Scanlon adds one additional and important item. Large inequalities in wealth can subjugate some people to others, and inequalities are objectionable when they lead some people to dominate others. Inequalities need not have this result, and domination does not always result from economic inequalities, but the connection is strong enough to provide an additional reason to condemn them.

Even though Miller and Scanlon reject Temkin's view that inequality is intrinsically bad, they show that the objectives to which equality is a means – or the principles from which a certain measure of equality follows – are not extrinsic to many of our concerns with equality. Equality bears more than merely a causal connection to fairness, self-respect, equal respect, fraternity, and freedom from domination. Equality in distribution is not merely a means to fairness, equal respect, or fraternity; under many circumstances, it manifests these values. Though equality is arguably only a means to self-respect or freedom from domination, the concern for individuals it manifests is linked to these values.

11.2 Equality of What?

Understanding why equality is morally important helps one to see why certain kinds of inequality are deeply disturbing while others may be permissible. Yet close attention to specific egalitarian proposals brings out troubling ambiguities. For example, should people whose ability to feel discomfort has been numbed be allocated fewer resources on the ground that deprivation has less effect on them? Should society compensate those who are born with poor eyesight or poor coordination or poor ability to discipline themselves to work? On the basis of a near consensus that egalitarians should aim to nullify morally arbitrary inequalities among individuals, most contemporary egalitarians have focused mainly on the question of *what* should be equalized. Should it be (1) welfare, (2) resources, (3) opportunity for welfare, or (4) capabilities or "access to advantage"? Other egalitarians have, however, grown suspicious of the underlying consensus in this literature on mitigating those differences among people for which the individuals themselves are not responsible. They have questioned the links between inequalities in goods or opportunities and the underlying concerns with stigmatization, domination, and fairness. In this section we will focus on disputes concerning what to equalize. In Section 11.3, we shall consider whether the underlying concerns of egalitarians support the goal of nullifying morally arbitrary inequalities among individuals.

11.2.1 Equality of Welfare

Economists might naturally interpret egalitarians as aiming to equalize welfare, because welfare is the central moral notion in normative economics. But to interpret the ideal of equality as involving equality of welfare also discredits the ideal in the eyes of most economists. To claim that two people are equally well-off is to make an interpersonal utility comparison. As discussed in Chapter 7, economists typically doubt whether interpersonal utility comparisons are possible, and they do not want policy conclusions to depend on them. Even if their qualms concerning interpersonal comparisons could be met, economists would warn those who seek to equalize welfare about the effects of such equalization on incentives. If those who work hard wind up no better-off than those whose idea of hard labor is a walk from the hammock to the beach, the economy will suffer.

Furthermore, equality of welfare is not a very attractive ideal. Why should one want to make good the losses of those who like to gamble? Don't those who work hard and choose prudently deserve to be better-off than those who are lazy and thoughtless? Why should those who are unhappy without expensive wine have more of a claim on social resources than those who are content with cheap beer? Why should Tiny Tim's sunny disposition and modest wants cancel out his claims to an expensive wheelchair (Cohen 1989, p. 918)? In terms of the underlying moral values that equality may serve, one might make a case that equality of welfare would promote fraternity or equal respect and bar subjugation. But it is hard to see equality of welfare as promoting fairness or self-respect. For reasons such as these, few authors defend equality of welfare.

11.2.2 Equality of Resources

Ronald Dworkin argues that egalitarians should be concerned to equalize resources rather than welfare. "Resourcist" views emphasize the fact that outcomes depend both on resources and on people's ambitions, choices, and characters. Egalitarians should aspire to provide equal resources and nothing more. Owing to their choices and characters, some people will then do well and others badly, but that's up to them. Any further equalizing undercuts individual responsibility. The proper concern of an egalitarian thus should not be welfare but rather the resources people have available to pursue their ends. Rawls's primary goods, as all-purpose *means,* are resources. Moreover, as we shall see in the next chapter, his two principles of justice constitute one variant of resource egalitarianism.

Resource egalitarianism involves many complexities, since it would be inefficient – indeed, ridiculous – for everyone to have exactly the same bundle of resources. Down parkas are not needed in Tahiti, and even with global warming there's not much demand for surfboards in Greenland. What resource egalitarians want is for everyone to have bundles of resources that are "equivalent," not identical. One might say that resource bundles are equivalent if nobody would want to trade her bundle of resources for anyone else's. This "no envy" test provides one definition of what it is for resources to be equal. Dworkin suggests that we imagine the equilibrium that would result from trade in a hypothetical competitive market starting from a condition in which everyone had an equal share of every resource. Owing to differences in their tastes, people's resources in equilibrium would differ, but no one would want to trade for anyone else's.

It is possible to show, for certain simplified cases, that any resource distribution satisfying this no-envy condition could be reached through a sequence of Pareto improving trades that start from a condition in which everyone has the identical resources. Following Kolm (1972), Hal Varian (1974, 1975) has proposed the label "fair" for envy-free Pareto optima and has defended this conception of fairness as ethically appealing (see also Foley 1967; Baumol 1986). An important stimulus to the study of envy-free allocations of resources has been their apparent ability to provide a standard of equity that does not involve interpersonal comparisons of welfare.

But there are many problems, even in theory. The ethical appeal of resource egalitarianism derives from the intuition that, once people are equipped equally with resources, how much welfare they achieve is their own affair and not a matter for society to try to equalize. But depending on social institutions such as insurance, taxation, and market structure, there may be very large or relatively small disparities in outcomes even if one begins with equality of resources. Should egalitarians be indifferent to inequalities in outcomes, provided that resources are equal? Furthermore, outcomes depend on physical and intellectual abilities as well as on external resources and individual character and choice. Shouldn't a person's abilities count as resources, too? Merely equalizing external resources seems not to go far enough. Equalizing wealth would, for example, leave paraplegics effectively with fewer resources. Productive abilities and consumption capacities are also resources – internal resources – and egalitarians should think about how they might be equalized or devise some compensation for inequalities in internal resources.

If one counts abilities as resources, then how can one possibly equalize resources? One approach is to say that total resources – both internal and external – are equal whenever people have equal amounts of goods and equal amounts of leisure. If everyone has the same tastes, then this conforms to Varian's definition (1974, 1985) of a fair distribution: it is Pareto optimal and envy free. But if people's tastes differ, then outcomes involving equal goods and leisure will generally not be efficient, and the competitive equilibrium resulting from trades made after this equal-division starting point will not be envy free. Starting with equal endowments of leisure and goods, a highly talented individual will be able to convert leisure time into additional goods at a much faster rate than will a less talented individual. In competitive equilibrium, the less talented will wish they could swap their goods–leisure bundles for those of the more talented (Pazner and Schmeidler 1974; Varian 1985). This way of extending the definition of resource egalitarianism to cover the case of internal resources doesn't have the nice property that competitive market exchange preserves equality. This difficulty has led economists to put forward two alternative definitions of equal resources, versions whose "envy freeness" will not be disrupted by trade.

The first of these is a "contribution fair" (Baumol 1986) or "wealth fair" (Varian 1974, 1975) distribution. Imagine equalizing people's external resources but allowing them to retain the fruits of their differential talents. The resulting competitive equilibrium will not be envy free in the sense we have defined because, as just noted, those with more talent will tend to have consumption–leisure bundles that others envy. The equilibrium is, however, envy free in a different sense. Suppose one asks individuals whether they prefer to consume what others consume and also to make the same contribution to production. Given how hard it would be for them to contribute as much to production as do the very talented, those with few talents will not envy the whole consumption–production package. In other words, agents will not want to trade places if doing so entails not only getting the other's consumption bundle but also producing as much as the other person produces (but without the other person's talents). Contribution fairness leaves talents or internal resources out of the package of things to be equalized, and it accepts as fair those inequalities that are due to competitive rewards for talent. This falls short of what many people would think of as equality of resources.

Alternatively, one can define an "income fair" distribution (Pazner and Schmeidler 1974; Varian 1974, 1975; Baumol 1986). Suppose everyone owned equal shares in the labor power (and hence the talents) of everyone. These shares and all external resources are divided equally. If people do not buy back shares of their labor power (which means that they work almost all the time), then everyone has equal resources and, if tastes are the same, everyone is equally well-off. But tastes are not all the same and people cannot work all the time, so people will want to trade some of their initial resources and will want to buy back shares in their labor power so that they can have some leisure. If we define people's "full income" as the combined market value of their external resources, the shares they own in people's labor, and their leisure, then the resulting competitive equilibrium equalizes full incomes – hence the term "income fair."

Because the labor time of the talented is a high-priced commodity, it will cost a talented person more to obtain any given amount of free time than it will cost an untalented person. If tastes are constant across the community, then the talented, whose leisure is more expensive, wind up with fewer external resources and will generally be worse-off. Equality of resources interpreted as an "income fair" distribution winds up overcompensating the untalented in terms of equality of welfare. Dworkin calls this the "slavery" of the talented (1981b, p. 312) because, unlike those whose leisure is cheap, the talented will have no choice but to work long hours.

The formal egalitarian criteria we have discussed – the notions of fair, wealth-fair, and income-fair allocations – are only a small and not particularly representative sample of a wide range of technical construals of egalitarian objectives that have been developed over the past two decades (see Moulin and Thomson 1996; Fleurbaey and Maniquet 2005). Since these formal construals of egalitarianism do not limit themselves to utility information (Fleurbaey, Suzumura, and Tadenuma 2005), there are lots of possibilities, many of which involve complex formal theorems.

Rather than attempting to equalize ownership of internal resources, Dworkin proposes a hypothetical insurance scheme to compensate for inequalities in internal resources (1981b, pp. 292–323). Dworkin's thought is that, if it were possible before birth to contemplate such matters, individuals would be willing to purchase insurance that would compensate them in the event they have the bad luck to be disabled. Redistribution of income to improve the life prospects of those who are less well endowed with internal resources can be conceived of as implementing such hypothetical insurance rather than as equalizing internal and external resources. Economists have criticized the details of Dworkin's analysis of such a hypothetical insurance market (Roemer 1985; Varian 1985; Fleurbaey 2002), but the idea remains intriguing.

Internal qualities such as talents resemble resources, and resource egalitarians seek some way to mitigate the consequences of inequalities in their distribution. Yet Dworkin maintains that egalitarians should not interfere with inequalities that result from differences in people's ambitions, preferences, or persistence. Not everyone agrees. John Roemer suggests (1985) that everything may lie on the involuntary side of the line. After all, a person's willingness to expend effort, like the person's talents, may ultimately be traceable to their biological and personal histories. Why not then treat the capacity to be ambitious as a resource? For that matter, some people may possess enzymes that allow them to extract more nutrition from a meal or even more satisfaction from a movie. Following this route, one might argue that any difference in achieved welfare levels resulting from apparently equal resources is actually the product of hidden differences in resources. In this way, equality of resources would lead back to equality of well-being (Roemer 1985, 1986a,b, 1987; Scanlon 1986).

Roemer raises two kinds of issues involved in distinguishing resourcist from welfarist views. One concerns the gap between ideal reasoning about just distribution and practical reasoning that takes into account limits on information and problems with incentives. Attempting to equalize welfare would require detailed knowledge of people's preferences (which they would have an incentive to misrepresent), and equalizing welfare would lessen incentives to work. Given these problems, some simplified version of resource egalitarianism may be the best way to lessen inequalities in welfare. The second issue is that one needs a theory of persons and of individual responsibility in order to decide which internal differences justify compensation and which do not. Should egalitarians compensate those who have the misfortune of being smokers, or should they regard smoking as a matter of personal responsibility? It is obvious that people choose whether to smoke. But it is also obvious that individual choices cannot explain why Italians are more likely to smoke than Americans or why working-class Americans are more likely to smoke than upper-class Americans. Roemer (1998) makes the intriguing suggestion that one can disentangle the effect of sociopsychological factors (for which individuals deserve compensation) from the effect of individual choice by comparing an individual's behavior to the distribution of behavior among a relevant social group. On average, then, smokers who belong to a social group within which smoking is almost universal bear little responsibility for smoking, whereas smokers who belong to a social group in which smoking is rare bear virtually complete responsibility.

Most egalitarians who favor compensating people for being born blind oppose compensating them for being born without ambition or for being stuck with religious beliefs that induce them to starve themselves. As Scanlon (1986, pp. 116–17) and Dworkin both emphasize, "It would be incoherent for me to regard some ethical conviction I have – that the only important thing to do with my life is to create religious monuments, for example – as a limitation on the goodness of the life I can lead" (Dworkin 1990, p. 108). The point of egalitarianism is to show equal respect for individuals, which rules out regarding their convictions and characters as misfortunes:

> The distinction required by equality of resources is the distinction between those beliefs and attitudes that define what a successful life would be like, which the ideal assigns to the person, and those features of body or mind or personality that provide means or impediments to that success, which the ideal assigns to the person's circumstances. (Dworkin 1981b, p. 303)

Articulating and defending these distinctions in a coherent way is a large and unfinished project. Why draw a line between abilities and preferences? It won't do to say that individuals have no control over their abilities while their characters are up to them, because our characters are to a considerable extent beyond our control. Cohen argues that, rather than drawing a line between abilities and preferences, egalitarians should distinguish between those disadvantages whose acquisition and retention are voluntary and those that are not (1989, pp. 920–34). In Cohen's view, the purpose of egalitarianism "is to eliminate *involuntary disadvantage*" (1989, p. 916; see also Arneson 1989, p. 177).

However, Cohen's way of drawing the line has its own implausible implications. As already noted, it would be inappropriate to compensate someone for holding religious beliefs that demand suffering, regardless of whether those beliefs are voluntary. Cohen concedes the point and revises his account to rule out compensation not only for disadvantages resulting from choice but also for disadvantages so intrinsically connected to their bearers that their bearers would not choose to be without them (1989, p. 937). But if involuntary disadvantages are not always of egalitarian concern, then the purpose of egalitarianism cannot be to eliminate involuntary disadvantage.

Resource egalitarianism, particularly in Dworkin's early formulations, derives from the value of equal respect. In his more recent work, Dworkin (2000) emphasizes instead its connection with the specifically political obligation that governments have to show equal concern for all citizens. If one can draw the line between resources and what people should be held responsible for, then equality of resources respects individual differences and responsibility. Rawls also stresses that, in making distribution insensitive to individual preferences, this standard achieves fairness toward different notions of the good. But the connection between resource egalitarianism and equal respect or equal concern is tenuous. Why does equal respect or equal concern demand equal resources? The connection between equality of resources and the other values underlying egalitarianism is also far from clear. Which (if any) version of equality of resources is fair? Do unequal resources undermine the fairness of social institutions? Though presumably as a matter of fact there would be more fraternity and fewer stigmatizing differences in status if resources were equal, equality of resources does not imply any particular pattern of outcome. What, then, is the connection between fraternity or stigma and resources?

11.2.3 Equality of Opportunity for Welfare

Like Cohen, Richard Arneson relies on choice to draw the line between those traits of an individual that demand compensation and those that don't. For Arneson, someone with involuntary expensive tastes may be as deserving of compensation as is someone who is handicapped. "If we put aside practical difficulties about information-gathering and measurement of hypothetical rational preferences, what further good reasons could there be for treating involuntary expensive preferences due to handicaps differently than involuntary expensive preferences due to tastes?" (1990, p. 190). This thought seems fatal to resource egalitarianism: surely someone whose prospects are diminished simply because their tastes are expensive to satisfy does not thereby have fewer resources. If egalitarianism nevertheless implies that they deserve compensation, then egalitarians must aim at something other than equality of resources.

Arneson proposes that egalitarians should aim to equalize opportunities for welfare rather than to equalize resources. "For equal opportunity for welfare to obtain among a number of persons, each must face an array of options that is equivalent to every other person's in terms of the prospect for preference satisfaction it offers" (1989, p. 85). Involuntary expensive tastes, like handicaps, limit opportunities for welfare. Spelling this view out in detail leads to surprising complications, though there may not be much practical difference between equal opportunity for welfare and equalizing resources. Arneson supports his proposal by arguing that focusing on opportunities for welfare enables one to explain why resources matter and gives one a principled method for weighting their importance. Yet the reference to welfare has costs, too. Although Arneson specifies that welfare is the satisfaction of "ideally considered" self-interested preferences (rather than the satisfaction of actual preferences), he still faces many of the difficulties with preference satisfaction views of welfare discussed in Chapter 8.

If welfare depends on preferences, then seeking equality of opportunity for welfare shows respect for preferences and thus, perhaps, links up to the value of equal respect. But we must question whether it is fair that differences in people's preferences should lead to differences in the shares of society's resources that they receive.

11.2.4 Equality of Capabilities

Equalizing welfare seems to equalize too much and to pay too little attention to individual responsibility. Some welfare differences are attributable to the choices and efforts of individuals and should not be equalized. On the other hand, equalizing external resources – however radical this may appear in comparison with actual societies – seems in theory to equalize too little. There are internal differences between individuals, such as disabilities, that are not the responsibility of individuals and so justify compensation. Rather than counting handicaps and talents as resources (like those who emphasize equality of resources) or focusing on their consequences for welfare (like those who argue for equality of opportunity for welfare), one might instead attempt to equalize what Cohen calls "access to advantage." In his view, egalitarians ought to look at what is "in between" resources and welfare. Resources and internal features of the individual determine the range of outcomes from which the individual chooses. Cohen argues that egalitarians should try to equalize this range of outcomes or, in his words, "access to advantage." This is similar to Sen's proposal (cf. Section 8.5) that egalitarians seek equality of capabilities.

Though attractive, these proposals of Sen and Cohen face many of the problems encountered by Dworkin and Arneson. The line between those disadvantages that egalitarians should be concerned about and those that are up to the individual is still hard to draw. How should one distinguish between those features of an individual that affect capabilities and those that determine what use individuals make of their capabilities? Specifying when access to advantage or capabilities are "equal" is also at least as difficult as specifying when resources or opportunity for welfare are equal.

Given the measurement difficulties, it might seem that the only way to implement Sen's or Cohen's proposals would be to focus on resources. Yet Martha Nussbaum argues that, by specifying a list of central human capabilities (see Figure 8.5.1) and insisting that these be available to everyone, one can formulate a practical egalitarian program.

On the other hand, one might argue that it is a strength, not a weakness, that the "access to advantage" approach or Sen's version of the capabilities approach leaves undecided the question of how to weight the relative importance of different access or capabilities. For as Scanlon has argued (1975, 1986), the value of resources, capabilities, or primary goods should be determined by public moral deliberation concerning their importance in facilitating a range of good lives. Whatever one concludes on this matter, it is clear that the main appeal of Sen's, Nussbaum's, and Cohen's proposals is to fairness and equal respect. These proposals appeal especially to the argument from fairness for improving the life chances of those who suffer from handicaps.

11.3 Complex Equality and Equality of Moral Status

In our view, which echoes the forceful critique by Elizabeth Anderson (1999), many of the recent debates concerning what to equalize and whether to favor egalitarianism or prioritarianism reflect a relatively superficial treatment of how the claims and interests of individuals should be weighted. Though an exaggeration, one can say that instead of locating the importance of egalitarianism within ethical theory and showing which sorts of inequalities are wrong and why, the parties to these debates have traded intuitions and detailed criticisms concerning which inequalities are intrinsically wrong, which disadvantages justify claims to compensation, or how much compensation is required. The result is an unsatisfactory compromise that in many cases is directed neither toward practical policy proposals nor toward fundamental theory and that, as Anderson claims, "has lost sight of the distinctively political aims of egalitarianism. The proper negative aim of egalitarianism is not to eliminate the impact of brute luck from human affairs, but to end oppression" (1999, p. 288). She argues that "luck egalitarianism" – whether formulated as equality of resources, equality of opportunity for welfare, or equality of capabilities – "fails the most fundamental test any egalitarian theory must meet: that its principles express equal respect and concern for all citizens" (1999, p. 289). Luck egalitarianism fails this test because it ignores disadvantages for which individuals are responsible, because it rests the case for redistribution on the inadequacies or bad luck of the unfortunate, and because it rests on intrusive judgments concerning the sources of inequalities. In place of luck egalitarianism, Anderson advocates what she calls "democratic equality," which focuses on securing the social conditions of freedom and the bases for equal respect and hence on preventing oppression.

Although we are sympathetic to Anderson's critique and to her positive proposals, those proposals are sketchy and open to criticism (see Arneson 2000). We shall focus instead on a related way of thinking about equality that has been developed most clearly by Michael Walzer (1983; see also Miller and Walzer 1995). Walzer denies that there is a single right answer to the question of what should be equalized. Instead he points out that the principles governing disparate aspects of social life appear to differ considerably. Fathers and mothers are supposed to be trusting, caring, altruistic, and partial. Entrepreneurs are supposed to be self-interested, distrustful, competitive, and acquisitive. Professionals such as doctors and nurses are supposed to be motivated by a concern for the well-being of those for whom they care. Citizens are supposed to be equal, impartial, and motivated by their views of the collective good.

In contrast, many economists are inclined to regard individuals as basically acquisitive and self-interested in all domains of social interaction. But whether or not people's motives are always the same, the principles governing their interactions still vary. Resources within families are not distributed as wages, profits, or rents. Licenses to practice medicine are not sold to the highest bidder. Political life is governed by principles designed to protect equal rights. Political rights are free, and they are neither distributed as rewards nor (except in the case of convicted felons) revoked as penalties; they cannot be bought or sold (Okun 1975, pp. 6–10; Radin 1987, 1996). Inequalities from economic life do, of course, spill over into political life. But this is widely thought to be undesirable, and nations erect barriers aimed at limiting the extent to which wealth leads to political power.

Walzer maintains that different distributive standards should prevail in different "spheres" of social life. Even within areas of activity that are recognizably economic, he argues that one should resist the supposition that they should all be governed by a single underlying standard. Thus, for example, the social significance of employment in modern industrial societies counts strongly, in Walzer's view, for equal opportunity in employment. (Implementing this will be a complicated matter and would arguably demand much greater educational opportunity.) On the other hand, Walzer would deny that our society should underwrite a similar commitment to equality of opportunity in consumption.

Walzer's egalitarianism does not rule out economic inequalities, inequalities in authority within particular institutions, or even political inequalities that arise from differing abilities to persuade, organize, and lead. These inequalities are acceptable so long as there are equal political rights and liberties and so long as the inequalities in one domain do not spill over into others – in other words, so long as the inequalities are not pervasive. What's wrong with the economic inequalities in the United States is that they lead to inequalities in status, in employment opportunities, in political power, even in health. Although Walzer argues that there is no reason to eliminate economic inequalities completely, he nevertheless provides good reason to reduce them. Furthermore, Walzer argues that some ostensibly economic inequalities, such as those involved in the authoritarian structure of economic corporations, are actually political in nature and are as inconsistent with equality in rights and liberties as are towns that are owned and controlled by corporations (1983, ch. 12).

Walzer's complex equality seems to follow naturally from concerns with fraternity, self-respect, equal respect, and the avoidance of domination. Indeed, the quotations from Tawney that we used to illustrate the concerns for fraternity and equal respect are in much the same spirit as Walzer's account. Walzer's more eclectic approach is not incompatible with inquiries into conceptions of economic inequality, such as those reviewed in the preceding sections. Since Walzer defends limitations on the permissible degree of economic inequality, his work would be complemented by clarifying the moral significance of inequalities within the economic sphere. But Walzer's approach points in interestingly different directions than the work of those who have been quarreling over what to equalize or whether to be an egalitarian or a prioritarian. In stressing the relativity of conceptions of equality and justice to the practices of particular societies, Walzer's outlook is more historically situated than are the more abstract efforts to clarify notions of equality of resources, welfare, and the like. He emphasizes the significance of actual inequalities with respect to power, freedom, and what one might call the authenticity of relations. Walzer's approach focuses attention on the social meanings of goods and on the influence of economic arrangements on the broader political and social relationships in a society.

Just as one can ask, "Why be concerned with equality per se?" and "Why weight the interests of the worse-off more heavily?" so one can ask, "Why care about equality of moral status or 'complex' inequalities?" Answers to ethical questions give out when one reaches fundamental notions, and we think that at this point one is close to the foundations of morality. As Griffin puts it, "to see things morally is to grant everyone some sort of equal standing" (1986, p. 295). As we mentioned at the beginning of this chapter, virtually all moral theories find their origin in some notion of equal respect owed to moral agents, and concerns about equality of moral status derive from a concern not about equality itself but rather about a particular kind of status. Everyone should have an equal moral status because everyone should have the moral status of a person; in possessing that moral status, all are equal to one another. If everyone has achieved the full moral status of a human being, then no one can have a higher status. One might also speculate that true status as a person can be possessed by all only if possessed equally. In a society that does not accord the same baseline respect to all its members, that does not support the self-respect of all, or that explicitly (or implicitly) admits of different "grades" of human beings, a certain sort of moral identity may not be possible.

Looking to the moral justification for a concern with equality does not quickly resolve the question of what should be equal. All of the contenders except equality of welfare are to some extent plausible, and none is without its difficulties. We believe that egalitarians should support something like Anderson's democratic equality or Walzer's complex equality and that they should adopt something like Sen's, Nussbaum's, or Cohen's construal of the nature of inequalities within the economic domain. However, as noted previously, there may be little practical difference among policies intended to decrease inequalities in resources, equalize opportunity for welfare, or increase equality of capabilities.

11.4 The Measurement and Importance of Inequality

Given some sense of what one might want to equalize and why, one must then say something about how important equality is. Arthur Okun makes this question vivid by means of an analogy to a leaky bucket (1975, pp. 91ff). Suppose that one carries goods from the rich to the poor in a leaky bucket. How much leakage should one accept before deciding that efforts at redistribution are too costly? Redistribution has administrative costs, effects on incentives, and effects on attitudes toward work and care of oneself. When there are trade-offs between equality and efficiency (which Okun stresses) or between equality and freedom (which libertarians emphasize), how much weight should be placed on equality? These are difficult questions. Clearly the extent of trade-offs will be sensitive to the conception of equality that is under examination, and it will vary depending on the society's degree of inequality relative to its level of welfare or freedom.

In order to pose these questions meaningfully, one needs some way to measure inequality. Otherwise, there is no way to talk about the trade-off between a reduction in inequality and an increase in welfare. But comparing the inequality of different distributions turns out to be a surprisingly complicated matter, and many different measures are employed (Atkinson 1970; Temkin 1993).

Most egalitarians have contended that inequalities of resources, opportunities, and welfare in modern industrial societies could be greatly diminished without serious loss of efficiency or freedom and without violating anyone's rights. Indeed, it is plausible that equalizing educational opportunity in the United States would increase, not diminish, both efficiency and freedom (though not in the libertarian's sense). Economies such as Japan, Austria, and Sweden demonstrate that inequalities in wealth and income that are small relative to those in the United States are compatible with freedom and economic growth. Some observers might argue that the increase in inequality observed in many countries beginning in the 1980s is evidence that the trade-offs are worsening, but this is not obvious. Increases in inequality have often been accompanied by disappointing economic performance.

The extent of these trade-offs between equality and economic growth is ultimately an empirical question, but how much weight to place on equality is a matter of values. In the United States over the past generation, increases in the payroll tax, which falls heavily on those who are relatively poor, coupled with dramatic decreases in income, dividend, and estate taxes paid by the wealthy, have accentuated inequalities. Economists who hope to contribute effectively to social deliberation about the wisdom and justice of these policies – and about how to make trade-offs between equality and other values – need to grasp what inequalities egalitarians object to and what values ground their objections.

Suggestions for Further Reading

There are three useful anthologies of works on egalitarianism: Pojman and Westmoreland (1997); Mason (1998); and Clayton and Williams (2000).

Crucial texts in the contemporary egalitarian debate are Anderson (1999); Arneson (1989, 1990, 2000); Cohen (1989, 1993); Dworkin (1981a, 1988, 1990, 2000, 2002); Rawls (1971); Roemer (1985, 1987, 1988, 1996); Scanlon (1986); Sen (1992a); and Williams (1962). For an accessible discussion of Roemer's way of understanding equality of opportunity, see the April/May 1995 issue of the *Boston Review,* ⟨http://bostonreview.net/BR20.2/BR20.2.html⟩.

Philippe van Parijs (1989), developing suggestions of Bruce Ackerman (1981), has put forward a weaker egalitarian notion of "undominated diversity of resources" designed to allow for the fact that different people may rank alternative bundles of resources differently. Arguing from a commitment to undominated diversity and other moral principles, van Parijs has made a persistent case for an unconditional grant to be paid to every adult citizen (1989, 1990, 1991). Ackerman and Alstott (1999) have defended the somewhat different idea of a stakeholder society in which each individual, upon adulthood, would be granted a sizeable "stake" to be used for investment, education, or security. Something like their proposal has actually been implemented in the United Kingdom in the form of so-called Baby Bonds (see Paxton 2002).

See Fleurbaey (1995, 2002) for a challenge to the direction in which the discussion of egalitarianism has gone. For reviews of technical work on fair allocation, see Moulin and Thomson (1997) and Fleurbaey and Maniquet (2005). For an argument that equality and efficiency go hand in hand rather than competing, see Bowles et al. (1999).

An extensive on-line archive of papers on egalitarianism can be found at ⟨http://mora.rente.nhh.no/projects/EqualityExchange/⟩.

TWELVE

Justice and Contractualism

The questions of justice that are important to economists concern the distribution of benefits and burdens among members of a community. What claims can persons legitimately make upon one another or upon the state? What burdens can the state place on its citizens, or can individuals place on one another? Because justice is concerned with the proper distribution of scarce resources in society, its subject matter overlaps with economics.

Economists cannot decide what principles of justice to rely on merely by consulting public opinion, because people are committed to many principles of justice, which are at differing levels of generality and are often ambiguous and conflicting. Even when there is a social consensus on a principle of justice, it may not be well considered, and economists may still not find much guidance. Consider, for example, equality of opportunity. Everybody is for it, but that's partly because there's so little agreement about what it is. Does equality of opportunity require that more resources be devoted to the education of the relatively disadvantaged? Does equality of opportunity require steeper inheritance taxes? Does equality of opportunity rule out affirmative action or require it? Questions about the extent to which economic policies contribute to equality of opportunity can scarcely be broached until the concept itself is clarified. Economic evaluation presupposes well-defined principles of justice.

Although issues of justice are controversial and complicated, we doubt that economists can avoid them. Suppose, for example, an economist is called upon to describe the economic consequences of alternative policies affecting families. Although some have argued that the family is an institution "beyond justice" (Sandel 1982), marriages involve both conflicts of interests and benefits from cooperation (Sen 1990). For instance, changes in divorce law in the United States in the direction of "no fault" divorce have reduced the need for the partner who wants to get out of a marriage to bargain with their spouse. Partly as a result, divorce frequently leads to a dramatic drop in the living standards of women and children and to a dramatic increase in the living standards of men. (The standard of living of divorced women and their children in California between 1968 and 1977 dropped on average by 73% during the first year after divorce, while the divorced men's standard of living increased on average by 42%! – see Weitzman 1985.) These facts about divorce in turn affect the relative bargaining power of married men and women. Evaluating the economic consequences of family law requires principles of justice. It is unlikely that relatively unreflective everyday principles of justice will suffice, because they so often conflict and because their rationale is not clear. In addition, as illustrated by the case of equal opportunity, accepted principles may pose such large problems of interpretation that they provide little guidance. So it seems that economists must look to moral philosophy.

Libertarian and utilitarian moral theories, the subjects of Chapters 7 and 10, imply theories of justice. For libertarians, justice is simply respecting the rights that libertarians recognize. For utilitarians, principles of justice are utility-maximizing general rules that facilitate cooperation and regulate conflicts of interest. Yet it is also natural to use the language of agreement, promise, negotiation, and particularly contract in order to talk about issues of cooperation and conflict, which are central to justice. This chapter describes the resurgent interest among economists and philosophers in theories of justice that build upon the idea of a social contract.

12.1 The Social Contract Idea

Social interactions involve conflicts of interest, but on the whole they are mutually beneficial. People would be in desperate straits without the companionship and assistance of others. Circumstances like these – falling between complete harmony, which would make justice superfluous, and the war of all against all, which would render it hopeless – are called by Hume and by Rawls "the circumstances of justice." It is tempting to think of the normative principles governing individual interactions as human contrivances to adjudicate conflicts of interest and to secure the benefits of cooperation. Norms might thus be regarded as arising from some sort of social contract. The idea is already set forth in Plato's *Republic* (though Plato denies that the principles to which individuals would agree in order to facilitate living together are truly principles of justice). In the seventeenth and eighteenth centuries, following the collapse of a medieval and Renaissance vision of society as naturally hierarchical, political philosophers revived the notion of a social contract as the source of political obligation and social norms, including principles of justice.

The basic notion of social contract theories is that one judges principles and institutions by asking whether people would agree to them in a "state of nature" in which individuals are not already bound by social obligations but are governed only by enlightened rational self-interest (which seventeenth- and eighteenth-century philosophers called the "law of nature"). Not all philosophers have found contractualist reasoning valuable. David Hume, for example, famously complained that no contract can explain or justify the duty to keep contracts. However, advocates of contractualism find the contract notion attractive because it links justice to consent and to rationality. Showing that a certain principle would be agreed to by people acting reasonably in the right circumstances increases its claim to guide people's actions, and showing that people in a state of nature would find it rational to agree to abide by certain principles supports the claim that the principles themselves are rationally defensible. All contract theories test moral principles by asking whether rational agents would agree to them under properly defined circumstances, but they differ in the interpretation of the hypothetical circumstances of choice, in their characterization of the agents' rationality and motivation, and in their view of *what* is to be chosen. These differences lead to large variations among contract theories.

Brian Barry (1989) has helpfully distinguished two broad categories of contract theory, especially as applied to theories of justice. The first, exemplified by Hobbes and Hume, links rationality to self-interest and agreement to the outcome of bargaining, leading to a view of justice as *mutual advantage*. The social contract is a compromise that enables individuals to pursue their separate aims more harmoniously and successfully. The second category, exemplified by Kant and Rousseau, links rationality to the autonomous pursuit of ends (which need not be self-interested) and agreement to consensus, leading to a view of justice as *impartiality*. The social contract determines what principles governing interactions would be endorsed by impartial agents. John Rawls, whose theory we will examine in more detail below, regards his own view as representing a third standpoint, one of *reciprocity*. Impartiality, argues Rawls, suggests an unconditional commitment to promoting the common good, whereas reciprocity implies a willingness to do one's part in a cooperative scheme provided that others also do theirs. Like impartiality and unlike mutual advantage, reciprocity may sometimes entail acting against one's own interest, but only in circumstances where others (if necessary) are disposed to do the same. However, in this chapter we will not pursue the contrasts between impartiality and reciprocity interpretations of contractualist theories.

Unlike utilitarians, who characterize equal respect as counting everyone's interests equally, or libertarians, who characterize equal respect in terms of equal rights, contractarians interpret equal respect as the equal moral status of the parties to the social contract. Thomas Scanlon maintains that a contractarianism based on impartiality requires one to "justify one's actions to others on grounds that they could not reasonably reject – reasonably, that is, given the desire to find principles which others similarly motivated could not reasonably reject" (1982, p. 116; 1998, ch. 5). An impartiality or reciprocity contractualist insists that moral principles be rationally acceptable to every agent who is willing to take up a perspective that recognizes the common need for agreement. Scanlon's formulation highlights a feature common to most writing in a contractualist vein: the interest not only in identifying defensible moral principles but also in giving an account of why people should feel bound by those principles. Since theorists of mutual advantage adhere closely to a view of rationality as self-interest, a major challenge for their views is to explain why rational persons would agree to restrain their pursuit of their interests and especially why they would adhere to agreements once made. Those who see contractualism as determining what is impartially best instead assert that people may be moved by a desire to regulate their conduct by principles that are morally acceptable from a suitably detached perspective. This assumption makes it easier to explain how moral considerations succeed in moving people, but it is more speculative than the self-interest assumption favored by theorists of mutual advantage.

12.2 Justice as Reciprocity: Rawls's Theory of Justice

The most influential work in modern moral and political philosophy is John Rawls's *A Theory of Justice* (1971), an epic exposition of a contractual moral theory adapted to contemporary circumstances. It is worth explaining Rawls's views at some length because of their pivotal role in contemporary philosophical discourse and also because Rawls draws on economics and offers a great deal to economists. The derivation of his principles of justice relies heavily on an analysis of rational self-interested choice in a special hypothetical situation that Rawls calls the "Original Position," and Rawls applies his principles of justice to basic political and economic institutions. Rawls sees the practical implications of his theory as interesting in their own right and also as an important way of testing his theory. In his view, the theory is acceptable only if it has implications that his readers can, on reflection, endorse.

12.2.1 Contractualism and the Original Position

The contractualist element in Rawls's theory derives from the claim that just principles are those that would be chosen by rational actors in the Original Position. We should suppose that we are choosing principles that would shape the basic institutions of our society while behind a "veil of ignorance" that deprives us of knowledge of what role we would occupy in that society. This veil of ignorance – which extends to one's social position, race, gender,† occupation, and even preferences – forces the choosers to be impartial. We cannot craft principles that will be to our own particular advantage, because we have no idea of what that particular advantage is. The veil of ignorance makes vivid the notion that arguments for principles of justice must not rely on such arbitrary factors. Because Rawls is in search of principles acceptable to rational individuals who are concerned to advance their interests, he assumes that (behind the veil of ignorance) agents are neither altruistic nor envious. They are also ignorant of the range of possible social positions and of the relative size of different groups. Thus the decision situation is one of radical uncertainty rather than calculable risk. This thought experiment is a method of determining what principles of justice should govern the basic institutions of society, which define people's rights and expectations.

Though Rawls defends the veil of ignorance as an appropriate method for addressing some ethical questions, there are others to which it should not be applied. For example, if one is deciding whether *P* ought to pay *Q* some money, then one needs to know what the rights and entitlements are; one cannot suppose that they are still to be determined. Nor does Rawls think that the veil of ignorance mechanism can be directly applied to individuals on a global scale. For that purpose, Rawls (2001b) proposes that the Original Position be recast as applying to the various "peoples" of the world, with each people deliberating over ground rules for global society from behind a veil of ignorance concerning its relative standing on the world stage.

† In *A Theory of Justice*, Rawls takes individuals behind the veil of ignorance to be heads of households. Okin (1989) argues that, like most theorists of justice, Rawls ignores the problems of justice within the family and the problems of justice posed by children and other dependents.

Rawls's framework borrows heavily from one developed independently by John Harsanyi and by William Vickrey to defend utilitarianism. These authors argued that, when faced with uncertainty about one's social role, the rational choice would be to maximize the expected value of one's utility, averaging over all the possibilities. The proper social principle is thus to maximize the average utility of everyone in society. Rawls objects that the principle of average utility is too risky to choose in this special situation. Maximizing utility for the whole society could require sacrificing the interests of some particular group, and it seems unjust to sacrifice the entire life prospects of some people to benefit others. Furthermore, the situation of uncertainty, as Rawls defines it, makes the calculation of expected utility irresponsible if not virtually meaningless: with no knowledge of the range of possible social groups or of the relative size of groups, one is pushed toward approaches to choice under uncertainty that don't require probability calculations. Finally, one cannot maximize utility (the satisfaction of preferences) without knowing what one's preferences are.

Rawls argues that, in the special circumstances of choosing principles of justice behind the veil of ignorance, it is rational to minimize the costs of winding up at the bottom of the heap. More specifically, Rawls argues that – for societies that are at least moderately well-off – rational choosers would seek first to safeguard their basic political and personal liberties and would then judge basic social and economic institutions according to their tendency to promote the primary goods of those in the least well-off social group. No one would find their basic liberties threatened, and social and economic institutions would be designed to advance maximally the interests of the worst-off group. It is important to note that these interests are to be judged not by the subjective standard of preference satisfaction but by the more objective notion of amounts of primary social goods, a notion that even individuals deprived of knowledge of their own view of the good would agree on as a standard for assessing basic institutions (see Section 8.5).

12.2.2 Rawls's Principles of Justice

In their full elaboration, Rawls's principles of justice are as follows.

> First, each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others. (1971, p. 60)
>
> [Second:] Social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged and (b) attached to offices and positions open to all under conditions of fair equality of opportunity. (1971, p. 83)

In addition, Rawls specifies that the first principle is, for any society with enough resources to provide the basic liberties for all, lexically prior to the second. That is to say, basic political and civil liberties are never to be traded off for economic advantage. Moreover, part (b) of the second principle is lexically prior to part (a) – the interests of the least well-off are not to be advanced at the sacrifice of the principle of equal opportunity. Rawls calls part (a) of the second principle "the difference principle." (In *Justice as Fairness: A Restatement*, Rawls presents modest revisions of the two principles that are intended to clarify misunderstandings and respond to critics – Rawls 2001a, pp. 42–3.)

The "priority of liberty," as the first principle is sometimes labeled, is not easy to capture in economic models, and the lexicographic ranking of this principle, which rejects the notion of trading off liberty for other valued ends, is not the sort of thing economists find easy to swallow (but see Cooter 1989). Rawls maintains that, after one attains a basic minimum of resources, further means to pursue one's goals are valueless unless one preserves the basic liberties that permit these goals to be effectively pursued. Not to secure the priority of liberty and thus to gamble liberties against resources betrays a failure to understand what it is to hold a conception of a good life.

Economists have also had doubts about the difference principle – that society should maximize the primary goods of the least advantaged – but they have been able to seize upon its formal interpretation as a "maximin" standard. If one simplifies and takes the difference principle as applying to welfare rather than primary goods, it implies that (subject to the first principle and part (b) of the second principle) society should maximize the well-being of the least well-off. This maximin rule is as simple as the familiar utilitarian rule to maximize average well-being, and economists have investigated implications of these two ideals (see e.g. Phelps 1973; Alexander 1974; Musgrave 1974).

There are serious disadvantages to reinterpreting the difference principle in terms of utility. Doing so makes Rawls's references to representative individuals or groups puzzling, and it leads to a view of justice that (like utilitarianism) is strictly patterned. On such a view, if anybody at the bottom could be made better-off then the social institutions are unjust. On Rawls's view, in contrast, a theory of justice that attempts to determine what justice demands as an outcome for each individual is not only unworkable, it fails to recognize that outcomes depend also on individual choices. The application of his two principles determines a framework of rights and institutions that will ensure maximal equal liberty and, insofar as possible, fair equality of opportunity. If social and economic inequalities benefit the least well-off group – as they may if they encourage effort and risk taking – then such inequalities will be permitted, but they will be arranged so that the life prospects of those in the worst-off group include the highest level of primary goods possible. A good deal thus turns on how one defines the worst-off group, and Rawls does not defend any specific definition. If the rights and institutions satisfy Rawls's principles, then the results are just, regardless of their details. The assessment of individual outcomes involves "pure procedural justice," just like the assessment of the results of a lottery. Provided that the rules were followed, the results are automatically just because there is no other criterion by which to judge those outcomes.

Though Rawls does not defend any specific characterization of the least well-off group, he does stipulate that its members are not ill or handicapped; he does not intend his theory of justice to address the moral problems raised by disabilities and health disparities. He is well aware of the thorny questions concerning the just distribution of medical care, which have become increasingly urgent as medical care has become increasingly efficacious. Rawls (to some extent following Daniels 1985) sees the distribution of medical care as governed as much by equal opportunity as by the difference principle (Rawls 2001a, pp. 173–4). Recent work has shown that there are pervasive and significant health differences between those of higher and lower social status. Not only do the poor have much worse health and much lower life expectancies than those who are not poor, but those who are relatively affluent and enjoy high status have worse health than those who are at a still higher status. Not only does life expectancy drop 20 years over the few miles that separate downtown Washington D.C. from an impoverished suburb, but in England engineers and doctors live on average three years longer than teachers and journalists. This effect was first established in a large study of British civil servants, all of whom were engaged in office work and had access to the same health care. The effect persists even after one controls for weight and smoking. One can see the same effect in the surprising fact that the life span of winners of Academy Awards is on average four years longer than nominees who did not win, or in the lower mortality of Swedes with Ph.D.s than of Swedes with B.A.s (Marmot 2004). Given how important health is to living well and given the magnitude of these effects, it is questionable whether issues concerning health and justice can, as Rawls hoped, be deferred until after the central principles of justice have been ascertained.

In translating Rawls's difference principle into the view that justice involves maximizing the minimum utility level of individuals, economists may thus have missed what is of the greatest potential value in Rawls's theory of justice. Although redistributions may be justified by particular injustices, concerns about justice should not be conceived as fundamentally redistributive. Reading Rawls should lead economists instead to think about the design of institutions that will minimize the need for redistribution, and his views should lead economists to focus on the means that individuals employ to construct their own goods instead of focusing directly on the satisfaction of preferences.

12.2.3 Implications of Rawls's Principles

Rawls's contribution holds special interest for economists. As we have already indicated, the argument from behind the veil of ignorance makes effective use of notions of rational choice, decision making under uncertainty, and maximization subject to constraints – all theoretical ideas with which economists are comfortable. His conception of individual distribution as a matter of pure procedural justice and his emphasis on primary goods suggest a new way to address problems of justice and a different set of units in which to talk about how groups are faring.

In addition, Rawls works out implications of his views both to address practical questions and to determine whether his principles yield results that, on reflection, informed observers find reasonable. Rawls investigates implications of his theory for the provision of public goods, the role of redistributive taxation, justice in saving and capital accumulation, the role of competition, and macroeconomic policy toward unemployment. In particular Rawls also considers the issue of whether capitalism or socialism is more compatible with justice, concluding that both systems can be just under the right circumstances.

As one might expect from a theory that focuses on the least well-off group, the institutional implications of Rawls's theory are egalitarian. He favors equalizing educational opportunities for persons of similar capacities, and he supports policies that discourage the accumulation of hereditary fortunes and block the influence of wealth on political decisions. Rawls's ideal society would, however, rely only minimally on redistributive taxes and transfers as an equalizing measure. On grounds of efficiency, liberty, and opportunity, Rawls emphasizes efforts to equalize the starting points of members of successive generations in preference to redistributing unequal outcomes.

The final feature of Rawls's work that should be of interest to economists derives from his concern to investigate what he calls the "stability" of principles of justice. Would a society that conformed to the principles of justice as fairness cause people to behave in such a way that those principles would continue to be upheld over time? The answer will depend on how psychological, sociological, and economic forces work themselves out over time. Will children raised under just institutions tend to embrace the principles of justice and regulate their lives properly by them when they become adults? Can institutions be created that are compatible with the basic liberties and that will ensure that inequalities do not accumulate over generations and so undermine the principles of justice over time? These are deep questions about political sociology and moral psychology, made all the more interesting in Rawls's treatment because he frames them explicitly as questions about justice and highlights their importance for assessing theories of justice.

Some major features of Rawls's work are more closely tied to his contractualist commitments than are others. In particular, one might find the Original Position framework helpful without being persuaded by the conclusion that Rawls's specific principles of justice would be chosen; conversely, the substantive principles may be appealing to those who are not persuaded by the Original Position argument.

12.2.4 Justice and Pluralism

A Theory of Justice provoked basic questions about the proper scope of the theory of justice. Over what range of time and space is the theory intended to apply? Does it make sense to appraise a medieval society or a nomadic tribe by the standards of Rawls's principles of justice? Is it plausible that the same principles of justice apply to representative democracies, feudal manors, and nomadic herders? How would a Rawlsian persuade an ancient Egyptian to enter the perspective of the Original Position?

An analogous problem of scope arises regarding the question of how comprehensive Rawls intended his moral view to be. Although one can read *A Theory of Justice* as a timeless and explicitly liberal moral theory, Rawls makes clear in his later writings that he prefers to see his work as a historically situated theory of justice for a pluralist society. A major goal of *Political Liberalism* (1993) and *Justice as Fairness* (2001a) is to clarify that his theory can be understood as a workable account of justice for a modern, pluralist democracy. Rawls depicts his theory of justice as a response to enduring questions concerning liberty and social and economic equality in diverse and developing commercial societies. He intends his principles of justice to provide a philosophically acceptable basis for social cooperation and also to provide grounds for what he calls "realistically utopian criticism" – "probing the limits of practicable political possibility" (2001a, p. 4). While not narrowly bound to contemporary controversies and immediate political opportunities, Rawls conceives of his theory as tailored to the issues of modern pluralist representative democracies.

For such a theory to achieve its purpose, it must not be anchored in a distinctively liberal view of the person and of morality, like those of Kant or of J. S. Mill. If it were, then the theory could command the adherence only of those who share a liberal vision of the person and of the good life, and it could not bring about the "overlapping consensus" required of a political philosophy that can hope to resolve enduring political disputes. Moreover, Rawls suggests, a theory that presupposes that an ideal society will be composed of people who share a common philosophical liberalism will be self-refuting. For, Rawls argues, it is a fact of modern life, which he calls the "fact of reasonable pluralism," that any society that secures the basic liberties will wind up producing people who hold a range of conflicting moral convictions and religious beliefs. Reason alone is inadequate to produce agreement; only coercion could maintain a society in which everyone's underlying beliefs were, say, those of a Kantian liberal or of a Roman Catholic. Given the impossibility of uncoerced agreement about moral and metaphysical foundations among free citizens, the problem for political philosophy in these circumstances is to seek out agreement on terms of fair cooperation among people who hold differing comprehensive moral, metaphysical, and religious views.

What is needed is not universal agreement but what Rawls calls "overlapping consensus." That is, people holding any particular comprehensive view, provided it is "reasonable," will come to see that they cannot expect all others to share their comprehensive view and must instead search for grounds of cooperation that are acceptable to everyone. Holders of any reasonable comprehensive view (as well as those who hold no comprehensive view but are reasonable in their outlook) will see that other views are also reasonable and will seek out fair terms of cooperation with those who hold them. The argument for justice as fairness, then, is that this is a set of principles on which all who seek fair agreement can agree.

For example, a conscientious Protestant can endorse a political conception of justice from which follows a commitment to liberty of conscience because coerced declarations of faith are of no value within the framework of her religion, while a skeptic can endorse the same political conception – including its implication that liberty of conscience should be upheld – out of a judgment that no right answers are to be had in this realm. They agree on general principles of justice to govern their common lives and accept the more particular principles they entail, but for different reasons. For political purposes they can agree, but they do not agree "all the way down." Rawls suggests more broadly that what may begin as simply a *modus vivendi* – a tactical agreement that toleration is preferable to continued warfare (as in the wake of the Reformation) – may grow into a more genuine consensus as the value of a free and fair political culture becomes apparent.

This casts the argument for the two principles of justice in a somewhat different light. Although one might regard the Original Position in *A Theory of Justice* as a device to help people to reflect on their most basic ethical and philosophical convictions, in Rawls's later works it serves as a device by which people who disagree (and who expect to continue disagreeing) about what constitutes a good life can arrive at a principled, just, and mutually agreed upon way to live together. The goal is not mere compromise but genuine justice – terms of cooperation that are justifiable to individuals who have their own views of the good but also a capacity for justice.

In articulating explicitly the historical context for the theory of justice and in distinguishing arguments about the political realm from arguments about comprehensive views of metaphysics and morality, Rawls raises questions about other apparently universalist theories, including those that appeal to economists. For example, is utilitarianism best understood as an attempt to resolve issues of policy and resource allocation as they arise in modern societies with representative democracies, or rather as a timeless doctrine applicable to both individual action and social policy and one that would be equally justified as the basis for ethical decisions in an agrarian society or a nomadic tribe? Or, again, must a defense of libertarianism invoke controversial views about the subjectivity of value and the prerequisites of autonomy, or could libertarianism instead be defended as a suitable basis for cooperation in a society with multiple conflicting metaphysical and moral views?

12.3 Justice as Mutual Advantage: David Gauthier

Theorists of justice as mutual advantage reject the features of the Rawlsian contractualist framework that limit the role of self-interest in contractarian reasoning. Thus David Gauthier, perhaps the most prominent of contemporary philosophers in this tradition, denies that agreements reached behind a veil of ignorance are rationally or morally binding. The challenge for a contractualist, in Gauthier's view, is to solve the problem Hobbes formulated: How can it be rational for a self-interested actor to conform his behavior to moral rules? According to Gauthier, one must consider what would be the result of bargaining among expected utility maximizers with full common knowledge of everyone's capabilities, endowments, and preferences. Unlike Rawls's Original Position, where mutual ignorance of differences eliminates the possibility of conflicting interests, Gauthier's framework leads us to conceive of the contractual interaction as *bargaining*.

No actor in Gauthier's setting will agree to accept less than she could obtain in the absence of agreement. Those who are advantaged initially will emerge with advantages from the bargain. Gauthier proposes that, in the absence of means to make interpersonal utility comparisons, agents in this situation would agree to distribute the gains from cooperation in accordance with a principle of "minimax relative concession": the largest relative concession anyone has to make should be as small as possible. Let $u_{\max}$ be the maximum utility the agent can get. Let u be the utility of the bargain for the agent, and let u_{na} be the agent's utility at the nonagreement point. An agent's relative concession is then $u_{\max} - u$ divided by $u_{\max} - u_{\text{na}}$. To define such ratios, the preferences of individuals must be representable by cardinal utility functions; however, since a comparison of relative concessions is a comparison of ratios, not utilities, there is no need to make interpersonal utility comparisons. In the technical literature on bargaining theory, which we discuss in Chapter 14, this is similar to the Kalai and Smorodinsky bargaining solution (see Section 14.5). In most circumstances, minimax relative concession coincides with equal relative concession. Minimax relative concession implies that those who are worse-off without agreement will have to give up more to gain agreement. Gauthier argues that – in comparison to the situation prevailing in the absence of agreement – this principle would distribute the gains from social interaction fairly. In Gauthier's view, the nonagreement point would reflect substantial inequalities, which are largely preserved in the just society (or else it would not be in everyone's interest). Consequently, the scope for defending equality-promoting institutions and policies is substantially less here than on Rawlsian or utilitarian views.

An important goal in developing theories of justice as mutual advantage is to show that a commitment to obey principles of justice is rational, even though rational actors are assumed to lack any independent desire to be just. Actors who are rational in this sense would not, for example, be moved by Scanlon's principle of grounding one's behavior on rules that could not be reasonably rejected by others. The reconciliation between morality and self-interest that mutual advantage theorists seek is hard to achieve, since the presence of rules of justice creates a classic prisoners' dilemma: even if you and I jointly benefit from the rules, I will do still better if you obey them and I allow myself to violate them when it suits me. Gauthier argues that it may be in my rational self-interest to cultivate the kind of character that will lead me to obey rules of justice, even if having that kind of character sometimes robs me of opportunities to free ride. It is crucial to this solution that other members of society are reasonably good at determining what kind of character I have, and that (depending on what they discern) they are able to reward or punish me suitably by their willingness to interact with me. The issue here – of when it may be in one's interest to cultivate dispositions that inhibit acting with explicitly self-interested motives – has received valuable treatment from economists (Akerlof 1983; Frank 1988).

12.4 Other Contractualist Views

James Buchanan and his colleagues argue that self-interested agents who are concerned about "constitutional" choices – choices involved in setting the general rules of the game – face so many uncertainties that they might as well be behind a veil of ignorance. Since the gains from constitutional design may be very large, self-interested agents should be concerned about the rules of the game and so have good reason to step behind "the veil of uncertainty" (as Brennan and Buchanan 1985 style it). Even though the motivational assumptions in this constitutionalist view are more self-interested than in Rawls's, the element of impartiality introduced by not knowing how alternatives will affect one's own situation provides a strong common element between the views.

In Buchanan's view, rational agents behind the veil of uncertainty will be more cautious about promoting equality through government action than Rawls's reasoning implies, because they will model both economic and political behavior as self-interested. Rawls is explicit in departing from this assumption: "ideal legislators do not vote their interests" (1971, p. 284). Brennan and Buchanan argue that agents behind the veil of uncertainty will endorse less egalitarian arrangements than Rawls does, partly on grounds of realism and partly on the Humean grounds that, in designing constitutions, "every man ought to be supposed a knave" (quoted in Brennan and Buchanan 1985, p. 59). The different assumptions that Rawls and Buchanan make about people's psychology lead them to different conclusions about the relative stability of societies organized according to various principles of justice. Even supposing that generous provision for the poor is desirable, political institutions such as majoritarian democracy may fail to provide it or may do so in a corrupt and inefficient way (Brennan and Buchanan 1985). On grounds such as these, Buchanan argues that individuals adopting a constitutional perspective would not want to leave redistribution to the discretion of self-interested politicians. If they favor measures to equalize resources, then they should prefer to embed particular distributive rules directly into the constitution – which might, for example, specify a flat tax coupled with a negative income tax (payments to low-income persons or families that decline with rising income). Such constitutional distributive rules (like the institutions mandated by Rawls's principles of justice) should not be considered *re*distributive, since they would figure into the underlying definition of property rights (see also Buchanan 1975).

The general idea that moral principles or social norms can be usefully interpreted as the outcome of self-interested bargains – that is, the research program suggested by the notion of justice as mutual advantage – is receiving wide attention. This sort of contractualism makes heavy use of the tools of game theory (see Chapter 14), and its findings may have both explanatory and normative force. In exploring mutual advantage models and in comparing their findings to contractualist work that emphasizes reciprocity or impartiality, it is important to remember their different starting points. What people would agree to if they were motivated as Scanlon assumes is a different matter from what self-interested people would agree to. Which approach is more fruitful depends on philosophical theses about the nature and point of reasoning about justice, which are not themselves answered from within either of these standpoints.

12.5 Conclusion: Social Contract Reasoning and Economics

Problems of justice can be addressed from many different perspectives, and the general moral theories we have already discussed – libertarianism and utilitarianism – imply theories of justice. Libertarianism appeals to many economists who share its devotion to the value of freedom and its hostility to government interference in economic matters. Utilitarianism appeals to many economists who share its focus on individual well-being. Contractualist approaches may have a harder time getting a hearing. But there is, we contend, good reason why students of economics should take them seriously.

All forms of contractualist reasoning are concerned with rational agreement. In an earlier formulation, Rawls went so far as to say that his approach made the theory of justice a branch of the theory of rational choice. Most contract theorists are also concerned to draw conclusions regarding matters of great interest to normative economics: the justification of property rights, the legitimacy of redistribution, and so on. Finally, forms of contract theory reflect well worked-out alternatives to "welfarist" theories of normative economics. At the same time, it is important for economists who rely on or contribute to work in contract theory to remember that the interpretation of particular bits of contractualist reasoning depends on the philosophical grounds for adopting a contractualist point of view. The reawakening of interest in theories of the social contract has been an important stimulus to economists' interest in normative theory, and we hope that its influence will grow.

Suggestions for Further Reading

Historical texts on the social contract are especially important for appreciating its purposes and significance. Landmark texts include Hobbes (1651), Locke (1690), Rousseau (1762 [1968]), and Kant (1785).

The literature on Rawls is exceptionally large and rich. The anthology edited by Daniels (1976) and the books by Barry (1973), Wolff (1977), Pogge (1989), and Kukathas and Pettit (1990) are good starting points. Rawls's economic framework was heavily influenced by James Meade's classic (1964). Scanlon (1998) develops a "reciprocity" view, including an intriguing account of moral reasons.

Along with Gauthier and Buchanan, Hampton (1986) and Kavka (1986) have outlined general theories of justice that take justice to be mutual benefit. Schotter (1981), Sugden (1986), and Taylor (1987) consider the recurrent problems of coordination and conflict that individuals face and the sorts of institutions and conventions they will come to accept to resolve them. Binmore (1990) argues that imperfectly rational agents will be led to a utilitarian solution to recurrent bargaining problems, and in his two-volume work (1994 and 1998) he treats the relationship between social contract theory and game theory in depth and offers an independent rationale for something closer to Rawls's views.

PART FOUR

MORAL MATHEMATICS

At first glance, morality and mathematics seem as unrelated as martyrdom and MTV. But it is possible to characterize moral notions formally and to prove theorems. Doing so does not banish controversy and cannot replace verbal argument concerning moral matters, because the abstract mathematical characterizations of moral notions require interpretation and defense. Just as there are disagreements concerning the formal definitions of rationality, so are there controversies about formal definitions of moral notions.

Over the last 75 years economists, decision theorists, and game theorists have made exciting progress not only in representing individual rationality but also in characterizing features of human interactions. This work is linked to moral philosophy. Formal models of rationality and game-theoretic studies of incentives hold out the hope of transcending some of the ancient conundrums concerning the relations between morality, rationality and self-interest discussed in Chapter 6. Concepts of "fair" or envy-free allocations, discussed in Chapter 11, facilitate the articulation of egalitarianism. Solution concepts in game theory may enrich the contractarian perspective (discussed in Chapter 12) that morality can be justified in terms of agreement. Theorems in social choice theory explore the implications of plausible axioms and test the consistency of traditional principles concerning how social policies should respond to individual interests. We face an embarrassment of riches, and we can only comment on a few of these developments.

THIRTEEN

Social Choice Theory

As we saw in Chapter 9, contemporary "Paretian" welfare economics is of relatively little help in evaluating economic institutions and outcomes. Welfare economists often espouse the Pareto principle (that A is better than B if somebody prefers A to B and nobody prefers B to A), but it's a rare day when there is this sort of unanimity. Welfare economists are sometimes tempted to say that A is better than B when A is only a *potential* Pareto improvement, but as argued in Chapter 9, this view is dubious. Economists need better bases for evaluating economic outcomes and institutions.

13.1 The Social Welfare Function and Arrow's Theorem

Social choice theory evolved out of an effort to construct better tools for evaluation. Following Bergson (1938) and Samuelson (1947), let us call any ranking of social states a "social welfare function." Normative principles can be regarded as constraints on social welfare functions. For example, resource egalitarianism demands that social welfare functions favor egalitarian resource distributions. Nazi principles of "racial purity" narrow the set of acceptable social welfare functions to those that give a low ranking to states of affairs in which Jews prosper. Hedonistic utilitarianism picks out social welfare functions that rank states of affairs in terms of total happiness.

The point of discussing social welfare functions is to provide a framework for exploring normative principles. Although the framework could in principle be applied to a broad range of evaluative criteria – including highly nonindividualistic ones – social choice theorists have in practice focused on social welfare functions whose arguments are individual preferences. The Pareto principle, for example, can be characterized as requiring that social welfare functions rank R over S if somebody prefers R to S and nobody prefers S to R. Such social welfare functions are "welfarist." Although there are several notions of "welfarism" in the literature, we take welfarism as restricting the inputs to social welfare functions to information about individual preferences. The formal characterization of Paretian social welfare functions has helped to establish the basic theorems of welfare economics and to clarify the normative claims underpinning cost–benefit analysis (see Chapters 5 and 9).

In the early days of investigating social welfare functions, it was hoped that thinking in terms of such functions could assist economists in identifying additional plausible normative principles that, like the Pareto principle, relate individual preferences and social welfare. If these principles had precise mathematical formulations, then economists would be able to deduce their implications. In this way, economists could carry out a sort of moral mathematics. However, much of the work on social welfare has not taken this form, owing to (a) the limitations of the basic Paretian perspective and (b) results of early formal inquiries into social choice and evaluation. A crucial problem with the welfarist framework is that it is relentlessly consequentialist: all that matters are outcomes. Procedural matters such as fairness or due process count only insofar as they affect outcomes.

A more specific obstacle to development of the social welfare function framework derives from a remarkable theorem proved by Kenneth Arrow in 1951. Rather than arguing for strong substantive ethical principles, Arrow put forward a set of apparently weak "procedural" principles that constrain acceptable relationships between individual values and social values. Arrow's aim was to identify a set of principles that would be consistent with the usual standpoint of normative economists and that would have widespread ethical appeal in their own right. Yet what Arrow discovered was that, in identifying a relatively short list of apparently innocuous constraints to whittle down the set of acceptable social welfare functions, he whittled it down to nothing! Arrow proved that no social welfare function could meet the conditions for acceptability that he had laid down.

What are the conditions Arrow imposed? First (see Arrow 1967), there is a version of the Pareto principle:

P If everybody prefers *A* to *B*, then *A* is better than *B* (weak Pareto).

By "better than" we mean that the social welfare function ranks *A* above *B*. Observe that **P** is implied by the view that Pareto improvements are moral improvements, which we called "the Pareto principle" in Chapter 9, but is weaker. Though more controversial than minimal benevolence (because it lacks an "other things being equal" qualification), **P** has seemed plausible, and social choice theorists often endorse it.

A second principle, which seems to us genuinely uncontroversial, is the following:

D Whether *A* is better than *B* should not depend on the preferences of a single individual regardless of what everybody else prefers (nondictatorship).

Speaking of dictatorship suggests issues about how to make social choices rather than issues about how to evaluate alternatives, and as we shall see, all of these conditions can be interpreted as constraints on decision-making methods rather than as determining the value of alternatives. But on either interpretation, it seems unobjectionable to stipulate that evaluations of all alternatives should not depend on a single person's preferences only.

The third condition is fundamental:

I Whether *A* is better than *B* should depend on how individuals rank *A* and *B* and on nothing else (independence of irrelevant alternatives).

This principle gets its name because it says that the comparison of two alternatives *A* and *B* should not depend on the presence or absence of other "irrelevant" alternatives. But the name is misleading. This version of the principle says that the goodness of the pair of alternatives being compared – the "relevant" alternatives – depends on how these alternatives stand in individual preference rankings, and that the goodness of these alternatives depends on nothing else. Condition **I** may be attractive to welfare economists because they so often take welfare to be the satisfaction of ordinal preferences, but it is hard to believe that nothing should be relevant to the social evaluation of *A* versus *B* except individual rankings of *A* and *B*. In actual political systems the choice between *A* and *B* typically depends on many other factors, such as precedent, constitutional provision, the extent to which *A* and *B* respect rights, and the intensity of preferences. All of these factors are ruled out by **I**. Indeed, **I** implies that the information that *A* is Pareto efficient and *B* is inefficient is irrelevant (unless *A* is Pareto superior to *B*) because the efficiency of *A*, like the inefficiency of *B*, rests on comparisons to irrelevant alternatives. Like many social choice theorists, we believe that **I** is unacceptable.

Arrow also includes two formal principles that constrain social welfare functions:

U No matter what the preferences of individuals may be, the social welfare function must always be able to rank alternatives (universal domain).

CR The social ranking of alternatives must be complete and transitive (collective rationality).

The rationale behind **U**, the universal domain condition, is that the method of evaluating social alternatives should not collapse when there are strange arrays of individual preferences. If the social welfare function is interpreted as a method of collective decision making rather than as a means of evaluating alternatives, then the universal domain condition seems obvious: a method of decision making must have some output or other. The plausibility of **CR** (collective rationality) is a complicated matter. Transitivity of social rankings seems as compelling as transitivity of individual preferences, but completeness of social rankings seems as questionable as completeness of individual preferences – at least if the social welfare function is interpreted as a way of evaluating alternatives.

Arrow proved that the only complete and transitive social welfare function that satisfies the Pareto principle and independence of irrelevant alternatives for all profiles of individual preferences is the dictatorial one where the evaluation of alternatives matches the preferences of a single individual! Or, to put it more dramatically (and much less precisely), the only universally applicable way to amalgamate individual evaluations of alternatives into social evaluations is to make one individual a dictator. However intuitively appealing these five conditions might appear to be, one cannot consistently endorse them all. At least one of them has to be abandoned. Mathematics here leads morality by the hand and gives it a rude shaking. These five conditions are not mutually consistent, so they cannot all be imposed.

13.2 The Interpretation of Arrow's Theorem

Sen helpfully distinguishes two broad purposes of aggregation: deciding and evaluating. For example, one might want an aggregation mechanism (1) to decide which movie a group will see or (2) to determine which among a set of movies was best. The criteria for judging choice mechanisms might be very different from the criteria for judging evaluation methods. Decision making and evaluating are not the same task. One reason why they differ is that decision rules may have procedural virtues. The best way to decide which movie to see might be by majority vote, since that shows respect for everyone's wishes, whereas the best way to evaluate movies might be to rely on the decision of an expert.

One might want to impose different conditions on decision rules than on methods of evaluation. For example, a ranking of x and y cannot express

	Purposes	
Inputs	Evaluation	Decision making
Preferences	Type 1: Preference-based evaluation	Type 2: Preference-driven decision making
Judgments	Type 3: Judgment-based evaluation	Type 4: Judgment-driven decision making

Figure 13.2.1. Interpretations of Social Choice Theories

the judgment that x is better than y unless the ranking is transitive. So it seems more imperative for social evaluations than for social decisions to be transitive. On the other hand, it is not obvious that a social evaluation must respect the preferences of individuals, so the Pareto principle is more compelling for social decisions than for evaluations (see Broome 1989).

As the movie example suggests, just what is being aggregated also differs among cases. Sen distinguishes among "interests, preferences, judgments, and views" as things one might want to "add up." This list could be extended to other items, including capabilities, resource endowments, and so on. Any of these properties of individuals might be aggregated for purposes of social decision making or evaluation. Arrow's theorem and the conditions discussed in the previous paragraph are typical of the social choice literature in treating the items to be aggregated as preferences, though, as a formal structure, Arrow's theorem applies to other possible inputs into social decision making.

To illustrate the range of possibilities, let us consider two interpretations of the individual attitudes that constitute the inputs into social choice. One possibility is that they are personal preferences among states of affairs – that is, rankings with respect to everything that matters to individuals. Another possibility is that they are judgments of policies or outcomes with respect to overall worth or with respect to some specific objective, such as security or economic growth. One can then draw the 2×2 table shown as Figure 13.2.1.

This figure illustrates four possible interpretations of work relating individual preferences or judgments to social decisions or evaluations. A social welfare function of Type 1 – one that attempts to provide a social evaluation based on profiles of individual preferences – is a branch of moral philosophy, and its theorems can reveal ambiguities or hidden implications of accepted principles. A simple illustration of an interest aggregation mechanism based on a normative principle (i.e., of a Type 1 theory) is a utilitarian social welfare function, which defines the socially best outcome as that which maximizes the sum of individual utilities (where utility is an index of preference).

It is important to be clear on the point that what is involved in classical utilitarianism is *not* an attempt to aggregate individual normative judgments about what is socially best into a social judgment; utilitarianism has not been espoused as a Type 3 theory. Preference utilitarians instead judge that the best state is that which maximizes aggregate preference satisfaction, and they reject all conflicting judgments. Hence there may be tensions within utilitarianism to the extent that people's preferences depend upon their commitment to nonutilitarian moral principles. Notice, by the way, that a utilitarian social welfare function satisfies four of the five conditions that give rise to Arrow's theorem (the Pareto principle, nondictatorship, universal domain, and collective rationality). It violates independence of irrelevant alternatives, since the total utility of an alternative *A* as compared to *B* does not depend exclusively on individual ordinal rankings of *A* and *B*.

13.3 Social Choice Theory and Moral Philosophy

Arrow's theorem has shaped a large portion of the field of social choice theory, and it has had a major impact on contemporary moral philosophy and on the way mainstream economists conceive of normative problems in general. Yet social choice theory is not exhausted by extensions and reactions to Arrow's theorem. Work we have already discussed – such as the welfare theorems of Chapter 5 or the formal work we alluded to in Chapter 11 concerning envy-free distributions – count as a part of social choice theory, as does a good deal of cooperative game theory (upon which we shall comment in the next chapter). Much of this work aims constructively to clarify principles of morality and rationality and to determine their implications.

However, the interpretation of theorems in social choice theory and even its very definition remain elusive and controversial. Sen suggests that social choice theory can be defined either by its approaches or by its subject matter, which he describes as follows: "Social choice problems arise in aggregating the interests, or preferences, or judgments, or views, of different persons (or groups) in a particular society" (Sen 1986, p. 214). Since aggregating interests and resolving conflicts is central to moral philosophy, a large part of the subject matter of social choice theory *is* moral philosophy. In more recent work, Sen (1999) argues that widening the informational basis of social choice theory beyond preferences or judgments may enrich social choice theory and lead to constructive results.

The methods of social choice theory are both formal and axiomatic. Works in social choice theory are devoted to proving theorems. Although one might regard social choice theory as anything that employs its approach *or* is concerned with its subject matter, what many people think of as social choice theory is in fact the combination of both – that is, the proof and interpretation of theorems concerning the social aggregation of preferences, judgments, and interests. In agreement with Sen's recent work, we believe this view is too narrow and that the subject matter of social choice theory can be broadened. The relevance of social choice theorems to morality depends on exactly what is being aggregated, and for what purpose.

Social choice theorems typically raise moral questions. The most obvious concern the axioms: Do they express plausible normative principles? But we may also ask questions about subtler features of a formal setup. For example, treating the relationship between profiles of individual preferences and a specific social choice as literally a *function* implies the uniqueness of social choice, which may be doubted. Furthermore, we can question whether a particular aggregation mechanism as a whole is morally justifiable, even when we cannot pinpoint the source of our objections. For instance, one may object to a process for aggregating interests for purposes of decision making (a Type 2 theory) as unfair, perhaps because of the way it responds to minority interests. Or one may object to a process for aggregating individual normative judgments into a collective judgment (a Type 3 theory) as unwise in that it fails to respond to the quality of the arguments supporting a particular judgment. Finally, one may ask moral questions about the interests, preferences, and judgments upon which the collective decisions or evaluations depend. In Chapter 8 we reviewed the likelihood that individual preferences may need to be "laundered" before they should be granted moral weight, and there may be good moral reasons to launder interests, views, or judgments as well. One might, for example, want to make sure that the informational basis of these items was adequate.

If the axioms and formal setup pass preliminary scrutiny, theorists can then argue from the results of social choice theory to moral conclusions. If one accepts the axioms of some proof, then one must accept the conclusions. Axiomatic arguments of this kind have been proposed for ethical conclusions as varied as utilitarianism, a utility analogue of Rawls's difference principle, and welfare egalitarianism (Roemer 1988). For example, in addition to making a veil of ignorance argument for utilitarianism, John Harsanyi (1955) demonstrates that if (1) personal and moral preferences satisfy the axioms of expected utility theory and (2) the Pareto principle is satisfied, then the social evalution of alternatives is a linear function of individual utilities. Harsanyi takes utility to be an index not of actual but rather of corrected and cleansed preferences (see Section 8.4). Harsanyi interprets this theorem as an argument for utilitarianism. Others, such as Sen (1976) and Weymark (1991), have questioned this interpretation of the result on the grounds that Harsanyi has not shown why the expected-utility representation of individual and ethical preferences is the right one to use for ethical comparisons. Harsanyi would counter this criticism by maintaining that only the expected-utility representation captures attitudes toward risk and, implicitly, information about preference intensity (1977a; see also Mongin and d'Aspremont 1998, pp. 434–5).

Although a good deal of social choice theory looks as if it consisted of formal arguments for moral conclusions, this is not the only way to appreciate this work. Logically valid arguments show only that one cannot both accept all their premises and reject any of their conclusions. If one is convinced of the correctness of their premises, valid arguments can persuade one to accept their conclusions, but if one is convinced of the incorrectness of the conclusions, then it is equally true that valid arguments can persuade one to doubt the premises. In general, people's confidence in the axioms will derive both from how reasonable they seem to be in their own right and from how reasonable are the conclusions they entail. If the axioms entail a disturbing moral conclusion, then people will be motivated to scrutinize them more closely and to reflect on whether their negative reaction is justified.

We suspect that the most important role of theorems in social choice theory has thus far been to reveal that apparently plausible moral principles are more ambiguous and problematic than they appear. Such ambiguities and difficulties have been highlighted most conclusively by demonstrations that sets of plausible principles may be logically inconsistent. Arrow's theorem is the most famous of these demonstrations. But it is not easy to say exactly what moral lessons should be drawn, because the theorem can be read in at least the four ways shown in Figure 13.2.1, depending on how the individual rankings are interpreted and on whether the point is to evaluate or to choose. In our formulation of the theorem, we supposed that the rankings specify individual preferences and that the point of the aggregation was evaluation. On this preference-based evaluation reading, Arrow demonstrates that individual welfare (as ordinal preference satisfaction) cannot provide an acceptable basis for evaluating alternatives. If one sticks to the interpretation of individual rankings as preferences but regards the social welfare function as a decision-making method, then one can read Arrow's theorem as showing that there is no acceptable method for reconciling conflicts among individual interests. But as Sen has observed, these results may not be as distressing as they appear because individual ordinal rankings provide little information concerning the relative importance of different interests. If interpersonal comparisons of preference intensity or nonpreference data (say, on comparative income or resource levels) are permitted (which violates **I**), then no analogous impossibility result holds (Sen 1970a, chs. 7, 8).

One might instead interpret the individual rankings as (moral) *judgments* about what is best for society to do. If the point of the aggregation is evaluation, then Arrow's theorem tells us that no acceptable scheme for aggregating individual moral judgments will produce a consistent evaluation. If the point of the aggregation is social decision making, then Arrow's theorem implies that no rational decision-making method respects individual moral judgments. It does not seem appropriate to avoid these impossibility results by comparing the intensity with which individuals hold their views of which alternatives are best for the society. Yet one can take some of the sting out of the impossibility results when one remembers that one of Arrow's conditions – universal domain – requires that an aggregation scheme work for all possible preference profiles. The Arrow conditions do not allow any sets of moral rankings to be ruled out as improper, and they do not allow for any modification of individual moral views in response to discussion. When constraints are imposed on preference profiles, the impossibility results may no longer hold (though it takes great optimism about the power of reason to be confident that they won't). Arrow's theorem need not, then, undercut practices such as weighing the importance of competing interests or deliberating over the persuasiveness of arguments (and Arrow has never suggested otherwise).

13.4 The Paradox of the Paretian Liberal

One crucial element in liberalism is the view that individuals should have a "private sphere" in which they can do as they please without social constraint. However unwise it may be not to brush one's teeth, the decision should be up to the individual. It is difficult to translate this aspect of liberalism fully into the formalism of social choice theory, but Sen has articulated what he takes to be a weak necessary condition, "minimal liberalism," that those who value freedom would endorse. Minimal liberalism states that there must be at least two individuals, each of whom is "decisive" over a pair of alternatives. An agent is decisive between two alternatives if her preferences among them determine their social ranking. Minimal liberalism would be satisfied if, for example: (1) Hausman got to decide between two alternatives that differed only in regard to whether Hausman brushed his

Lewd	Prude
(b) both read the book	(n) neither reads the book
(p) only Prude reads it	(p) only Prude reads it
(l) only Lewd reads it	(l) only Lewd reads it
(n) neither reads it	(b) both read it

Figure 13.4.1. Lewd's and Prude's Preferences

teeth; (2) McPherson were similarly decisive over a pair of alternatives that differed only with respect to whether McPherson brushed his teeth; and (3) no one else were decisive over any other pairs. So minimal liberalism is indeed minimal!

Sen (1970b) then demonstrates that one cannot consistently espouse (1) the Pareto principle, (2) minimal liberalism, (3) unrestricted domain, and (4) a variant of collective rationality.

An example that Sen uses to illustrate the conflict involves two individuals, "Lewd" and "Prude," and a copy of D. H. Lawrence's novel, *Lady Chatterley's Lover.* Lewd most prefers that both he and Prude read *Lady Chatterley's Lover.* If only one should read it, however, Lewd wishes it to be his prudish friend, since it would loosen him up. Prude would most prefer the book not be read at all, but would rather read it himself than further his friend's corruption. So their preferences are as in Figure 13.4.1.

Given minimal liberalism, Lewd should be decisive between either (b, p) or (l, n) and Prude should be decisive between either (n, p) or (l, b). Suppose Lewd is decisive over (b, p) and Prude is decisive over (l, b). Since Lewd is decisive between b and p and prefers b to p, it follows that b should be socially preferred to p. And since Prude is decisive between l and b and prefers l to b, l should be socially preferred to b. So if social choice respects minimal liberalism then the outcome should be l, that only Lewd reads the book. But both prefer p – that Prude read the book – to l, and the Pareto principle consequently implies that p should be socially preferred to l. If one respects both minimal liberalism and the Pareto principle, then there is a cycle and collective rationality is violated.

Allan Gibbard pointed out that if one strengthens Sen's minimal liberty condition to permit two individuals to be decisive over *all* pairs of alternatives that differ only with respect to purely personal matters (such as whether Lewd reads *Lady Chatterley's Lover*), then certain combinations of preferences will rule out the possibility of any social choice, regardless of questions of Pareto optimality. In Gibbard's example of the nonconformist and Mrs. Grundy (1974, p. 389), the nonconformist wants his bedroom walls to be a different color from Mrs. Grundy's, while she wants her bedroom walls to be the same color as the nonconformist's. If each is decisive over social alternatives that differ only with regard to the color of his or her bedroom walls, then every possible outcome will be vetoed by somebody.

Sen's example and his theorem are mainly concerned with personal liberty over private matters (Riley 1989, 1990). One essential feature that generates the paradox is that people are assumed to have preferences over outcomes that properly – according to the liberty principle – belong in someone *else's* private sphere. Indeed, Blau (1975) shows that Sen's paradox does not arise in the two-person case unless at least one person has "meddlesome" preferences in a technically defined sense. The Pareto principle, as applied in this context, treats these preferences on a par with personally oriented ones.

Gibbard (1974), Nozick (1974), Gärdenfors (1981), Sugden (1985), and Gaertner, Pattanaik, and Suzumura (1992) suggest that the paradox arises from a misconstrual of rights. If rights imply decisiveness, then Prude's and Lewd's rights must be violated if Prude reads *Lady Chatterley's Lover* and Lewd does not. For Lewd is decisive between p and b and prefers b. But there seems to be no violation of their rights if they voluntarily reach an agreement whereby Prude promises to read the book and Lewd promises not to read it. In that case the decision whether to read the book or not would still be up to each of the individuals. So without sacrificing their rights, Prude and Lewd can improve on the suboptimal outcome where only Lewd reads the book. In the view of Sen's critics, there appears to be a conflict between rights and the Pareto principle only because rights have been mischaracterized in terms of decisiveness. Rights should instead be modeled as sets of strategies in suitable game forms. What this means, roughly, is that an individual has a right to read or not to read a book if the choice is up to that individual.

This game-form "dissolution" of the paradox faces two serious problems. First, Riley has shown (1989; see also Basu 1984) that, in games corresponding to the social choice problems in which there is a clash between rights and Pareto optimality, the Pareto optimal outcomes will not be "Nash equilibria" (see Section 14.5). Each agent has an incentive to break his promise unilaterally. So Lewd's and Prude's promises would give rise to enforcement problems, and enforcing them would require illiberal interference in the private affairs of individuals. As Sen notes, it is thus questionable whether such contracts or their enforcement are morally desirable. Second, as Sen (1992b) argues, the game-form interpretation of rights seems too weak. For on this interpretation, Hausman has the right not to brush his teeth even if his dentist will shoot him otherwise. Sen argues that it is hard to see how any system of rights that provided genuine protection to individual liberty could fail to satisfy his minimal liberty condition – to make at least two individuals decisive over a single separate pair of social alternatives. So the paradox seems real. In Sen's view, its moral is that normative social policy cannot be based only on preferences.

One might instead try to resolve the liberal paradox through relaxing the Pareto principle. Why should preferences regarding other people's private activities be accorded the same moral importance as those regarding one's own activities or public matters? Indeed, acknowledgment of a private sphere of activity for a person creates a presumption that nosy preferences should receive less weight. Other moral arguments have been offered for discounting other-regarding preferences in making public decisions (see Dworkin 1977, ch. 12). Sen's paradox underscores the importance of permitting other things besides preferences to count. In particular, one may want to examine the reasons behind preferences in deciding on the role they should play in social decisions and moral judgments – a point we have made elsewhere in this book.

13.5 The Range of Social Choice Theory

The discussion so far gives a meager idea of the range of issues that social choice theory considers. As the three examples we discuss in this section illustrate, the tools of social choice theory have many applications and have given rise to many surprising results.

13.5.1 The Logical Coherence of Social Judgments

Suppose society consists of three individuals who make the judgments shown in Figure 13.5.1 concerning the truth or falsity of the propositions *P* and *Q*, and suppose that social judgment follows the majority. The judgments of each of the individuals are consistent with the principles of logic, while social judgments violate them (List and Pettit 2002).

Unlike the material discussed earlier, no ethical principles are explicitly invoked here; the issue concerns rationality. As the example shows, majority decision making can easily lead to logical inconsistency. How important is it that social judgments be consistent with the principles of logic? What can be done about the disturbing possibility of inconsistency?

13.5.2 Formal Representations of Freedom and Opportunity

In our discussion of freedom in Section 10.1, we talked about the factors that influence the extent of an individual's freedom. How can this be measured?

	P	If P then Q	Q
Individual 1	True	True	True
Individual 2	False	True	False
Individual 3	True	False	False
Society	True	True	False

Figure 13.5.1

Similarly, in Chapter 11 we discussed an interpretation of egalitarianism in terms of equality of opportunity. That too raises the question of how to understand and measure the extent of the opportunities available to individuals. Social choice theorists have put their formal tools to work in attempting to clarify the notion of the extent of opportunity or freedom.

Let X be the set of all possible alternatives among which people might choose. Let $A, B, \ldots$ be subsets of X which are the actual sets of alternatives among which particular individuals choose. Some normative economists have been interested in characterizing when $A \geq B$, $A > B$, or $A \approx B$, where "$A \geq B$" is to be read "A provides at least as much freedom or opportunity as B," "$A > B$" means "A provides more freedom than B," and "$A \approx B$" symbolizes "A and B provide the same amount of freedom." These economists have attempted to formulate some plausible conditions comparing the freedom provided by different sets and to investigate their implications. For example, it seems plausible that (i) opportunity sets that permit no choice at all – that contain only one member – all provide the same (zero) amount of freedom. It also seems plausible to maintain that (ii) if A contains all the options in B and some others in addition, then $A > B$. Finally, it seems plausible that (iii) if some option x is not in sets A or B and if A^* and B^* are the sets that result from adding x to A and B (respectively), then $A \geq B$ if and only if $A^* \geq B^*$. But these three conditions imply that the opportunity or freedom of a set can be measured by the number of alternatives it contains.† As Sen (1990) has argued with particular force, this implication is clearly unacceptable. It means, for example, that a worker has much

† Condition (i) tells us that $\{x\} \approx \{y\}$ and $\{z\} \approx \{w\}$ for any four distinct alternatives. If we add z to the set $\{x\}$ and also to the set $\{y\}$, then (iii) implies that $\{x,z\} \approx \{y,z\}$. If we add x to the sets $\{z\}$ and $\{w\}$, then (iii) implies that $\{x,z\} \approx \{x,w\}$. So $\{y,z\} \approx \{x,w\}$ – that is, all two-member opportunity sets provide the same amount of freedom (which is, by (ii), more than the freedom in one-member opportunity sets). In the same way we can show that, if all n-member opportunity sets provide the same amount of freedom, then all $(n+1)$-member opportunity sets provide the same amount of freedom, and by mathematical induction it follows that every two sets with the same number of alternatives provide the same amount of freedom. See Pattanaik and Xu (1990).

more freedom if she can only work at the counter of any one of fifty different McDonald's restaurants than if she has a choice between ten entirely different jobs. So at least one of the three plausible premises that imply this conclusion must be mistaken.

Sen emphasizes that the *quality* of the alternatives bears on the extent of the opportunities a set provides. "We find it absurd to dissociate the extent of our freedom from our preferences over the alternatives" (1990, p. 470). One simple way of taking this into account (which Sen does not espouse) is to measure opportunity by the utility of the best alternative that a set makes possible. But proceeding this way simply collapses the concern with freedom back to a concern with welfare. Freedom is not a matter of how well-off one is. A military conscript who is assigned a great job may be much better-off than a civilian who has a choice among a variety of worse jobs, but the conscript does not have a wider range of opportunities or more freedom.

Others concerned with the formal representation of freedom have focused on the *diversity* of the alternatives available to an individual. The concern with diversity links up to a concern with the quality of alternatives insofar as diversity depends in some way on preferences. One might, for example, consider not just an individual's actual preferences but also a range of different preference rankings than an individual might have or might have had. Arrow (1995) thus links our interest in freedom to our interest in *flexibility*. One might then argue that adding an alternative to an opportunity set increases freedom or opportunity if and only if the alternative comes out best in at least one of the admissible preference rankings. However, this sort of view can easily lead to the conclusion that if every admissible preference ranking ranks some alternative x in the opportunity set A higher than every alternative in B, then A provides more freedom than B (Sugden 1998). But A may contain very few alternatives while B contains a large number of highly diverse alternatives.

These remarks provide only a small taste of a growing literature. The axioms we presented are simple, and some philosophers might condemn them as simplistic. But the efforts of social choice theorists to characterize the extent of freedom or of opportunity have highlighted a serious lacuna in the philosophical literature. Without some coherent concept of more or less freedom, of greater or lesser opportunities, it is fruitless to talk about maximizing freedom or equalizing opportunity.

13.5.3 Should Egalitarians Aim to Equalize Welfare?

John Roemer (1988, pp. 160–72) offers a plausible set of axioms that together imply equality of welfare. Roemer invites us to consider the distributional principles that should govern a hypothetical economic environment in which there is a single output, "corn," which is produced and consumed by two individuals, Able and Infirm, who have the same preferences for bundles of corn and labor but very different skills: Able can produce much more corn per hour of labor than can Infirm. If egalitarians want to compensate Infirm yet remain fair to Able, what overall egalitarian principle should they follow? Just as social choice theorists – in their attempts to define a social welfare function – sought conditions relating individual preference and social welfare, so Roemer proposes moral principles governing the allocation of corn to Able and Infirm. His conditions are as follows.

1. (Pareto Optimality) The corn allocation should be Pareto optimal.
2. (Land Monotonicity) If the amount of land increases, neither individual should be made worse-off.
3. (Technological Monotonicity) If it becomes possible to produce more corn with the same inputs of land and labor, neither individual should be made worse-off.
4. (Limited Self-Ownership) Able should not be worse-off than Infirm.
5. (Protection of Infirm) Infirm should be at least as well-off as he would be if Able had no more skills than Infirm has.

These moral principles seem attractive, though they are certainly not uncontroversial. Greater productivity should not harm either individual. Able should not do worse than Infirm; Infirm should not lose in absolute terms from Able's greater skills. And the allocation of corn should be Pareto optimal. Yet Roemer proves that the only method of allocating corn that satisfies these five rules is to equalize Able's and Infirm's utility! The weak egalitarianism that makes rules such as protection of Infirm or the monotonicity conditions attractive, when combined with limited self-ownership and the Pareto principle, implies an incredibly strong egalitarian conclusion. One reaction would be to go back to Chapter 11 and rethink the arguments against equalizing welfare. Perhaps the criticisms described there were premature and equalizing welfare is, after all, an appropriate ideal. Alternatively, one can reconsider Roemer's five conditions and their role in the proof of his theorem.

The point of discussing these examples is not to explore the logical consistency of social judgment, to establish a measure of freedom, or to reopen the discussion of egalitarianism. Instead, our object has been to show how the techniques of social choice theory can be employed with different inputs and for different purposes. The techniques of social choice theory yield surprising results concerning a wide range of issues.

13.6 Conclusions

What do examples such as Arrow's impossibility theorem and Sen's paradox, explorations of implications of axioms concerning freedom, and Roemer's surprising demonstration tell us about the fruitfulness of social choice theory in illuminating moral issues? Its typical subject matter – especially the problem of assembling individual judgments or interests into social evaluations – is of undeniable moral importance. But we are ambivalent about the value of its axiomatic approach.

We are not ambivalent about the value of precision and rigor. There is no doubt that the formal approach can lead to important clarifications and can stimulate a new awareness of the depth of problems – as the literature on the liberal paradox or the measure of freedom shows. At the same time, the abstractness of these methods has pitfalls. To quote Sen: "Arrow was undoubtedly right in saying that 'one of the great advantages of abstract postulational methods is the fact that the same system may be given several different interpretations, permitting a considerable saving of time' (Arrow 1951, p. 87). But *that* probably is *also* one of the great disadvantages of these methods" (Sen 1982a, p. 190). Considerable confusion has been generated, particularly in discussions of Arrow's theorem, by a failure to distinguish among alternative interpretations of the objects being aggregated and the purposes of aggregation.

It is difficult to spend much time in the social choice literature without becoming aware of the disparity between the large amount of attention lavished on the derivation of theorems and the comparatively small amount devoted to their interpretation. Unless guided by ethical sophistication, social choice theory can degenerate into formal exercises. But it would be unfair to single out economists for blame. Moral philosophers ought to be more involved with this work. If formalizations of plausible moral principles lead to inconsistencies or to unacceptable conclusions, then philosophers and economists need to understand what has gone wrong. It may be that the formalizations fail to express the informal principles they attempt to restate, or it may be that there are ambiguities or problems within those principles themselves. Only careful interpretation of both the informal moral principles and their formal restatements enables one to appreciate the significance of a particular theorem. Social choice theorists have left lots of philosophical work to be done.

Suggestions for Further Reading

One of the best general introductions to social choice theory is still Sen (1970a). For more recent introductions and surveys see Arrow, Sen, and

Suzumura (1996, 1997); Barbera, Hammond, and Seidl (1998, 1999); Sen (1999); and Arrow, Sen, and Suzumura (2002). Arrow's theorem was first presented in Arrow (1951), a classic. For a valuable book-length philosophical discussion see MacKay (1980). A range of critical and constructive essays on social choice is presented in Elster and Hylland (1986).

Sen first presented his paradox in 1970. For some of the major discussions, see Sen (1976a, 1983, 1986, 1992b); Gibbard (1974); Nozick (1974); Gärdenfors (1981); Riley (1989, 1990); and Gaertner et al. (1992). For recent developments of a game-form construal of rights, see Fleurbaey and van Hees (2000).

Vickrey (1945), Fleming (1952), and Harsanyi (1955) present axiomatic arguments for utilitarianism. Re-interpreted and spruced-up contemporary versions of something like Harsanyi's argument can be found in Broome (1991b), Mongin and D'Aspremont (1998), and Mongin (2001). Harsanyi's interpretation of his theorem as an argument for utilitarianism is challenged in Sen (1976) and Weymark (1991).

For axiomatic arguments for a utility analogue to Rawls's difference principle, see Hammond (1976), Strasnick (1976), and d'Aspremont and Gevers (1977). Issues concerning the logical consistency of social judgments are explored in List and Pettit (2002, 2004).

There is a large literature on formal representations of the extent of freedom and opportunity. See particularly Jones and Sugden (1982); Pattanaik and Xu (1990); Sen (1990); Arrow (1995), Laslier et al. (1998); Carter (1999); and especially Sugden (1998). For further developments, see Nehring and Puppe (2002) and Bossert, Pattanaik, and Xu (2003).

For some formal investigations of egalitarianism and its consequences, see Fleurbaey (1995, 2002); Kolm (1997); Roemer (2002); Fleurbaey and Maniquet (2005); and Fleurbaey et al. (2005).

FOURTEEN

Game Theory

Game theory was conceived as a tool to explain, predict, and guide behavior in strategic interactions among individuals, and its main applications still lie in this domain. But game theory is less wedded to a particular set of questions than is social choice theory, and it has applications to such disparate phenomena as computer programming and biological evolution. Game theory might indeed be regarded as a branch of mathematics rather than as specifically a theory of interactive decision making. However, in this chapter we will focus exclusively on applications of game theory to strategic interactions among individuals. An interaction among individuals is "strategic" if the payoffs to some individuals depend on the choices of others.

14.1 What Is a Game?

A game is defined by its *players,* the *strategies* that are available to the players, the *payoffs* to the players from each combination of strategies, and the *capacities* of the players, including especially the *knowledge* each player has concerning the other features of the game. Each of the italicized terms has a technical meaning. In the case of the games we shall consider, players are simply people. To keep things simple, we shall for the most part discuss games with only two players. A specific strategy for a player is a complete specification of what actions the player will take whenever the player gets to act. The specification of the strategies available to a player defines all possible actions that the player can take.

Before explaining the notion of a payoff, it will help to provide an example. One way to represent strategic interactions is via an "extensive form" or a game "tree," like the one shown in Figure 14.1.1. This interaction involves two players, player I (Jill) and player II (Jack). Jill moves first at the node labeled *a* and plays Up or Down; these two moves are her two strategies.

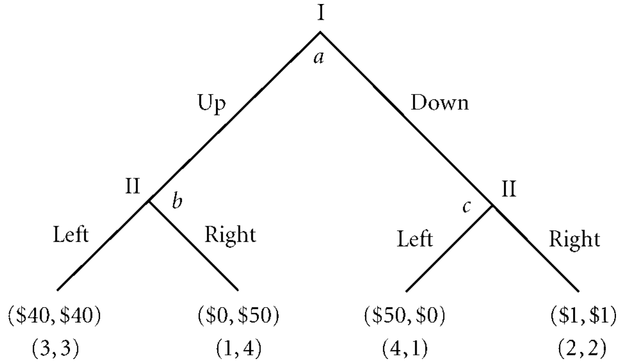

Figure 14.1.1

Player II (Jack) knows how Jill has chosen before he gets to act, so he knows whether he is choosing at node *b* or at node *c*. Although Jack can choose between just two actions at each node, he has four (not just two) possible strategies. His four strategies are: (i) Left regardless (no matter what Jill does); (ii) Left if Jill plays Up, Right if Down; (iii) Right if Up and Left if Down; and (iv) Right regardless. What we are calling "strategies" here are called by game theorists "pure strategies" as opposed to "mixed strategies." A mixed strategy involves playing different pure strategies with different probabilities. If Jack resolved to flip a coin and to play Left or Right depending on whether the coin lands heads or tails, regardless of what Jill does, then he would be playing the mixed strategy of playing Left regardless of what Jill does with probability 0.5 and playing Right regardless of what Jill does with probability 0.5.

In the interaction depicted by Figure 14.1.1, the material results of different combinations of strategies are monetary payoffs. The first of the two dollar amounts at each terminal node represents the earnings of the first player, Jill, and the second amount is how much Jack receives. So, for example, if Jill plays the strategy Up and Jack plays either strategy (ii) (Left regardless) or strategy (iii) (Left if Up and Right if Down), then both receive $40. But knowing the dollar outcomes is not sufficient to define the game. The "payoffs" that actually define the game are the player's *preferences* over the outcomes. The pairs of numbers along the bottom of Figure 14.1.1 are the utilities that the two players attach to the outcomes of the game. As discussed in Chapter 4, utilities are indices of preference. In this illustration, the utilities are ordinal utilities. That the outcome of Jill playing Up and Jack playing Left has a utility of 3 for both players means that it is the second-best outcome for both. The utility Jill attaches to an outcome can depend on the result for Jack as well as on the result for herself, and it may reflect preferences over features of the game besides the monetary payoffs at the end. In this particular example, Jill's and Jack's preferences over the outcomes match what they would be if they cared only about their own monetary return, but this need not always be the case. For example, even if Jack might prefer a state of affairs where an anonymous donor gives him \$50 and Jill gets nothing to a state of affairs where a donor gives both \$40, Jack might believe that in this interaction he has an obligation to reward Jill's trust or kindness in playing Up. For this reason Jack might prefer the state of affairs where Jill plays Up, he plays Left, and they both receive \$40 to the state of affairs where Jill plays Up, he plays Right, Jill gets nothing, and he gets \$50. The crucial point is that the payoffs in games are preferences or utilities, not physical outcomes or quantities of money. The monetary amounts shown in Figure 14.1.1 are redundant and irrelevant to the game, strictly speaking.

The final ingredient in the definition of a game is a specification of what the players know. In Figure 14.1.1, for example, Jack knows whether Jill has played Up or Down before he chooses, and Jill knows that Jack knows this and that Jack knows that Jill knows that Jack knows this, and so forth. In the terminology of game theory, the game shown in Figure 14.1.1 is one of perfect and complete information. Everything about the game is *common knowledge.*

The games with which we will be concerned involve players who are rational and intelligent, who have perfect recall of earlier moves in the game, and who possess common knowledge of the extensive form of the game and of their common rationality, intelligence, and memories. Game theorists usually take rationality to be expected utility maximizing. The players are "intelligent" in the sense that they can figure out anything about the game that the game theorist can. In advanced work, game theorists have found ways to relax many of these assumptions, and in other applications of game theory (such as evolutionary game theory) players can be unintelligent entities hardwired to particular strategies.

To show how sensitive the treatment of standard games is to the knowledge players have, let's say that Jack does *not* know how Jill has chosen before he has to move. We suppose again that Jill can play Up or Down and that Jack can play Left or Right but that the payoffs – that is, the preferences or utilities – are now as shown in Figure 14.1.2.

In the figure, choice nodes *b* and *c* are enclosed within a dashed oval; that oval is an "information set." The player who must choose does not know at which of the nodes in the information set he or she is choosing. In this

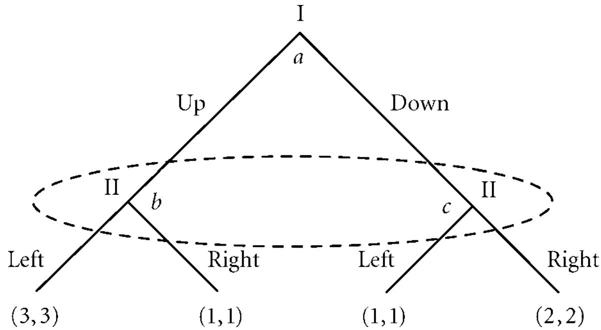

Figure 14.1.2

case, Jack does not know whether Jill has played Up (which would put Jack at node *b*) or Down (which would put him at node *c*). Hence Jack's strategies cannot specify different moves at node *b* and at node *c*. Whereas in the game of Figure 14.1.1 Jack had four strategies, in the game of Figure 14.1.2 he has only two strategies: to play Left or to play Right.

In addition to extensive-form representations of games, such as those shown in Figures 14.1.1 and 14.1.2, games can also be represented in "normal" (or "strategic") form as in Figures 14.1.3a and 14.1.3b. Figure 14.1.3a shows the player's preferences for the strategy combinations in the game depicted in Figure 14.1.1, while Figure 14.1.3b shows the player's preferences for the strategy combinations in the game depicted in Figure 14.1.2. ("Lu" abbreviates "play Left if Jill plays Up," etc.) The strategic form contains information concerning possible strategies and their payoffs, but it is silent about the order of play.

Game theorists are concerned with the "solution" to different kinds of games. That is, they seek to determine what are the best strategies for players to employ and consequently what the outcomes of games for rational and intelligent players will be. In thinking about the solutions to games, theorists distinguish cooperative from noncooperative game theory. The difference is that, in cooperative game theory, the players are able to make binding agreements. This makes possible more advantageous outcomes, because the players need not pay the costs of enforcing agreements and do not face the problem that sometimes agreements cannot be enforced at all. On the other hand, cooperative game theory does not explain *how* individuals are able to make binding agreements and is thus regarded by most game theorists as less fundamental. Except for the discussion of bargaining theory in Section 14.5, we shall focus on noncooperative games.

Player II (Jack)

		Lu Ld	Lu Rd	Ru Ld	Ru Rd
Player I (Jill)	Up	3, 3	3, 3	1, 4	1, 4
	Down	4, 1	2, 2	4, 1	2, 2

Figure 14.1.3a

Player II (Jack)

		Left	Right
Player I (Jill)	Up	3, 3	1, 1
	Down	1, 1	2, 2

Figure 14.1.3b

The most important solution concept in noncooperative game theory is that of a Nash equilibrium. A Nash equilibrium consists of a set of strategies such that each strategy is a "best answer" to the strategies the other players have chosen. In Figure 14.1.3a, for example, the strategy pair (Down, [Ru, Rd]) is the unique Nash equilibrium strategy pair. If Jill plays Down then Jack has no better move than to play Right (and no better strategy than [Ru, Rd] – though [Lu, Rd] is equally good), and if Jack plays [Ru, Rd] then Jill has no better strategy than Down. Even though (Down, [Lu, Rd]) has the same payoffs it is not a Nash equilibrium, because if Jack plays [Lu, Rd] then Jill's best reply is to play Up rather than Down. The appeal of the notion of a Nash equilibrium is that neither player can do better in terms of his or her own preferences, given the strategy the other player employs.

In the game shown in Figures 14.1.2 and 14.1.3b, there are two Nash equilibria: (Up, Left) and (Down, Right). The latter might seem irrational, but given that Jill is playing Down, Jack would do worse by playing Left, and given that Jack is playing Right, Jill would do worse by playing Up. As both the games illustrate, the outcomes of Nash equilibrium strategies need not be Pareto optimal. Some games have many Nash equilibria in pure strategies, while others have none. If one also takes into account mixed strategies then, as Nash (1951) proved, every game has at least one Nash equilibrium set of strategies.

14.2 Moral Philosophy and Some Simple Games

Much of game theory is not concerned with moral questions, and moral philosophy bears less directly on many games than on social choice theorems. But the subject matter of the game theory with which we are concerned – human interactions – is also the subject matter of much of ethics, and the problems faced by players are often moral problems. So game theory and moral philosophy have a good deal to do with one another.

Human flourishing requires social cooperation but, as philosophers have long recognized, such cooperation is problematic. In Hobbes's state of nature, where there is little cooperation, "the life of man [is] solitary, poor, nasty, brutish, and short" (1651, p. 107). The theory of games is particularly relevant to moral philosophy in its revelation of the sorts of interaction problems with which moral principles must deal. For example, Hardin (1988) argues that the complications of strategic interactions throw a wrench into both applications and criticisms of utilitarianism. A sophisticated utilitarianism that recognizes these factors may be able to defend rights assignments as devices for simplifying strategic interactions and avoiding suboptimal outcomes. Kantian moral philosophy may similarly profit from game-theoretic analyses of the consistency of everyone following the same rule (Schelling 1978; Elster 1989a).

Problems of social cooperation (or social cooperation "games") are often complicated, and it can be enlightening to think about recurring patterns. Consider the game shown in Figure 14.2.1. Jill and Jack both have two strategies, and the payoffs (utilities) are as stated. Notice that Jill and Jack both have "dominant" strategies: no matter what Jack does, Jill does better by confessing; and no matter what Jill does, Jack does better by confessing. Yet the result seems paradoxical, for if there were some way for Jill and Jack to cooperate and each not confess, then both would be better-off.

The game in Figure 14.2.1 is called "the prisoners' dilemma" because of a story that goes with it. Jill and Jack are two prisoners to whom the district attorney separately offers the choice between confessing and not confessing. The best outcome for Jill occurs if she confesses and Jack does not. Second best is if both refuse to confess. Third best is if both confess. Worst of all is not to confess when the other confesses. And so each, pursuing his

		Player II (Jack)	
		Don't confess	Confess
Player I (Jill)	Don't confess	3, 3	1, 4
	Confess	4, 1	2, 2

Figure 14.2.1. Prisoners' Dilemma

or her own interests, winds up confessing. Rational self-interest leads to an outcome that is suboptimal for the two criminals – though not necessarily to the society.

Prisoners' dilemmas vividly represent problems of social cooperation, free riding, and public goods provision. Individuals deciding whether to contribute to the production of a public good seem like players in an *n*-person prisoners' dilemma game. Free riding – that is, enjoying the public good but not contributing to its provision – is best of all. Second best is enjoying the public good and contributing one's share. Third best is not contributing and doing without the public good. Worst is having others free ride on one's own contribution. Whether public goods provision is best modeled as a prisoners' dilemma is, however, controversial (Hampton 1987). There are two main grounds for doubt. First, doing without the public good may be worse than having others free ride on one's contribution, in which case one has a game of "chicken" (see Figure 14.2.5) rather than a prisoners' dilemma. Second, problems involving public goods provision arise in a context of continuing social interactions, and it is often misleading to model them as one-shot or single-play games. For the "supergame" consisting of a sequence of moves that by themselves would be prisoners' dilemmas typically does not itself have a prisoners' dilemma structure.

Prisoners' dilemmas can also model problems of moral cooperation (Parfit 1979). For example, each mother trying to save her own child from drowning in a beach accident may do less well than all would do if they cooperated in saving all the children. The collectively self-defeating character of moral commitments that are subject to this problem is disturbing, and one might want to require that acceptable moral views not be collectively self-defeating. But many familiar moral maxims seem to fail in this way.

Not all problems of social interaction have the structure of a prisoners' dilemma. There are also pure coordination problems (Lewis 1969; Schelling

		Player II (Jack)	
		Drive left	Drive right
Player I (Jill)	Drive left	1, 1	0, 0
	Drive right	0, 0	1, 1

Figure 14.2.2. Pure Coordination

		Player II (Jack)	
		Dinner	Theatre
Player I (Jill)	Dinner	2, 1	0, 0
	Theatre	0, 0	1, 2

Figure 14.2.3. Battle of the Sexes

1978), such as determining which side of the road to drive on. It does not matter if everybody drives on the right or if everybody drives on the left, but it matters a great deal that everybody drive on the same side of the road! The normal form of a pure coordination problem is given in Figure 14.2.2. Pure coordination problems are "easier" in the sense that they have self-enforcing solutions, which can arise spontaneously. Once a custom becomes established, everybody has an incentive to adhere to it. But pure coordination problems also have less bearing on moral problems, in which there are conflicts of interest to resolve.

Closely related to pure coordination games – but of more relevance to moral philosophy – are games with the structure illustrated in Figure 14.2.3. Such games are called "battle of the sexes" games because of the following story. A man and a woman want to go out together for an evening, and they face a choice of what to do. Both would prefer to be together at either activity than splitting up, but Jill would prefer to go out for dinner while Jack would prefer to go to the theatre. As in a pure coordination game, each player wants to do as the other does, but in this case which strategy they jointly decide upon is not a matter of indifference. Many conflicts of interest resemble battle-of-the-sexes games, and though genuine conflicts they

		Player II (Jack)	
		Fix bridge #2	Do nothing
Player I (Jill)	Fix bridge #1	3, 3	0, 1
	Do nothing	1, 0	1, 1

Figure 14.2.4. Assurance Problem

are much easier to resolve than prisoners' dilemmas, because here it is in both player's interests to coordinate their actions.

Now consider the rather different game in Figure 14.2.4. Jill and Jack might both want to do their part toward providing some collective good that requires both their efforts. It is best for each to cooperate if the other cooperates, but it is worst to cooperate if the other does not. For example, a road might be unusable because two bridges are out. Jack is able to repair only one bridge, and Jill is able to repair only one bridge. The road is very valuable to each if both bridges are repaired and worthless otherwise. All it takes in this case to make cooperation individually optimal is "assurance" that the other will cooperate, so this is known as the assurance problem (Sen 1967). In assurance problems, the obstacles in the way of social cooperation are slighter than in a prisoners' dilemma and the prospects for reconciling individual rationality and social optimality are brighter. One strategy for coping with prisoners' dilemmas is to convert them into assurance problems by changing either incentives or personal motivations.

Suppose instead that Jack and Jill face the problem depicted by Figure 14.2.5. They are approaching an unregulated intersection in their cars. Both prefer to speed up if the other slows down. Second best is if both slow down. Third best is deferring to the other driver. By far the worst is if both speed up and crash into each other. As in the assurance game, Jill and Jack have no dominant strategy. This game of chicken can arise when failing to provide a public good is worse than having others free ride on one's contribution or when conflicts over possessions are costly. In such games, cooperation may be individually rational; or, as Sugden (1986) has argued, if the positions of the interactors are not fully symmetrical then conventions can arise directing those who occupy different positions to play different strategies. Consider, for example, the convention of yielding to the car on the right. It enables drivers to avoid accidents while at the same time

		Player II (Jack)	
		Slow down	Go fast
Player I (Jill)	Slow down	3, 3	2, 4
	Go fast	4, 2	1, 1

Figure 14.2.5. Chicken

permitting one of the two cars to avoid slowing down. In addition, unlike a convention such as "smaller cars should yield to larger cars," it regulates traffic fairly by imposing the burden of having to slow down more or less equally. Property rights might be regarded as conventions that solve problems with the structure of chicken, although property rights have also been shaped by explicit legislation.

As illustrated by these last games, circumstances that feature conflicts of interest as well as advantages in cooperation have a variety of structures, and they are certainly not always prisoners' dilemmas (Hardin 1982; Hampton 1987; Taylor 1987, ch. 2).

14.3 Cooperation and Justice

Social life is full of circumstances in which there are enormous gains to cooperation but also considerable conflicts of interest. In some cases, such as one-shot prisoners' dilemmas, the gains to cooperation cannot be achieved by rational individuals, but social life is more often characterized by continuing rather than one-shot interactions and so the possibilities for rational cooperation are not usually that bleak. As we hinted previously, even a structure of interactions resembling a prisoners' dilemma need not prevent self-interested rational agents from cooperating if the interactions are repeated. Suppose that players in what would be a prisoners' dilemma game (if it were played only once) are likely to play again. The game has then changed and, in such *iterated* prisoners' dilemmas, defection is no longer a dominant strategy because cooperative moves can elicit like cooperative moves from others. Though the formal analysis of repeated games is quite complicated (Taylor 1987, chs. 3, 4), simple strategies such as tit-for-tat (cooperate on the first move, then do whatever your opponent did on the previous move) do well against a wide range of alternatives (Axelrod 1984).

A strategy that does strictly better against itself than does any other strategy is an "evolutionarily stable" strategy (Maynard Smith 1982). The biological terminology is not merely metaphor. Animal traits and behavior can be regarded as strategies, and game theory can be applied to the study of evolution. Tit-for-tat is not an evolutionarily stable strategy in an iterated prisoners' dilemma, since other strategies (e.g., the strategy of always cooperating) do as well against tit-for-tat as tit-for-tat does when played against itself. But it can be proved that, when the probability of repeated play is high enough, tit-for-tat is "absolutely (collectively) stable" – that is, it does no worse when played against itself than does any other strategy (Axelrod 1984, pp. 59, 207–9). When the probability of repeated play is high enough, a group of individuals who play tit-for-tat can "invade" a population of players who defect on every move and, if survival depends on success, eventually drive to extinction those who always defect. So cooperation can evolve among rational self-interested agents, and one might be tempted to regard much of morality as the conventional outcome of repeated games.

For example, people sometimes compete for shares of divisible goods. Suppose that (1) the competition is between just two individuals and (2) if their claims are compatible then they both get what they claim, but if they collectively claim more than the total amount of the good then they wind up fighting. Most people would judge that the fair thing to do would be to divide the good 50–50. Why? Where do such judgments of fairness come from, and how are they justified? Although the question of justification seems to be one more for philosophers to address, questions about the origins of norms of fairness seem to be as much sociological as philosophical. It seems that game theory might have something to contribute to both philosophical and sociological accounts of fairness.

To simplify, suppose (unrealistically, for the purposes of illustration) that everybody in the population is limited to one of four strategies: in the case of limited goods, (a) ask for all of the good, (b) ask for 2/3, (c) ask for 1/2, or (d) ask for 1/3. People then meet randomly in circumstances where they compete for goods. Let Jill and Jack be two members of the population who happen to meet. The normal form of the game they are playing might then be as shown in Figure 14.3.1. As before, the numbers within the matrix are utilities: 2/3 of the good is best and fighting over the good is worst. This game has four pure-strategy Nash equilibria. Rather than attempting to figure out what would be the most rational strategy for Jill and Jack to employ, one might suppose instead that they are already wedded to particular strategies. For example, Jill might be the sort of person who asks for

		Jack			
		Ask all	Ask 2/3	Ask 1/2	Ask 1/3
Jill	Ask all	0, 0	0, 0	0, 0	0, 0
	Ask 2/3	0, 0	0, 0	0, 0	3, 1
	Ask 1/2	0, 0	0, 0	2, 2	2, 1
	Ask 1/3	0, 0	1, 3	1, 2	1, 1

Figure 14.3.1. Divide the Good

1/2 and Jack might be the sort of person who asks for 1/3. Thus, when they compete with one another for a good, Jill gets half the good and Jack gets a third (and one sixth is wasted).

Suppose that, in the population as a whole, everybody is at the given moment of time wedded to one or the other of the four strategies. In the course of agent interaction, some strategies prove to be successful and others not. Suppose, in addition, that there is some sort of mechanism that selects for those whose strategies are successful or that permits individuals to update their strategies by comparing how they fared against the average payoff. This is a sketchy example of evolutionary game theory (borrowing from Skyrms 1996, ch. 1). Depending on the details of the dynamics, on whether the pairings are random, and on whether mutuations are possible, one can show that populations are likely to evolve to the point where everyone has the strategy of asking for 1/2. This strategy is evolutionarily stable: any other strategy will do worse against it than the strategy does against itself.

Although abstract and speculative, the account of an equal-division norm that is provided by evolutionary game theory shows how such a norm could have come into existence, why it is stable, and why it is efficient. If stability and efficiency are desirable properties for moral principles, then this account goes some distance toward justifying a norm of equal division, too.

14.4 Paradoxes and Difficulties

Game theory does not always have such reasonable results, and one may question how broad its uses really are for the purposes of moral philosophy. Suppose two players are playing a hundred-move game, where the payoffs for each move are those of the prisoners' dilemma game described by Figure 14.2.1. If they both play conditionally cooperative strategies such as tit-for-tat, the outcome for both will be 300. If they both defect, then the outcome for both will be only 200. Surely they should be able to find a way to cooperate in at least some of the hundred games. Yet there is an argument that apparently shows that, if the number of repetitions is known, then defecting (refusing to cooperate) on every move is the only rational strategy. The argument begins by noting that the hundredth repetition is just like the one-shot game. Cooperating in order to elicit further cooperation from the other player is senseless, because there will be no further repetitions. So defection is the only rational action. Consider then the ninety-ninth repetition. Since cooperation is rationally excluded on the hundredth repetition, cooperation on the ninety-ninth round cannot elicit later cooperation, and defection cannot be "punished." So defection is the only rational course in the ninety-ninth repetition, too. Proceeding one move at a time, this "backwards induction" argument yields the conclusion that the only rational strategy is to defect on every move.

This conclusion seems hard to accept, and it seems that ordinary people have the last laugh on game theorists, for they manage to cooperate quite a lot and consequently do a great deal better in their interactions than would rational game theorists with common knowledge of their rationality. The practical irrelevance of the backwards induction argument may conceivably be explained by uncertainty about the number of plays in which the parties will be engaged. But experiments (see Dawes, van de Kragt, and Orbell 1990) show that, if individuals are permitted to form personal connections (which do not, however, involve any promises or contractual arrangements concerning their play), then they typically cooperate even in what seems to be a one-shot prisoners' dilemma game! Appearances may be deceptive, however, and experimental subjects might prefer the cooperative outcome and hence not be playing a prisoners' dilemma. In the case of a genuine one-shot prisoners' dilemma, cooperating pairs do better than pairs who both defect because each player is benefited by the other. Individuals would do even better by defecting themselves while having a partner who cooperates. To understand how people play requires a sophisticated theory of choice that recognizes the role of norms – including moral norms – in decision making (Bicchieri 2002).

The backwards induction argument also runs into paradoxical difficulties. Suppose Jill fully accepts the backwards induction argument and defects in the first game in the series. But for some reason or other, Jack

Player II (Jack)

Player I (Jill)		Lu Ld	Lu Rd	Ru Ld	Ru Rd
	Up	5, 5	5, 5	0, 0	0, 0
	Down	9, 1	0, 0	9, 1	0, 0

Figure 14.4.1

cooperates. What should Jill do on the next move? Jack's first move decisively refutes Jill's view of the game. It demonstrates that Jack is not rational (or does not understand the game) or that Jack does not believe that Jill is rational (or understands the game) or that Jack does not believe that Jill believes that Jack is rational or.... Moreover, Jack knows that making a cooperative move will pose this perplexing problem for Jill and that such a move may thus induce Jill to cooperate in order to take advantage of Jack's apparent "irrationality"! Is it then obviously irrational after all to cooperate for some of the moves of a hundred-repetition prisoners' dilemma? It is hard to tell a convincing story of how a player should work out a suitable strategic response to contingencies that ought not to arise if the other player is rational and well-schooled in game theory.

Ethical applications of game theory are bound to be precarious, because game theory is in turmoil. Not only are there the problems involved in incorporating imperfect knowledge and imperfect rationality, but there is disagreement about what solution concepts are appropriate even in cases of perfect knowledge and rationality. The traditional solution concept for noncooperative game theory is the Nash equilibrium, and most game theorists would hold that it is a reasonable necessary condition for rational play that solutions should be Nash equilibria. But it is not a sufficient condition: not all Nash equilibria are rational solutions.

Consider the game whose normal form is shown in Figure 14.4.1. Player I (Jill) plays Up or Down and Jack then plays Left or Right. Suppose that we have a cardinal representation of their preferences that is proportional to their own dollar payoffs (which are shown in Figure 14.4.2). There are three Nash equilibria: (Up, [Lu, Rd]), (Down, [Lu, Ld]), and (Down, [Ru, Ld]). The strategies [Lu, Ld] (play Left regardless of what Jill does) and [Lu, Rd] (play Left if Jill plays Up and Right otherwise) are each best replies by Jack to Jill's playing Up; but only the latter is a Nash equilibrium, since if Jack

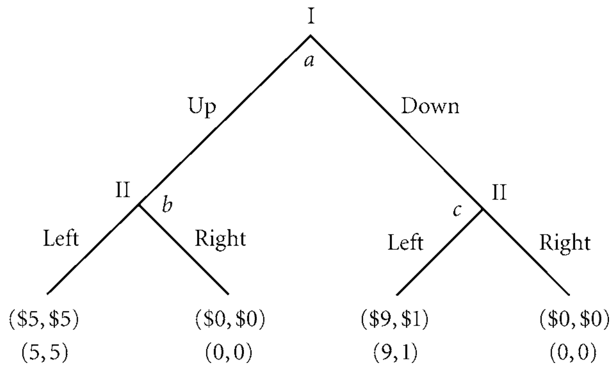

Figure 14.4.2

adopts the strategy of playing Left regardless of what Jill does, then Up is not Jill's best reply. Both [Lu, Ld] and [Ru, Ld] are best replies to Jill playing Down, and Down is a best reply to both these strategies. So there are three pure-strategy Nash equilibria. The normal-form representation makes it easy to recognize the equilibria – just look for cells where the first number is a maximum in its column and the second number is a maximum in its row.

However, as one can see more easily from the extensive-form representation of Figure 14.4.2, one of these Nash equilibria seems to be irrational. If Jack adopts the strategy of playing Left in response to Up and Right in response to Down, then it is rational for Jill to play Up. But Jack's threat to play Right in response to Jill's playing Down is not credible. For once Jill has played Down, Jack has a choice between getting $1 and getting nothing, so if he prefers more money to less then he will play Left after all. Therefore, Jill should play Down. Only the two equilibria involving Jill playing Down are rational. They are the only "subgame perfect" equilibria (Selten 1975): they are not only Nash equilibria in the game as a whole, they are also equilibria in subgames (games that begin with intermediate nodes). The Nash equilibrium strategy pair (Up, [Lu, Rd]) is not subgame perfect because, in the subgame that begins after Jill has already played Down, Jack's best move is not Right. Solution concepts that are stronger than the notion of a Nash equilibrium (such as the notion of a subgame perfect equilibrium) are needed, but there is little agreement on what they should be. These controversies are important to the ethical application of game theory, since they bear on such questions as the rationality of retaliation.

Notice that the extensive-form representation of games is crucial to identifying subgames. Whether or not the two representations are ultimately equivalent (Harper 1991), they do have different virtues. Normal-form representations are more compact and make it easy to identify Nash equilibria. Extensive-form representations make explicit the order of play, the player's knowledge, and the structure of subgames.

The game presented in Figures 14.4.1 and 14.4.2 is of particular interest owing to experimental results that concern so-called ultimatum games. In these experiments, subjects are paired via computer in the laboratory (so that they never actually meet). One member of each pair is told to offer any division of a sum of money – of \$10, for example – to the other player. If the second player accepts the proposed division then both players receive the amounts proposed; if the second player rejects the offer then both players receive nothing. There are many Nash equilibrium: for each \$*X* less than \$10, the strategy pair (Offer \$*X*, Accept \$*X* and no less than \$*X*) is a Nash equilibrium. But the only subgame perfect equilibrium involves offering the smallest amount of money (since even a minimal payment is better than nothing). What happens in the experimental setting is that most of the offers are close to a 50–50 split, and very unequal divisions are almost always rejected. The game shown in Figures 14.4.1 and 14.4.2 is a more restricted version in which only two proposals – an equal division or a \$9–\$1 split – are possible. If Jill naively accepts the theory of subgame perfect equilibria then she will offer a \$9–\$1 split, and if Jack is like most people then she will wind up with nothing. Subjects who play ultimatum games regularly carry out what game theorists regard as empty threats.

In our view, two factors are responsible for the experimental results. First is the fact that people care about many things besides their own financial payoff. People also care about fairness (particularly when they get the short end of the stick), and they are willing to spend a dollar in order to "send a message" to somebody who makes an unfair offer. (If the unequal division is generated by a chance mechanism rather than proposed by the first player, people are much more willing to accept it – see Blount 1995.) Since the second player thus actually prefers the outcome where both get nothing to the outcome where the proposer gets \$9 and the second player gets \$1, the game shown in Figures 14.4.1 and 14.4.2 misspecifies the preferences of the experimental subjects and so does not correctly model their strategic circumstances. The experimental results highlight the enormous difficulties in modeling a strategic interaction as literally a *game*. And until economists have succeeded in specifying what game it is that people are playing, game theory has little to contribute.

Second is the fact that it may be reasonable for people to adopt "rules" or "policies" or to cultivate dispositions that in the real world of repeated encounters prevent them from being identified as or played for a sucker. Although the experimental situation is specifically constructed as a one-shot interaction, people may still behave according to these habits or rules; and given that it may be difficult or costly to distinguish one-shot from repeated environments in normal experience, this behavior may be reasonable.

Even if game theory were not wracked by controversy and paradox, there would be grounds to hesitate before employing it to address ethical questions. First, there are qualms about its heavy reliance on information about the utilities of outcomes. Game theory is not strictly welfarist because it does not rely only on information concerning preferences. For example, nonpreference information determines what the possible strategies are and what their utility consequences will be. In addition, in many game-theoretic applications that aim to explain and predict strategy choices or to advise people on which strategies to choose, questions of evaluation simply do not arise. But one might reasonably say that evaluative applications of game theory are quasi-welfarist, for other ethically relevant factors – rights, capabilities, needs, and reasons – make no explicit appearance, and their role is heavily truncated (Roemer 1986a). For example, if the strategic problem is to distribute some set of commodities, then the solution will be the same regardless of what the commodities are as long as the feasible allocations and utility functions are the same. But people's judgments about just distributions do not depend only on information about preferences or welfare; they depend also on what the goods to be distributed are. As we shall see, the outcomes reached in experimental tests of bargaining theory do not depend on preferences alone. Indeed, Yaari and Bar-Hillel (1984) have shown that people hold different ethical views about how to distribute avocados and grapefruits depending on *what* the fruits are wanted for, not just on how much the fruits are wanted. Many people would insist that justice requires a more egalitarian distribution if the fruits are wanted for their vitamins than if they are wanted for their taste. To apply game theory sensibly, one needs to grasp the complexities of individual motivations and evaluations; and to do this, one must understand people's moral commitments.

A further complication in individual motivation is that people are capable of looking at situations from the perspective of the group. Rather than asking "What is best for me?" people sometimes ask "What is best for *us*?" Game-theoretic analyses of social interactions implicitly rule out the possibility of people adopting such group perspectives. Sympathy, altruism, even "commitment" (see Section 6.5) might still be allowed, but not a perspective that considers what *we* should do and what the consequences will be for *us*. In ruling out such a collective perspective, game-theoretic analyses thus beg some of the deepest questions in social and political philosophy. It has been argued that such a collective perspective is essential to the very notion of social interaction (Gilbert 1990; Tuomela 2000, 2003; Sugden 2003; Bacharach 2005) and also to the explanation for why real individuals play simple games so differently from how game theory would prescribe and predict (Dawes et al. 1990; see also Regan 1980, ch. 8; Hurley 1989, p. 145). What better explanation is there for the fact (see Bicchieri 2002) that experimental subjects are much more likely to cooperate when they have an opportunity to talk with one another, even on irrelevant subjects? In our view, the application of game theory to moral philosophy remains controversial, and the suitability of its assumptions must be addressed explicitly in each application.

14.5 Bargaining Theory and the Social Contract

In order to clarify the difficulties involved in applying game theory to moral philosophy, let us focus briefly on some game theory that has actually been applied to ethical issues: two-person cooperative bargaining theory. In John Nash's classic analysis (1950), the two parties are supposed to have cardinal utilities that are not interpersonally comparable. The bargaining problem is to select a distribution of utility, u^* to the first party and v^* to the second party, from a closed convex set S of possible utility outcomes. (A set is closed if it includes its boundaries, and it is convex if, for any two points x, y in the set, it includes every point on the straight line segment between x and y.) The set S of possible outcomes contains a "threat point," $d = (u_d, v_d)$, consisting of the utilities the two individuals receive if no bargain is reached. Rational individuals will, of course, not settle for less than they can get if no bargain is made. So $u^* \geq u_d$ and $v^* \geq v_d$. Nash proved that if the solution to the bargaining problem satisfies four further explicit conditions, then the solution maximizes the product of the utility gains from bargaining. That is, (u^*, v^*) is the Nash solution when $(u^* - u_d)(v^* - v_d) > (u - u_d)(v - v_d)$ for all points (u, v) in S other than (u^*, v^*) itself. Nash's four conditions are as follows.

1. (Pareto Optimality) For all points (u, v) in S, $u^* \geq u$ and $v^* \geq v$.
2. (Invariance) Suppose (u^*, v^*) is the solution to the original bargaining problem in which the agents have utility functions U and V and that the utility functions are transformed to U' and V', where $U' = aU + b$ and $V' = cV + d$ (here a, b, c, d are real numbers and a and c are positive). Then $(au^* + b, cv^* + d)$ is the solution to the transformed bargaining problem. This means that positive affine transformations of the utility functions are irrelevant to the solution in material terms.

3. (Symmetry) If $u_d = v_d$ and if (v, u) is in S whenever (u, v) is in S, then $u^* = v^*$. If the agents are in symmetrical positions then the bargaining solution must also be symmetrical.
4. (Independence of Irrelevant Alternatives) Suppose that S' is a subset of S and that both the threat point (u_d, v_d) and the solution (u^*, v^*) to the bargaining over the larger set S are also in S'. Then the solution to the bargaining game restricted to S' should also be (u^*, v^*). (Notice that this is quite different from Arrow's independence of irrelevant alternatives axiom **I**, discussed in Chapter 13.)

These conditions seem plausible, and the Nash solution has many defenders. Yet it has awkward consequences, for if the feasible set of utility outcomes expands in person *A*'s favor then the result may be that *A* actually receives less than before (Kalai and Smorodinsky 1975, p. 515). The Nash solution may also be morally objectionable in the extent to which it disadvantages both the poor (whose u_d will be low) and the risk averse (whose utilities increase more slowly as a function of their physical shares).

A further problem is that, in empirical studies, the bargains that individuals reach diverge systematically from the Nash solution even when both individuals have perfect knowledge of their utility functions. In "binary lottery" games, individuals bargain over the distribution of lottery tickets that give them a chance of winning separate prizes. Since the bargaining solution does not depend on interpersonal comparisons, one can stipulate that each individual's utility is zero when she has no tickets – that is, no probability of winning – and that the utility of having all the tickets – that is, of winning her prize for sure – is 1. The utility for each player of having X% of the lottery tickets will then be X times 1 (the utility of getting the prize) plus $(1 - X)$ times zero or simply X, the percentage of lottery tickets the player has. The Nash solution is a 50–50 split of the lottery tickets.

The Nash solution follows regardless of what prizes the individuals can win. If both *A* and *B* will win \$5 in their separate lotteries then the Nash solution, a 50–50 split of the lottery tickets, seems reasonable. But if *A*'s lottery pays off \$5 while *B*'s lottery pays off \$20, then the Nash solution of a 50–50 split of the lottery tickets now seems less appealing. When the prizes differ, the actual experimental outcomes of the bargaining games are "bimodal," with both the Nash solution and the solution that equalizes monetary returns as the two modes (Roth and Malouf 1979). Roth, Malouf, and Murnighan (1981) argue persuasively that *moral norms* concerning just distributions affect the bargaining solutions. The person who stands to win the larger prize is seen as having a weaker moral claim on lottery tickets than the less advantaged person. Yet experimental work by Binmore et al. (1993) suggests that people's moral views are fragile and depend on their experience in bargaining. It seems that even purely positive bargaining theory must take cognizance of the moral commitments of the bargainers, and it is by no means obvious how this is to be accomplished formally (Gibbard 1990, esp. p. 262; Pettit 1990).

An alternative to the Nash solution developed by Kalai and Smorodinsky (1975) forms the basis from which Gauthier (1986) constructs his contractarian derivation of justice from rational choice. Kalai and Smorodinsky conceive of the bargaining problem much as Nash does, and they accept Nash's first three assumptions. But in place of Nash's fourth condition, his independence of irrelevant alternatives constraint, Kalai and Smorodinsky impose a monotonicity requirement. This condition states that, if the maximum utility available to one of the parties increases, then the utility that party gets in the bargaining solution must not decrease.

With the monotonicity condition in place of independence of irrelevant alternatives, Kalai and Smorodinsky derive for the two-person case an "equal relative concession" solution. Recall that the relative concession of each bargainer is $(u_{\max} - u)/(u_{\max} - u_d)$, where $u_{\max}$ is the largest feasible utility attainable by the individual, u is the utility for the agent if a given bargain were made, and u_d is the utility at the threat point. Because the equal relative concession solution need not be Pareto optimal when there are more than two bargainers, Gauthier substitutes "minimax relative concession" – the solution that minimizes the largest concession anybody has to make. When equal relative concession is Pareto optimal, it coincides with minimax relative concession. Gauthier (1986, ch. 5) argues that rational individuals who are aware of their own rationality and the rationality of others and who are aware of the utility consequences of their bargaining will accept the minimax relative concession solution. In Gauthier's view this argument suffices to establish that such a solution is just, although he also maintains (ch. 8) that this solution satisfies conditions that have been thought reasonable to impose on conceptions of justice, such as impartiality. Gauthier argues further (ch. 6) that it is individually rational to adhere to such bargains.

Although this extension of the Kalai and Smorodinsky solution avoids some awkward implications of the Nash solution, it still disadvantages the poor and the risk averse, and it is equally inconsistent with the experimental evidence concerning bargaining over shares of lottery tickets. Moreover, the Kalai–Smordinsky solution has the following disquieting consequence. Suppose that the solution to a particular three-person bargaining problem is for the first person to receive half the good while the second and third persons each receive a quarter of the good. If the third person were guaranteed his one-quarter share and the first two were to bargain over the remaining three quarters, one would suppose that they would agree to a division whereby the first gets two thirds (half the original) and the second gets one third (a quarter of the original). A bargaining rule with this property is called "consistent." The Kalai–Smordinsky solution is not consistent. Lensberg has proven that the Nash solution is the only bargaining rule that is efficient, impartial, scale invariant, and in this way consistent (see Young 1994, p. 122).

Gauthier's bargaining theory argument for his principles of justice has been strongly contested, and its difficulties highlight the complex interconnections between game theory and moral philosophy. The characterization of individual rationality that Gauthier relies on is controversial, the relevance of solutions to a two-person bargaining problem under conditions of complete information is questionable, the particular solution concept is dubious, and the limitation to quasi-welfarist information is objectionable. This list of difficulties is not intended as a criticism of Gauthier's trailblazing efforts or as an argument against employing game theory to tackle moral problems. Indeed, it is only through Gauthier's efforts that these dimensions of a contractualist argument for principles of justice could be clearly seen.

This example illustrates the present relations between game theory and moral philosophy. Game theory cannot now solve serious moral problems because there are difficulties with its notions of rationality, its solution concepts, and its strong knowledge assumptions. Even so, game theory has provided a valuable and influential conceptual framework in which to think about moral problems; as game theory continues to develop, it may yet help to solve them, too.

Suggestions for Further Reading

There are many introductions to game theory. Luce and Raiffa (1957) is a classic, and it is still very helpful. Harsanyi (1977b) is also accessible, though it defends some controversial positions. Good texts are Binmore (1992) and Osborne (2003). Most authors accept the view that solutions to noncooperative games should be Nash equilibria, but some (such as Bonnano 1991) disagree.

The prisoners' dilemma has been discussed in hundreds of books and essays. Axelrod (1984) and Parfit (1984, ch. 2) are particularly helpful discussions. Frohlich et al. (1975), Hardin (1982), Taylor and Ward (1982), and

Hampton (1987) all question whether problems of collective goods provision are always best understood as prisoners' dilemmas. For Sen's discussion of the assurance game see Sen (1967, pp. 112–24).

There are other kinds of mixed coordination/conflict games than those described here. For excellent discussions of how coordination can arise in various circumstances, particularly when there are asymmetries in the positions of different players, see Schotter (1981) and Sugden (1986). For a pathbreaking account of how people sometimes manage to solve apparently insoluable coordination problems, see Schelling (1960). For an attempt to provide a deeper theory of how coordination can be achieved, see Vanderschraaf (2001).

For discussions of the backwards induction argument and its difficulties, see Kreps et al. (1982); Kreps and Wilson (1982); Binmore (1987, 1988); Bicchieri (1988, 1990); Pettit and Sugden (1989); and Bonnano (1991).

An excellent account of how game theory bears on multifarious equity questions is Young (1994), and a wonderful introduction to bargaining theory and its applications to moral philosophy is Barry (1989). Other useful discussions of bargaining can be found in Nash's own short paper (1950); Braithwaite (1955); Luce and Raiffa (1957, pp. 124–37); and Thomson and Lensberg (1989). Binmore (1994, 1998) is an important application of game theory to moral philosophy.

Critical discussion of Gauthier's theory can be found in Roemer (1986a); Kraus and Coleman (1987); Nelson (1988); Paul et al. (1988); and Sugden (1990). See also Gauthier's collection of essays (1990). Skyrms (1996, 2003) offers excellent applications of evolutionary game theory to social norms.

CONCLUSIONS

We opened this volume with four examples of work in economics that reflects ethical views. If this book has served its purpose as an analysis of the ethical presuppositions of economics and as an introduction to a range of ethical concepts and theories, it should make the ethical commitments implicit in these cases easier to understand, and it should suggest how they might be influenced by a broader and more self-conscious ethical perspective. We have also – especially in the last two chapters – said a bit about tools that economists have developed that may be of use to moral philosophers, and we wish also to say something about ways in which economists and philosophers can together help to address some of the daunting problems facing the nations of the world.

Accordingly, in the concluding chapters we shall revisit the four cases discussed in Chapters 2 and 3 and offer some remarks about ways in which grasping both ethics and economics might help in identifying good policies and principles for citizens and governments to adopt. In Chapter 15 we shall return to the two examples of normative economics discussed in Chapter 2 – the defense of school vouchers and Larry Summers's proposal that the World Bank encourage polluting industries to locate in poor countries. In particular, Chapter 15 will show how the discussions in Parts II and III of welfare and of nonwelfarist evaluative considerations help one to understand and appraise these arguments. In Chapter 16, after a brief discussion of the ways in which a better knowledge of ethics can help clarify the controversies concerning involuntary unemployment or Samuelson's defense of a biological rate of interest in his overlapping generations model, we shall say a few words about how the tools of economics and the concerns of moral philosophers can be united to address pressing social problems.

FIFTEEN

Pollution Transfers and School Vouchers: Normative Economics Reconsidered

Both Lawrence Summers's argument that the World Bank should facilitate the transfer of polluting industries to poor countries and the efficiency argument for school vouchers are instances of standard welfare economics. If parents are permitted to choose among competing schools then their preferences will be satisfied at least as well as in a state-sponsored system, and competition among schools will induce cost savings and improvements in quality. Concerns about the distributional consequences can be addressed by a system of vouchers, which can be as generous or stingy toward those who are poor as the citizenry wishes. Similarly, the massive inequalities in incomes, wages, and environmental quality between rich and poor countries and the fact that pollution damage is not a marketable good create opportunities for Pareto improving trades (or industry relocations) between rich and poor countries, which the World Bank can facilitate. Uncompensated transfers of pollution to LDCs (like the simple elimination of public support of education) cannot be justified in this way, since doing so would not be beneficial to all parties. Yet even uncompensated pollution transfers would apparently represent potential Pareto improvements and would be viewed favorably from the perspective of cost–benefit analysis.

We went on to suggest a number of objections to these arguments. In addition to distributive concerns, voucher plans must be responsive to the paternalistic and political reasons why governments support education. In Chapter 2 we suggested that, although these considerations do not necessarily defeat the argument in support of a voucher system, they do support constraints on its structure. More importantly, we argued that the economic arguments in favor of vouchers ignore crucial values that are not matters of either equity or efficiency in the satisfaction of preferences – values such as social solidarity and a cultivated citizenry.

With respect to the Summers memorandum, we mentioned five objections. The first and third do not challenge the normative framework that Summers employs. Relocating polluting industries to poor countries might increase the total amount of pollution in the world so much that, even if the distribution were more efficient, the overall effect might be harmful. Considering only one alternative to the status quo risks overlooking still better possibilities. The other three objections aimed more directly at the moral basis of Summers's argument and of welfare economics as a whole. The second objection was that the gross disparity in bargaining position of citizens of rich and poor countries makes the exchange in question unfair, whether or not it is mutually beneficial. The fouth objection concerned doubts about the measure of welfare on which Summers's argument relies, and the fifth objection questioned the reliance on market measures of value. As we pointed out in Chapter 9, welfare economists do not claim that efficiency in satisfying preferences is the only relevant consideration, and they explicitly concede that considerations of equity may sometimes limit the pursuit of efficiency. So welfare economists could grant the second objection, though not of course the fourth and fifth. But in focusing exclusively on efficiency, welfare economists often wind up exaggerating its importance as compared to alternative moral considerations; and the case for school vouchers and the Summers memorandum both illustrate this risk.

The important objections to both vouchers and pollution transfers point toward alternative frameworks for moral evaluation. Concerns about fairness might profitably be considered from the standpoint of social contract theories, the subject of Chapter 12. Furthermore, one might think through the issues raised by vouchers or exporting pollution from the perspective of freedom and rights (presented in Chapter 10) and from that of equality (discussed in Chapter 11). Different views of welfare also bear on the assessment of proposals for school vouchers or pollution transfers. One might approach the question of whether a voucher system should replace government provision of education – or whether polluting industries should be shifted to poor countries – from the standpoint of utilitarianism, which permits the interpersonal comparisons of welfare that are ruled out in the Pareto efficiency framework, as well as from the standpoint of "objective" measures of welfare considered as alternatives to the preference satisfaction view of welfare which is implicit in standard efficiency arguments. Although a careful treatment of the issues from any of these moral perspectives is not possible here, some rough sketches follow.

15.1 Do Vouchers and Pollution Transfers Make People Better-Off?

From a standard efficiency perspective, the question concerning vouchers is whether government provision or a voucher system makes people better-off, where the standard of well-being is the satisfaction of preferences. The question is: Which satisfies preferences better? But whose preferences are we talking about? Is the point of schooling to satisfy the preferences of parents, the preferences of students, the preferences of both, or the preferences of the citizenry as a whole? Nobody thinks that the point of schooling is principally to make parents better-off: schools serve children, and the extent to which they benefit or harm students is central to their assessment. On the other hand, the point of schooling is obviously not only or mainly to satisfy the existing preferences of students. If the preferences students will later have as adults were given, independent of schooling, then one might consider to what extent schooling increases the capacity of students over their lifetime to satisfy these given preferences. But those preferences are not given and, on most accounts (including Milton Friedman's), one crucial objective of schooling is to *shape* the preferences of children: to instill "values" and to lead young people toward healthy and increasingly sophisticated tastes. Assessing policies that change preferences is a difficult task within a preference satisfaction view of welfare. Moreover, schools also have effects on those who are not parents and who are no longer students – as, for example, if schooling helps create citizens who are more law abiding.

Once one recognizes that efficiency in economics is efficiency in promoting welfare and that economists take welfare to be the satisfaction of preferences, one can then see that the efficiency argument in defense of vouchers is not as obvious as it may appear. Since those choosing among schools are only some of those whose preferences the schools are attempting to satisfy, one cannot conclude that a system that is responsive to their choices will best satisfy all the relevant preferences. Indeed, the whole preference satisfaction framework seems inappropriate. If, as the identification of welfare with the satisfaction of preferences implies, people's preferences are directed toward their own well-being, then it is inconsistent to assume that the preferences of parents are directed toward the well-being of their children (unless the well-being of parents and children happens to coincide). More fundamentally, one cannot understand the extent to which a system of schooling promotes the well-being of students if one takes well-being to be the satisfaction of preferences. Which preferences schools nurture may be more important than how well they satisfy the preferences students may have at any moment in time.

Summers similarly reduces the question of whether LDCs are "underpolluted" to the question of whether shifting more pollution to the LDCs better satisfies people's preferences. Cost–benefit analysis measures preference satisfaction in terms of willingness to pay or to accept compensation. This method is problematic, since the extent to which people prefer to avoid pollutants depends on their wealth and on their (possibly mistaken) beliefs concerning the consequences of the pollutants. Summers's memorandum avoids these problems, since he focuses on the actual effects of additional pollution and since the revenues polluting industries would provide are supposed to compensate those who may be harmed by the relocation of polluting industries. Regardless of what people in LDCs believe about the effects of pollution, its welfare costs will, Summers argues, in fact be less there.

We mentioned in Chapter 2 that Summers, unlike many welfare economists, measures all costs and benefits – including injuries and deaths – in terms of willingness to pay. In addition, he takes for granted the status quo distribution of wealth. Thus, a permanent injury to a child lowers the welfare of the child's family much more if the family happens to live in a rich country, because the cost of the child's care, the extent to which the child's prospective earnings are diminished, and the amount that the family would be willing to pay to avoid the injury are all much greater in rich countries than in poor.

Measuring preference satisfaction this way seems morally unacceptable. It is just as morally important to avoid crippling a poor child as to avoid crippling a rich one. We shall later note how a careful formulation of Summers's argument for compensated pollution transfers may avoid wealth bias in the measure of preference satisfaction. Furthermore, it is possible and indeed common to impute single wealth-independent values to injury and death. Yet doing so creates conceptual strains. If dying or losing a limb is just as bad whether one is rich or poor, then isn't everything that influences health or longevity equally valuable regardless of one's wealth? And since almost everything influences health and longevity, it is incongruous to impute values to injury and death that are independent of an individual's willingness to pay (or to accept compensation) while still taking the values of other things to depend on prices.

Given the difficulties of applying a subjective preference satisfaction view of welfare to the cases of vouchers and pollution transfers, there may be advantages in appealing instead to objective conceptions of welfare such as

Sen's and Nussbaum's capabilities or Rawls's primary goods. In the case of schooling, thinking in terms of capabilities is very natural because schooling is relevant to so many central human capabilities. Rawls does not list education as a primary good, but clearly it is relevant to equal opportunity and to virtually all the primary goods on his list. In order to simplify the task of thinking about pollution in terms of capabilities and primary goods, let us focus on the negative health consequences of pollution. The capabilities affected by pollution would include the capability to breathe freely (or to continue breathing at all!), the ability to see well, and so on. Although, as we noted, Rawls does not include health among the primary goods on the grounds that it is a natural rather than a social good, many have argued that health is heavily and increasingly influenced by social factors and that health should be included among the primary goods. We agree. Pollution would then, like schooling, have direct consequences on primary goods.

In the case of some pollutants, such as those that increase the risk of prostate cancer (a disease that affects mainly elderly men), a given exposure will have a relatively minor effect on the capabilities or primary goods of people in LDCs, who are likely to die young from other causes. A given dose of a particular pollutant may also have fewer negative health consequences if there is very little pollution than if there is already a great deal. On the other hand, the interaction between pollution effects, the generally worse health status of people in poor countries, and deficiencies in the systems for detecting and treating environmentally induced illnesses might render some of these effects more serious. Similarly, the numbers of people affected by pollution might sometimes be higher and sometimes lower in poor countries. There is no good reason to believe that *uncompensated* transfers of pollution to poor countries would result in a net increase in primary goods or capabilities in the world.

Analysis of the consequences for capabilities or primary goods of *compensated* transfers is more complicated. The Pareto efficiency argument in favor of compensated transfers doesn't require any comparison of the pollution's welfare effects in rich and poor countries, and in this sense it avoids the direct wealth bias noted earlier in the argument for uncompensated transfers. The argument relies instead on the fact that people in poor countries are willing to give up environmental quality in exchange for items like food and basic health care on terms that make the transfer of polluting industries from rich countries attractive. These trade-offs appear beneficial in terms of capabilities and primary goods, not merely in terms of preference satisfaction. As we pointed out in Section 9.4, however, even an actual Pareto improvement may depend on the distribution of income and could evaporate if the distribution of income changed. We will have more to say about this in the next section.

There is yet another problem here that we did not mention in Chapter 2. The idea of compensating a *country* is a cheat: transferring pollution is a Pareto improvement only if the preferences of every individual are at least as well satisfied after the transfer as before. Even on its own terms the argument may be in trouble here, because the compensation may fail to reach the individuals who are harmed by the pollution. In this respect, objective standards of well-being may have a practical advantage. In terms of an objective and interpersonally comparable standard of welfare, a country that accepts a polluting industry in exchange for public health measures that serve the country's poor will experience, on balance, a gain in welfare – without requiring that every individual harmed by the pollution be compensated. However, important issues of fairness would arise if different people benefited from the public health measures than were harmed by the pollution.

The frameworks of capabilities and primary goods seem particularly useful in thinking about the provision of elementary education. Basic literacy and numeracy are central capabilities in modern societies and certainly can be construed as the kind of "universal means" that Rawls considers primary goods to be. As one moves beyond primary education to high schools and especially to colleges, the capabilities and the goods involved are diverse and their relative values harder to weigh. At the elementary school level, it seems clear that the most important defect in American schooling is the failure of large numbers of students to achieve levels of literacy and numeracy that allow them to function effectively in high school and in later life, both as economic agents and as citizens.

Whether a voucher system would reduce the incidence or severity of such failures compared to a public school system (other things being equal) is a difficult question to judge. Would parental choice among schools lead schools to compete to nurture students' capabilities? Would such a system break the "tyranny of low expectations" that permits students from impoverished homes to pass through school without learning much? Would the lessened influence of teachers make primary education less conceptually, emotionally, or cognitively demanding? Or would a voucher system produce larger gaps than now exist in the development of basic capabilities between those whose parents are attentive and those whose parents are not? These are hard questions, but there may be a better prospect of approaching answers to them through the analysis of relatively objective capabilities like reading and calculating than through the one-dimensional analysis of preference satisfaction.

Objective welfare criteria aim to measure central human concerns, which should also show up in subjective evaluations if people are well informed and largely self-interested. So there is some presumption that Pareto optimal policies are genuinely welfare promoting. Still, the presumption is clearly rebuttable, and in the case of education it is dubious. Whether it is plausible to maintain that Pareto improving pollution transfers are genuinely welfare enhancing is also questionable, since circumstances and preferences are changing rapidly and since so many of those affected are children.

15.2 A Utilitarian Perspective on Pollution Transfers

A central difference between utilitarianism and standard welfare economics is that utilitarians are willing to make interpersonal comparisons. What difference does this fact make to Summers's analysis of pollution transfers? For simplicity of comparison, assume that the utilitarian agrees that welfare is the satisfaction of preferences. Granting Summers's empirical assumptions, a compensated transfer of polluting industries to poor countries would enable the citizens of all countries to satisfy their preferences better and thus would increase utility overall. In this sense, utilitarians who endorsed Summers's empirical assumptions and accepted his view of welfare would agree that compensated transfers of polluting industries to poor countries are an improvement on the status quo.

However, it does not follow that a preference utilitarian would support the pollution exchange policy, for there are more than two alternatives. In addition to the status quo and to compensated transfers of pollution, consider the alternative discussed in Section 9.4 of simply redistributing wealth. On plausible utilitarian assumptions, the marginal utility of income is much lower in rich than in poor countries. Transferring income or resources from the rich to the poor will, then, be welfare enhancing in utilitarian terms, for the gains in utility to persons in poor countries will outweigh the losses to those in the rich countries. (Utilitarians recognize that one needs to consider long-run consequences and that technical and educational aid are often better than cash, but these complications are not germane here.) Are such one-way transfers of resources from rich to poor countries more or less effective in promoting world welfare than are exchanges like the one Summers contemplates?

Imagine breaking the compensated transfer of pollution into two parts. Part 1 is the transfer of compensation (jobs, medical facilities, etc.) from rich to poor countries; Part 2 is the transfer of pollution toward poor countries (or the transfer of environmental quality from poor to rich countries). Part 1 is clearly, on standard utilitarian assumptions, welfare enhancing for the reasons just discussed. However, there is no general reason to believe that Part 2, the movement of pollution toward poor countries, is welfare enhancing. In rich countries, clean air is worth more in dollar terms, but is it worth more in interpersonally comparable utility terms?

We can take this a step further in a way that clarifies the relationship between preference utilitarian and Paretian views. If it were still possible to satisfy somebody's preferences better without lessening the preference satisfaction of someone else, then utility would not yet have been maximized. So the best utilitarian outcome will be Pareto optimal. In addition, a Pareto improvement increases total utility. But it does not follow that any particular Pareto improvement is a good way to increase total happiness. Given the present disparities in resources between rich and poor countries, the kinds of Pareto improving transfers between rich and poor countries contemplated by Summers would in fact produce small gains in total utility compared to the large gains that well-designed wealth transfers from rich to poor countries would provide. In addition, if inequalities in standards of living were greatly reduced, it would no longer be obvious that movements of pollution toward LDCs would be welfare enhancing. For then the principal force pushing in that direction – namely, the very low incomes and wages in poor countries – would no longer be operating.

A parallel argument can be made in the case of vouchers. In the United States, the most urgent problems and greatest failures of education are found among poor children. Indeed, a frequently heard argument in favor of vouchers is that they provide the children of poor parents a "way out" of failed public schools. But the most obvious problem that needs addressing is that educational failure itself, which perpetuates the inequality. Transfers from the rich to the poor (providing the efficiency losses are not too large) move resources from people for whom they provide little utility to people for whom they provide much utility. Redistributing income and educational resources is probably a substantially more important matter from a utilitarian point of view than is the question of vouchers versus public schools. It is not entirely clear how the comparison of public schools and a voucher system would be affected by a considerable reduction in material inequality. Under more egalitarian conditions, there might be few failing public schools and little interest in vouchers. On the other hand, the concern that vouchers might disproportionately benefit affluent families would be eased by greater equality in background conditions.

15.3 Other Ways of Evaluating Vouchers and Pollution Transfers

We noted in Chapter 2 that some would question whether Summers's proposal is fair, since it apparently permits rich countries to take advantage of the vulnerabilities of poor countries, and we considered nonwelfarist concerns about vouchers, too. But we did not raise explicitly questions concerning rights, liberty, or equality.

15.3.1 Rights, Freedoms, Pollution, and Vouchers

One might think that there would be no freedom- or rights-based objection to compensated pollution transfers. After all, Summers argues that the transfers are mutually beneficial and would be agreed to voluntarily by both parties if there were no barriers to such exchanges. But this way of thinking is mistaken. There is a big difference between, on the one hand, George trading his watch to Mary in exchange for her briefcase and, on the other hand, some third party taking George's watch and Mary's briefcase and then giving the watch to Mary and the briefcase to George. What is at stake in the Summers memorandum are not voluntary agreements among individuals in different countries to accept pollutants for compensation, but rather efforts by governments and the World Bank to enforce transfers that would approximate the results of voluntary exchanges. As Nozick stresses, the libertarian is concerned with processes rather than outcomes. One does not respect freedom by forcing people to do what they would have chosen to do, and one does not respect rights by forcing on people risks that they would voluntarily have chosen to impose upon themselves. Since vouchers allow for actual choices, they are in this regard more responsive to concerns about freedom than are proposals for transferring pollution.

Furthermore, compensated transfers of pollution across nations need not be Pareto improvements across individuals, because those who incur the costs of the additional pollution will not necessarily receive the compensation. So those who value rights and freedom most of all will have serious qualms about Summers's proposal. Libertarians like Nozick are prepared to reject the entire welfare-improving perspective of welfare economics and to restrict policies to the protection of individual rights. Those who value both freedom and welfare and who have a more expansive view of rights would take a more nuanced view. To pursue the issue further from this perspective raises complicated problems (that we cannot address here) concerning what rights people have to a clean and low-risk environment.

At first glance, it might seem easy to make a libertarian case for vouchers as opposed to government provision, since libertarians are eager to limit the size and power of government. But it is not so simple to apply ideas of freedom and rights to the voucher issue. A vital question is what kind of rights parents hold over their children and what rights children have (Brighouse 2000). If parents have very strong rights that amount to something close to the ownership of their children, then parents should have the sort of control over their children's education that a voucher system permits and that public provision of schooling tends to frustrate. If instead one views the rights of parents as more contingent and limited – as essentially instrumental to promoting children's welfare rather than having some independent ground – then the case for or against vouchers will turn on the more contingent factors we have discussed elsewhere.

One such factor may be children's (in contrast to parents') freedom. In some interpretations of liberalism as a political philosophy, an essential aspect of a good life is that one is able to reflect on one's values and commitments and to endorse them for oneself. This essential feature of a good life requires a social framework in which important freedoms are guaranteed and in which information and active criticism of alternative points of view are encouraged. This kind of liberal view may severely constrain the kinds of education that parents may choose for their children on the grounds that certain types of education (e.g., fundamentalist religious education) discourage rational criticism of favored beliefs and judgments. Liberalism of this variety might impose substantial limitations on the kinds of educational programs for which vouchers could be used without necessarily ruling out the use of vouchers altogether. Thus, libertarians might actually discourage parental choice of schooling on grounds of children's freedom.

15.3.2 Equality, Pollution, and Vouchers

Consider briefly how issues of equality bear on vouchers and pollution transfers. The central point with respect to the Summers memorandum is that the proposed transfer takes place against a background of extreme inequality. Uncompensated transfer of pollution to poor countries, which might be recommended on cost–benefit grounds, would worsen these inequalities. Whether compensated transfers would increase or diminish inequalities is unclear. The crucial point is that efficiency-oriented policy analysis takes as given the status quo distribution of resources. Exactly the same point holds with respect to private purchases of education, though most people endorse some notion of equal opportunity in the case of education.

Those committed to egalitarian values such as equal respect, solidarity, and fairness will reject Summers's proposal and his arguments. Showing equal respect precludes assigning greater values to the lives, health, and comfort of those who are richer. Solidarity points toward protecting those who are least well-off and toward sharing the burdens of technological advances rather than taking advantage of the poverty of the LDCs. And we have already questioned whether Summers's proposal was fair.

The relations between voucher proposals and an egalitarian perspective are complex. Schooling is a key instrument for determining social and economic status in modern societies. Broadened educational opportunity is therefore a key strategy for promoting more equal opportunity generally. At the same time, unequal "starting conditions" among people from different social classes create large inequalities, which schooling alone cannot overcome. There is evidence, for example, that as much as half of the observed differences across social classes in student performance on standardized tests is present by the time students begin elementary school (Rothstein 2004).

Strong egalitarians such as John Rawls take as their ideal a state of affairs in which life prospects of individuals are independent of the social and economic status of their parents. Equalizing schooling is just one of many reforms that are needed, and egalitarians might well support well-designed voucher plans that expanded the opportunities of poor children. The crucial question would be whether voucher systems do so more effectively than improvements in public education. Though concerns about equality ground objections to the details of particular voucher schemes, they do not determine the case for or against vouchers.

15.3.3 Justice, Pollution, and Vouchers

In order to narrow the discussion, we shall focus on contractualist theories of justice. Some contractarian theorists, notably including Rawls, have been cautious about extending their theories to international settings, and the problems in working out a contractarian view in those settings are formidable (see Beitz 1985; Rawls 2001b). However, it is possible to examine cautiously the issues at stake from the perspectives of impartiality and mutual advantage that underlie different contractarian views.

Consider some alternative U.S. perspectives on the issue of what constitutes adequate funding of schooling (an issue that arises whether the funding is through vouchers or through direct provision). Many state constitutions in the United States guarantee every child a right to an "adequate" education, and in some states the courts have become heavily involved in enforcing this right. Interpreting and enforcing this right have been complicated by the fact that in most states education is locally controlled and locally funded mainly by property taxes. Hence affluent communities can afford to provide more expensive education than poor communities.

In this context, courts and legislatures have faced the problem of deciding how much funding is required for an adequate education. A libertarian might argue that the state's responsibility is limited to enforcing the duties families have to educate their children and providing minimal education for destitute families. Call this the *minimum* principle. A more ambitious requirement might be that the state provide an education that is adequate to prepare students for satisfactory performance in their roles as citizens and economic actors. Call this the *adequacy* principle. Going further, a much more egalitarian system might be imposed which would likely require moving away altogether from using local property taxes to finance education. Call this the *equal funding* principle. (A difficulty in judging the worth of any of these principles is evidence that the correlation between school spending and student learning is low.)

One might try to decide which of these interpretations might reasonably be rejected by impartial contractarians, who seek a fair agreement or a ranking of alternatives from behind a veil of ignorance. It seems likely that the minimal principle would be rejected: from behind the veil of ignorance, the risk of being fated to receive just a minimal education, with meager life prospects, simply seems too high. It seems plausible that persons facing that choice might instead opt for the adequacy principle, with deliberation focusing on how high the requirement for adequacy should be set. Presumably, in this context an "adequate" education is one that opens up for people a reasonable range of good lives within the society in question. (Thus the requirements for an adequate education are higher in an "information society" than in a traditional agrarian society.) To opt for the equal funding principle might seem an unnecessary intrusion on families' freedom, assuming that the adequacy bar is set high enough.

Similarly, one can list a variety of principles that might govern pollution transfers and assess their justice. A principle that would justify Summers's proposal is this: locate polluting industries so as to minimize economic costs. Let us call this the *cost-minimizing* principle. What are some plausible alternatives? A second might be: locate polluting industries where they will produce the least human suffering. Let us call this the *pain-minimizing* principle. A third might be: locate polluting industries so that those who derive the largest benefits from the industries endure most of the pollution costs. Let us call this the *pay-your-way* principle.

These principles are too crude to underpin real policies, and they are intended only to suggest a range of alternatives whose justice can be considered. Not knowing whether one were a citizen of a rich or poor country, one could reasonably reject the cost-minimizing principle from behind a veil of ignorance. Choosing between the pain-minimizing and the pay-your-way principles is less simple. If it turns out that additional pollution causes relatively less pain when it is concentrated in LDCs, then advocates of the pain-minimizing principle face the tough problem of explaining why it is reasonable for people in poor countries to endure the costs of industries that mainly benefit those in rich countries. The pay-your-way principle, on the other hand, provides some assurance that costs cannot be inflicted by one group upon another. It also has attractive incentive features. (It resembles in this regard the semi-facetious proposal of those concerned about the safety of workers in chemical plants to circulate air from the shop floor through corporate offices.)

One cannot, however, assess principles governing schooling or pollution transfer in isolation from other principles. One cannot suppose that the determination of fundamental rights and duties is up for grabs, as it is in deciding on the principles of justice to govern basic social institutions. On the contrary, one needs a prior specification of rights and general moral principles. One needs to know, for example, what rights parents and children have or what rights individuals have to clean air and water and how property rights are defined. The reasonableness of a particular policy will depend on the whole fabric of the relationship between the parties. From an impartial perspective, one would want to work out a whole scheme of distributive and allocative principles governing opportunities and families or governing relations between rich and poor countries.

Mutual benefit contract theories might come closer to endorsing something like the Summers proposal or proposals for distributing education that fall short of the adequacy principle. Theories in this vein constrain principles of justice by the requirement that they be in the rational self-interest of all parties from the standpoint of their situations prior to agreement, and such theories allow inequalities in the pre-agreement situation to influence the outcome of the rational bargain that yields principles of justice. Whether one should accept the large existing inequalities between families or between rich and poor countries as pre-agreement data for working out principles to govern education or international relations is a central question for those who adopt the mutual benefit approach. If one goes back a few hundred years, inequalities among families were in many ways worse and international inequalities were much less, but neither that historical status quo nor the interactions in the intervening years within societies or between rich and poor countries meet any plausible standard of mutual benefit.

Suppose, though, that we put qualms about starting points aside and ask simply what kinds of bargains might be just – concerning either vouchers or transfers of pollution – given the huge existing disparities in incomes and living standards. If just agreements are defined by the rational self-interest of all concerned, then it may be that compensated pollution transfers, like a wide range of voucher systems, would be just. But even on this view there is one important caveat, which involves the distribution of costs and compensation within poor countries. One segment of the population of a poor country might benefit from a deal to import polluting industries while other segments suffer. In order for the mutual benefit standard to be upheld, compensation would have to extend to those within the country who are losers in the deal and not simply to the country as a whole. In the same vein, acceptable voucher schemes would need regulations to guard against the possibility of serious harm to children from unsuitably or badly chosen schools.

15.4 Conclusions

We have not attempted in this chapter (or in the book as a whole) to argue for any specific policy conclusions, and our discussions of alternative construals of welfare and the relevance of nonwelfarist evaluative considerations do not determine whether, all things considered, it is right or wrong to encourage the migration of a polluting industry to some less developed country or whether a particular voucher system would be an improvement over current systems of government provision of schooling.

As it happens, our view is that such pollution transfers would be very bad policy. We believe that serious worldwide environmental problems would be aggravated if pollution could be more readily exported. For LDCs have fewer resources for monitoring and controlling pollutants, and those who benefit from polluting industries would no longer have to bear as many of the costs. Exporting pollution would be morally objectionable because of these bad consequences, because it is unfair, and because it callously fails to show equal concern for those who happen to be born in poor rather than rich countries. Furthermore, because of its unfairness and callousness, it would undermine global cooperation, which is needed in order to protect the environment.

We are more ambivalent about vouchers and other policies that promote competition among schooling alternatives, largely because there are so many different varieties of school choice programs. On the one hand, the status quo in the United States is so unsatisfactory that we are inclined to welcome alternatives, particularly when they are unlikely to be more unfair to poor children. On the other hand, we take seriously the worries about solidarity and the balkanization of society that the collapse of public schooling may encourage. In either case, the economic efficiency argument does not seem to go to the heart of the issues.

Rather than resolving these issues, what we have tried to show is that a better understanding of the concept of well-being – including the peculiarities and limitations in the construal economists prefer – and a better appreciation of alternative ways to evaluate economic outcomes, institutions, processes, and policies permits a more sophisticated and informative appraisal of policy proposals. Mainstream normative economics is narrowly focused on preference satisfaction. That fact by no means makes it useless or silly, but it can be extremely misleading if economists – and those who listen to their advice – do not understand its foundations and its omissions.

SIXTEEN

Economics and Ethics, Hand in Hand

We have repeatedly extolled the virtues of interdisciplinary understanding. A better appreciation of ethics can contribute to economic analysis, and at the same time the tools and distinctions that economists have fashioned can contribute to moral philosophy. We hope that the examples we have already discussed and the detailed comments in the previous chapter concerning applications of mainstream welfare economics have shown that these are not empty slogans – mere reminders that more knowledge and more sensitivity are almost always better than less.

In this our last chapter, we would like to do more to show how ethics and economics can complement one another. We will begin by discussing briefly the examples presented in Chapter 3 concerning involuntary unemployment and the use of overlapping generations models to study retirement savings. Then we will sketch a range of urgent problems that cannot be tackled successfully without an understanding of both ethics and economics.

16.1 Involuntary Unemployment and Moral Baselines

While emphasizing the complexities of unemployment, we argued in Chapter 3 that controversies concerning the existence and analysis of involuntary unemployment result in part from conflating voluntary choice and rational choice and in part from normative disagreements among economists. The conflation leads some economists to regard as fully voluntary what are in fact paradigmatically involuntary choices, such as surrendering one's wallet to a robber who threatens death otherwise (Colander quoted in Snowdon and Vane 1999, p. 233). As Chapter 3 pointed out, a fully deliberate and rationally defensible choice need not be a voluntary choice. If handing over one's wallet to an armed robber is voluntary, what genuine choices could be involuntary? The fact that someone who is looking for a job could take up selling apples, stealing cars, or even suicide does not make his unemployment voluntary.

Whether a choice is voluntary – or, since voluntariness comes in degree, the extent to which it is voluntary or involuntary – seems to depend on two things: the process that determines the alternatives among which one chooses, and whether there are "acceptable" alternatives. The choice the robber gives the victim is almost totally involuntary because of the morally indefensible constraint on the set of alternatives and also because the victim is left without any good options. This means that whether someone is regarded as unemployed involuntarily or as having voluntarily chosen not to accept a worse job depends on social norms governing what sorts of jobs people can normally be expected to accept and what sort of processes can legitimately limit an individual's job choices. When economists commit themselves to describing unemployment as voluntary or involuntary, they endorse those norms. So a better appreciation of moral philosophy can shed light on what appears to be a controversy in positive economics concerning the nature and existence of involuntary unemployment. Moral philosophy helps by clarifying the concept of voluntary choice and by making explicit the conflicting moral norms that contribute to the controversy.

The particular day-to-day choices that consumers make in a market economy seem to be voluntary. Toyota versus Honda, Tide versus Cheer – the alternatives are surely acceptable and often nearly indistinguishable. For someone to say, "I had no choice but to buy a Toyota" would in ordinary circumstances be bizarre. In normal market circumstances, the choice of whether to take this job or that job, to prepare for this career or that career, has many of the aspects of ordinary consumer choice. The stakes are higher than in most consumer choices and the risks may be greater, but what is at stake are comparable alternatives, and the market mechanism seems a quite reasonable way to frame the choices. Therefore, such choices may reasonably be seen as voluntary.

But note that, even outside the context of job loss, the constraints on a person's economic choices may make them less than wholly voluntary. Consider, for example, the choice between owning and not owning a car. The Manhattan executive who prefers cabs and limos is voluntarily choosing not to have a car. For inner-city residents who need to spend two hours on three buses to get to the only available jobs, voluntariness is in doubt. By moving in with relatives or working even longer hours, a person in this situation could own a car. So there is a choice, but it is hard to see it as fully voluntary. Moreover, whether one views the way the alternatives are limited as legitimate depends on the facts. If public policies favor auto travel over public transportation or if they favor public transportation linking affluent locations over inner-city trips, then it becomes harder to view these inner-city residents as facing a voluntary choice. For one of them to say "I have no choice but to take the bus" seems perfectly sensible.

Economists who invoke the notion of involuntary unemployment resist reducing the "choice" to be unemployed to a normal market choice among alternative jobs. A person doesn't generally choose to become unemployed; he loses his job and can't find a new one that is roughly comparable. It strains language, such economists might argue, to say that the choice not to accept a much worse job like selling apples makes the circumstance of unemployment voluntary. Clearly, individual episodes of unemployment may differ from this paradigm case. Someone may lose her job because of willful poor performance or may refuse a job because of objections to the political views of prospective co-workers. When people speak of involuntary unemployment as a widespread social phenomenon, they presumably mean that there is a significant increase in job losses that is not explained by any coincident increase in people's fussiness about amenities or their tardiness or incompetence. The people haven't changed but circumstances have, and that leaves many folks involuntarily unemployed. Or so say a certain group of economists.

"Voluntariness" is here unmistakably a moral as well as a descriptive notion, with connections to other moral notions like blame, coercion, and responsibility. Those who regard the unemployment as involuntary want to emphasize that this worker is out of a job "through no fault of his own." They also want to suggest that the alternatives facing the worker are unacceptable and that perhaps different institutions or policies could provide those who are unemployed with better choices – or at least that such policies or institutions should be sought. A claim that there is widespread involuntary unemployment at a particular time is thus a claim that there has been a "market failure" and that some remedy is called for.

Critics of this view will contend that spells of unemployment are a normal part of a worker's life in a market economy and that substantial fluctuations over time in the average length of those spells is only to be expected. Workers are presumed to be making sensible choices about how long to keep searching for a high-quality job as opposed to taking a worse job sooner. Within the framework of a market economy, there is no fix for this: no pattern of government intervention can reliably overcome these fluctuations without imposing unacceptable costs of other kinds. The ups and downs of job holding and job seeking over a career are a "natural" part of the game.

Which is the right way to think about unemployment? What attitude should one have toward these market outcomes? This is partly a question for positive economic theory, but it also rests on moral judgments concerning what societies owe to their members and concerning the moral importance of security and solidarity as compared to freedom and independence. Although economists can remain neutral and confine themselves to stating whether members of a society regard unemployment as voluntary or involuntary in the light of their own norms, it is hard to divorce descriptions of market processes and outcomes from attitudes of approval or disapproval or evaluative comparisons with alternatives. These linkages should not be permitted to lie in the shadows, since much confusion and needless controversy could be avoided if economists were self-conscious and explicit about the values that so many of them share. Economists need not be embarrassed to admit that they value freedom, that they are committed to minimal benevolence, that they are anti-paternalist, or that they are admirers of markets rather than merely students of them.

The controversy over involuntary unemployment thus depends on attitudes toward markets in general and labor markets in particular. An unemployed worker who resists taking work shining shoes or selling apples on street corners is involuntarily unemployed if and only if she is justified in expecting better work. If there is a social commitment to providing work that is commensurate to skills and experience – by government intervention, if need be – then a good deal of unemployment is involuntary and so unemployment reflects a social failure. If there is no such social commitment, then those who are unemployed must learn not to expect anything better than they can get. Like cattle farmers, who cannot charge more than the market price for their steers no matter how well fed or well bred and no matter what they were worth in the past, workers need to anchor their employment and wage expectations on the actual opportunities and the going wages. Even so, those who think that the unemployed are not justified in expecting better work than what is available need not and typically do not believe that unemployment (or the changes in demand for labor that gave rise to it) is insignificant. Unemployment may, for example, indicate and contribute to diminished economic output and growth. But those who endorse market outcomes will not see unemployment as a market failure. On the contrary, they will regard markets as working exactly as they ought to, or anyhow as well as can be expected, and they will take unemployment to reflect voluntary choice. Expectations on the part of the unemployed for something better are unreasonable. Confusions aside, this is what is meant by claiming that unemployment is always voluntary and what is denied by those who insist that a great deal of unemployment is involuntary.

16.2 The Overlapping Generations Example

As we saw in Chapter 3, attitudes toward markets are also central to the controversies concerning Paul Samuelson's influential article on interest rates in an economy where saving is driven by the need to provide for retirement. Samuelson's article makes some theoretical points that might by themselves be regarded as without any ethical content. He shows that in a barter economy of the kind that underpins much economic modeling, market competition in the context of overlapping generations would lead to an inefficient profile of overconsumption in the first period of life followed by penury in the second and third periods. In order to overcome this market imperfection, the economy requires an institutional structure that relies on trust – the necessary institution being either a social contract in which younger generations acknowledge an obligation to support the elderly, or fiat money whose value is taken on faith by the members of the community. Yet, as Abba Lerner points out, the pattern of savings and consumption that would result from this trust is also open to ethical objection. If one permits interpersonal comparison, the pattern of saving and consumption in a growing economy leads to a smaller average utility in every period than a pattern that equalizes consumption across the periods of a life. The first two points were made by Samuelson and the third by Lerner with regard to a highly stylized and abstract depiction of an economy, and their application to practical problems would require much careful work.

We find this abstract theorizing instructive in three regards. First, although the above points can all be regarded as "positive economics," their interest and the way they are presented and debated depend on their relevance to normative issues. Samuelson is himself intrigued by his model because it presents a case in which perfectly competitive markets are inefficient. Since most mainstream economists regard perfect competition as an ideal (other things being equal), this was a striking result, which brought an immediate response from William Meckling.

Second, the model obviously bears on the issue of whether one should conceive of Social Security as a savings program or as a transfer program. Without an appreciation of the normative relevance of the abstract models, one would have a hard time understanding why Samuelson presents the model as he does and why he is challenged by Lerner and Meckling. At the same time, Samuelson's model also shows how much economists have to contribute to the understanding of normative issues. To draw clearly the distinction between a savings and a transfer program and to distinguish among different optimality criteria require the specialized technical competence that economists possess.

Third, Samuelson is in fact quite oblique in making the normative points highlighted here, and indeed his critics did not initially formulate their objections directly in ethical terms but instead in the guise of analytical disputes. We suggest that progress on both normative and positive issues would be more sustained if economists more frequently acknowledged the normative dimensions of their work. One main hope for this book is that it will make available concepts and tools that will facilitate their doing so.

16.3 Pressing Problems

There are many mundane detailed issues in economic theory, in moral philosophy, and in everyday life at both an individual and a social level in which ethical considerations influence economics and which are best tackled with a combination of ethical sensitivity and economic modeling. But to dramatize the importance of ethics to economics and of economics to ethics, we turn here to some of the grave problems now facing the entire world. For example, conflicts between religious and national groups have repeatedly led to war and genocide and threaten to lead to more. Greater firepower promises greater horror. These conflicts meld into the so-called war on terror, which threatens to play out the dynamics and destruction of ethnic and religious conflict on a global scale with the use of weapons of mass destruction. Nuclear proliferation threatens to transform even localized conflicts into threats of annihilation for whole populations and ecosystems. AIDS has killed millions of people and threatens to depopulate some nations, leaving them with masses of orphans whose needs will overwhelm their collapsing institutions. Widespread deprivation spawns famines and epidemics, and grotesque disparities in wealth and economic growth fuel conflicts. Economic growth and unrestrained consumption in wealthy nations combined with increasing population in the world as a whole have placed increasing strains on the environment everywhere and threaten massive climate changes with unknown but potentially nightmarish effects.

Let us pare down this litany – which is far from complete – to three serious problems:

- ethnic and religious conflict and terrorism;
- global distribution, famine, and epidemics;
- environmental degradation and global warming.

These are difficult and frightening problems, and one reaction is despair. But despair is a luxury. The problems cannot be avoided, and despair does not help to tackle them. To prevent the coming decades from going down in history as the era of disaster and annihilation will require four things of us.

1. (Recognition) We – the peoples of the world – must recognize the problems. Otherwise, people will not search for solutions or be willing to put up with the costs that solutions may require.
2. (Will) We must have the will to address the problems. The peoples of the world must be willing to devote resources to studying the problems, and they must be willing to sacrifice some of their immediate interests in order to lessen the risks these problems create and to prevent and alleviate suffering and destruction.
3. (Knowledge) We need knowledge to understand and address the problems. It is fortunate that – at the same time as we humans have been slaughtering one another, despoiling the planet, and turning our backs on human misery and degradation – our knowledge has been accumulating rapidly. We don't know nearly enough, and some of this knowledge only makes the mess worse by increasing our ability to kill and destroy. But increasing knowledge makes it possible to solve problems, too.
4. (Cooperation) A recognition of the problems coupled with the will to address them and the knowledge of how to address them will prove futile unless we collectively are able to cooperate. No single nation can solve these problems. Peoples and nations must be able to work together.

Recognizing that these problems require recognition, will, knowledge, and cooperation may make the prospects for solution seem even worse, but we take a more hopeful view. Some of the prerequisites for solutions are already in place. The problems are widely – though certainly not universally – recognized. Finding the will is more of a problem, and shortsighted selfishness will never be in short supply.

Nevertheless, we conjecture that enough people who are in a position to contribute to the solution to these problems are in fact willing to do so or would be willing to do so with the right leadership. In our view, the unwillingness to face the problems stems more from failures of leadership and uncertainties about what to do than from myopic individual selfishness. At the same time, advances in medicine and technology have been enormously helpful. We are not supposing that technology is now or ever will be able by itself to solve the problems. Technology is of little help with problems such as ethnic conflict, and the opportunities technology provides may come with risks that are sometimes hidden and sometimes all too visible. But advances in technology are, all the same, impressive. They have, for example, greatly diminished the risks of epidemics and famines, and they offer the prospect of greatly lessening our dependence on fossil fuels and greatly diminishing the environmental degradation caused by their extraction and use.

Where we see the greatest problems is in understanding the social dimensions of these problems and in conquering the barriers to cooperation in addressing them. We shall comment on specific challenges to which the three highlighted problems give rise and on how collaboration, or at least increased mutual understanding, among economists and philosophers can help people to understand and cope with the difficulties.

16.3.1 Ethnic and Religious Conflict

Consider first the problem of religious and national conflicts, the problems one sees played out in the Israeli–Palestinian conflict, in terrorist attacks on both Catholics and Protestants in Northern Ireland, in ethnic cleansing in the former Yugoslavia, in the slaughter of Hutus and Tutsis in Rwanda, or in the rapes and massacres in Darfur. In the popular media, these conflicts are usually attributed to extremist religious or nationalist hatred spawning a murderous cycle of revenge and anger. There is some truth in such depictions, but they do not explain how the people who are killing one another this week managed to live amicably as neighbors for decades or even centuries, as in many cases they have, or why the institutions that maintain the peace sometimes break down catastrophically. Furthermore, it is misleading to regard the parties to such conflicts as inevitably "extremists" unmoved by more quotidian desires for wealth and power. Neither Catholicism nor the varieties of Protestantism common in Northern Ireland are typically regarded as extremist religions. Though there are some religious extremists among both Jews and Palestinians, their influence on both parties has, at least until recently, been relatively minor.

Each of these conflicts is unique, and it would be foolish to suppose that some single theory will permit one to understand the many different ethnic conflicts, let alone to resolve them. Indeed, conflict in many cases may have deep roots in economic forces, including for example economic inequality and dependence on natural resources (World Bank 2003). But patterns of conflict are precisely the sort of subject matter that game theory is supposed to illuminate. So, for example, Hardin (1995) argues that ethnic conflicts can be explained by rational self-interested interactions among leaders and followers within competing groups. Although Hardin's book is a useful antidote to accounts of ethnic conflict as nothing but murderous lunacy, we believe that its application of rational choice modeling should be seen as only a suggestive sketch of what needs to be done. Understanding ethnic conflict requires models of strategic interactions in which important roles are played by motives other than material self-interest and the pursuit of political power. Hatred and longing for revenge as well as more positive aspirations for one's faith or nation matter, too. Commitments to social norms, including moral principles, also govern the actions of the conflicting parties. Some of these commitments, such as the endorsement of particular religious ideals, may be divisive; others, such as the recognition of the moral powers of human beings and the need for an overlapping consensus on principles of justice, offer a hope of a lasting and just peace.

Also of relevance in understanding ethnic conflict are the incentives that influence whether individuals emphasize or downplay conflicting ethnic or religious identities and whether they flaunt or conceal beliefs that may engender conflict. Kuran (1995, 1998) offers a rational choice analysis of such alternatives, one that complements the more sociologically and psychologically oriented accounts of related phenomena in Hirschman (1982) and Sunstein (2002).

We have no theory of ethnic conflict to offer and no solution up our sleeves, but we suggest that only by working together will moral philosophers, economists, and other social scientists enable us to understand these conflicts. Clearly these phenomena have a highly complex strategic structure, and game theory is the best theory we have of strategic interactions. Yet ethnic conflicts also reflect a complex set of motivational considerations with considerable fluidity and instability in people's preferences. So rather than proceeding on the basis of given preferences – as mainstream economists so often do – models of ethnic conflict will need to address how preferences form and shift and how they depend on moral commitments and social norms as well as on individual tastes and longings.

16.3.2 Global Inequalities

Let us turn now to the second set of problems, the unequal distribution of the world's wealth, the disparities in rates of per capita income growth, and the overall problems of economic development. One reason these problems are so hard is that economists have not yet figured out what determines economic development. However, quite apart from this task confronting economists, there stand a number of questions that lie on the boundaries between economics and philosophy.

1. What kind and level of aid do justice or benevolence demand of rich nations and their citizens? Which inequalities are injustices?
2. How strong are the claims of national sovereignty and noninterference when corrupt and repressive governments fail in their most basic responsibilities to their citizens?
3. What moral respect is due to cultures or ways of life that are threatened by global economic relations?
4. How should one understand and measure deprivation and development? Are economic measures such as per capita gross domestic product sufficient?

All of these are moral questions, yet moral philosophers have no chance of answering them without the help of the concepts, tools, and measurements that economists provide. For example, resolving the questions of Chapter 11 and determining what should be equalized require some knowledge of how resources interact and contribute to well-being, and egalitarianism cannot be translated into policy without knowing what can be measured and with what accuracy.

Problems of global economic justice or of economic justice among nations, which we pointed to in question 1, pose deep challenges to economists and moral philosophers. Much of mainstream international trade theory seeks to justify free trade by showing its benefits for all participating countries. Such arguments may fit naturally into versions of social contract theory that define justice in terms of mutually advantageous bargaining from a given (nonmoral) starting point, although they may beg questions about the winners and losers from free trade within individual countries. Pogge (2002) argues in contrast that the rich countries are causally involved in the poverty of poor countries – through the history of colonialism and through present-day rules of the international trading system – and that these causal relationships impose a moral obligation on citizens of rich nations to alleviate extreme poverty in poor countries. A different outlook is to find an appropriate way of interpreting "impartiality" in the context of relations among nations. Rawls's (2001b) approach, deriving from the social contract tradition, is to reason from the standpoint of deliberation among representative citizens of various countries who are ignorant of where their country stands in the global pecking order. Any useful attempt at reasoning about fair principles for international relations in that setting requires knowledge of economics: fair rules surrounding international transfers, for example, require weighing the efficiency costs of such transfers. Progress in understanding these matters and in crafting workable institutions and policies to further desirable policies can only benefit from the efforts of economically informed philosophers and morally informed economists.

Likewise, consider the fourth question, concerning measurement. Within standard welfare economics, well-being is preference satisfaction; and wealth (or income), which provides the means to satisfy preferences on the market, is a measure of overall well-being. But income, as measured for example by gross domestic product (GDP) per person, means different things in different contexts, and measures of income and wealth can be seriously misleading. Family incomes may tell us little about how resources are distributed within families; for instance, women may be destitute even though family income is not desperately low. In other cases, focusing on wealth may exaggerate how badly off people are. For example, GDP per person in the Indian state of Kerala is about the same as the average in India as a whole, but life expectancy there is more than ten years longer than in other Indian states and nearly everyone is literate. Indeed, literacy and life expectancy in Kerala are better than in U.S. inner cities, where GDP per capita is enormously higher even after adjusting for differences in purchasing power. It seems obvious that people in Kerala are better-off than people in several other Indian states, even though no richer.

How then should one measure deprivation? One way, which has been developed by the United Nations with the help of Amartya Sen, is to define a "human development index" as a numerical measure of the capabilities (see Section 8.5) that are available to the populations of nations of the world. Although its details are still under development, the index depends on life expectancy, knowledge (as measured by adult literacy and school enrollments), and GDP per capita. Such an index is obviously crude, and many of its features are determined by data limitations. But the index also reflects philosophical judgments concerning individual well-being or, more generally, what really matters to people. There is no way to compare the well-being of people in different countries and of different classes without (a) possessing some grasp of economic life, including an understanding of the relevance of particular data and the difficulties of acquiring them, and (b) committing oneself to some view of individual well-being. The choice between alternative possible indices depends on both economics and ethics.

16.3.3 Environmental Protection and Global Warming

Let us turn now to the last set of questions, those that concern environmental protection and global warming. Problems of economic measurement loom large here as well, since the consequences of human activities for climate, landscape, plants and animals, and air and water all have to be valued in some way. In considering how environmental economists attempt to quantify these predicted consequences, we return to the problem of articulating the proper role of mainstream normative economics – of finding some way to make use of its analytical techniques and the opportunities for quantifying costs and benefits that it offers without allowing them to distort or overwhelm other ethical considerations.

One might distinguish three different kinds of environmental problems: *pollution* of land, water, and air; *depletion* of natural resources, particularly farm land, fuels, fish, and forests; and *preservation* of species of plants and animals, places of natural beauty, and cultural landmarks. These three problems are not distinct from one another and are causally interrelated. One can also distinguish environmental problems by whether they are important mainly because of their consequences for people's material interests. Preventing or lessening the effects of global warming is important mainly because of the grave harm global warming is likely to do to people (Broome 1992). Preventing the extinction of the Siberian tiger, in contrast, won't save human lives or increase GDP.

Within the framework of mainstream normative economics, the treatment of problems of pollution, depletion, and preservation that bear on people's interests is apparently straightforward: one compares the extent to which various alternative policies satisfy preferences, using market data and willingness to pay as measures of the extent to which preferences are satisfied. In doing so, however, one immediately confronts a serious ambiguity that arises because people disagree about the consequences of activities and policies for pollution, depletion, and preservation and about the consequences of these for health, income, and all the other things people care about. Should one measure the extent to which global warming will affect welfare by imputing its costs from the amount people are willing to pay for increased fuel efficiency in their cars, or should one look to science for a prediction of the consequences of global warming and elicit information on people's willingness to pay with respect to those consequences? Although there are arguments on both sides, in our view the reasons in favor of satisfying preferences do not justify adhering to mistaken popular beliefs or to preferences that reflect mistaken beliefs (Hausman and McPherson 1994). Welfare economics should instead rely on the best supported estimates of the consequences of activities and policies and should attempt to screen out the effects of popular misapprehensions by eliciting preferences only for the consequences one has good reason to expect. As we noted in Chapter 2, this is what Summers does in his memorandum.

Such a framework for examining the significance of pollution, depletion, and preservation and for assessing possible policy responses obviously has its limits. Its measure of welfare is questionable both because well-being is not preference satisfaction and because willingness to pay is not a good measure of the extent to which preferences are satisfied. Its typical tools for dealing with the future (by "discounting" future benefits and future costs) are problematic in the context of long-lasting environmental change, since effects on the distant future are effectively discounted away. Insofar as environmental economics employs the framework of mainstream normative economics, it explicitly abstracts from distributional concerns, which it regards as separable though not unimportant. Rights enter only as devices to secure efficient outcomes, and mainstream normative economics makes no space for the view of pollution as a violation of fundamental rights to clean air and water.

Nevertheless, the mainstream approach has two crucial virtues: it quantifies costs and benefits and it does so in a way that bears a tangible (albeit imperfect) relation to human well-being. The quantification is crucial because it provides a way to make trade-offs between the benefits that polluting, depleting, or despoiling activities provide and the costs of pollution, depletion, and failure to preserve. However, these numbers are useful only if they are related to what we care about. Though the relation is not direct or unproblematic, it exists; and what better alternative do we have? Yet the danger is not that the virtues of quantification will be ignored but rather that they will be exaggerated and that the numbers will determine the policies. As even its staunchest proponents concede, welfare economics captures only part of what matters, and it captures that part imperfectly. Net benefits must be weighed against other factors. But how does one "weigh" numbers against unquantified matters of equity, rights, freedom, or solidarity?

These difficulties become more poignant when one turns to those environmental issues where human well-being is not the main consideration. One can try to shoehorn these problems into the framework of mainstream economics, but not without doing violence to the concerns at stake. Many people care about preserving endangered species, old-growth forests, canyons, or marshes. In some cases, this preservation will serve people's interests. For example, marshes purify water, flora and fauna may yield life-saving drugs, and people enjoy hiking, boating, and bird-watching. But people's concerns about preservation are often not self-interested. Many features of our planet and of the creatures that live on it are magnificent in themselves, and there is something worthy or even obligatory about preserving them – quite apart from the ways in which they may tangibly benefit people. Some people hold similar attitudes toward historical and cultural monuments and indeed toward cultures themselves. How should such "aesthetic" or principled concerns be understood, and what influence should they have on policy making?

Many environmental economists have been tempted to describe such concerns in terms of individual preferences and to quantify them in terms of willingness to pay. If someone thinks that it would be a bad thing if Siberian tigers became extinct or if a shopping mall were built where the Auschwitz crematoria stood, then presumably they would prefer states of the world in which Siberian tigers are not extinct and there are no shopping malls in Auschwitz. And it is surely the case that many people would be willing to pay something in order to preserve species, landscapes, monuments, or cultures. Since there are no markets for species, landscapes, monuments, or cultural preservation, one can treat the fact that aesthetic and principled views about the environment have little influence on market outcomes as if it involved an ordinary market failure. Cost–benefit analysts can then use willingness-to-pay information to value environmental preservation and thereby to respect consumer sovereignty.

In this way one can quantify aesthetic and principled concerns to protect species, landscapes, watersheds, historical monuments, or cultures, and by quantifying these concerns one provides a way of trading them off against other values. But this tactic has huge conceptual costs (Sagoff 2004). By a sort of definitional alchemy, non–self-interested concerns with these things have been transmuted into individual interests. Because principles influence preferences, which mainstream normative economists in turn identify with welfare, considerations of principle are treated as if they were individual interests. By conflating principles and interests, the measure becomes individual satisfaction or willingness to pay rather than the force of the *reasons* why it is or is not important to preserve something. For example, preserving some wilderness area may serve people's interests by providing recreational opportunities, and at the same time it may also have a principled or aesthetic importance. Its worth as a source of well-being to people is a matter of how it affects the people who use it and is reflected in the extent to which it satisfies their preferences or improves their well-being by some more objective standard. Its aesthetic, historical, cultural, and intrinsic worth is not a matter of how it much it advances the interests of the people who use it. That many people who do not individually benefit from Siberian tigers would nevertheless be willing to pay to preserve them does not show that they individually benefit from the tigers all the same. Nor does their willingness to pay track the moral importance of the principles they appeal to. Instead it reflects their belief that these animals are magnificent and that the earth would be poorer without them. It is difficult to say how important this magnificence is, and it is easy to see why the shortcut of looking to willingness to pay is so attractive. But willingness to pay is at best an imperfect measure of individual benefit, and what is at stake with respect to preservation is not individual benefit.

It may be helpful to emphasize this point by considering the related problem of animal suffering. Some environment policies and many policies regulating agriculture influence the amount of pain and suffering that animals endure. A mainstream economist might argue that concern about such suffering should be viewed in terms of its effect on human well-being. The importance to be attached to animal suffering depends only on the degree to which people prefer that animals not suffer, and this might be measured by their willingness to pay. On certain views of utilitarianism, in contrast, the suffering of animals counts just because it is suffering, and if a change in policy can raise the utility of animals without causing a larger reduction in the utility of humans then that change should be undertaken – whether or not the suffering of animals enters into the utility functions of people. Similarly, on some rights-based views (including certain kinds of social contract views), the well-being of animals commands moral concern because of their status as feeling and thinking creatures, and their interests therefore have moral weight regardless of the degree of concern that people might have for them.

Taken as an overarching moral view, the neoclassical economic outlook can be seen as making three large claims: (1) the only objects of intrinsic moral importance in the world are human beings – everything gains its moral importance only through its effects on human beings; (2) the measure of this moral importance is the subjective preferences of human beings; and (3) there is no way to compare the intensity or importance of the preferences of different human beings. Few economists, if presented directly with those three claims, would leap to endorse them, but they are implicit in the efforts that many economists make, when presented with moral considerations that don't naturally fit into the scheme of mainstream economics, to force them into that framework.

By recognizing the limitations of mainstream normative economics and showing a greater sensitivity to the many moral issues that do not fit and cannot be made to fit within standard welfare economics, economists can contribute to the understanding and resolution of problems of ethnic conflict, global inequalities, and environmental destruction without distorting and burying competing values. Economics informed by ethics and ethics empowered with economic modeling have a huge task ahead of them in diagnosing the causes of these problems and in developing and assessing alternative solutions. Economics and ethics can together contribute also to solving the hard problems of cooperation that beset efforts to implement solutions to these problems.

Obviously there is plenty of work for others, too; and it would be absurd to place the salvation of the world in the hands of economists or moral philosophers. But we hope that the speculative words of this section heighten your appreciation of the ways in which economics and ethics can fruitfully inform each other.

16.4 Conclusions

After working through the arguments and analytical perspectives reviewed in this volume, different readers will arrive at different judgments about practical policy options such as exporting polluting industries to poor countries. We have not defended any single view of ethics, and we have not attempted to weight the differing considerations. Furthermore, policy conclusions depend also on judgments concerning the likely consequences of alternatives, and there are many disagreements concerning these matters. But we hope that our discussion encourages readers to reflect on this question: Should one rest content with evaluating policies solely according to their consequences for satisfying people's preferences (with perhaps a few adjustments to allow for equity concerns), or should one strive to include in one's evaluation the consequences for freedom, equality, justice, solidarity, and quality-of-life indicators that go beyond preference satisfaction? Although the former path may keep economics neater and perhaps in appearance more "scientific," the latter will, we contend, make it more useful and defensible.

As we have stressed repeatedly, and as our examples of involuntary unemployment and the overlapping generations model illustrate, the relevance of ethics is not limited to the context of evaluating policy. Ethics enters directly into the development of social welfare theory and into the modeling of strategic interactions as games. It is needed in order to formulate relevant questions for "positive" research. It is implicit in the standard theory of rationality that lies at the core of contemporary economics. It is implicit in theories of how labor markets work and of how social cooperation in general is possible. Economics is certainly not a wayward branch of moral philosophy, but neither is it unrelated to moral philosophy. One enriches both economics and moral philosophy by acknowledging their interdependence.

APPENDIX

How Could Ethics Matter to Economics?

We hope in this book to have shown how knowing moral philosophy can help one to do economics better. The most persuasive way to make this case is the one we pursued in the main text: by describing important aspects of moral philosophy and showing their bearing on economics. But many economists are inclined to deny that moral philosophy has anything to do with economics. Why? In this appendix, we shall explore the reasons and reaffirm the conclusion of this book – that ethics *is* relevant to economics.

As explained in Chapter 1, one can distinguish between "positive" and "normative" inquiries. Positive inquiries address factual questions, whereas normative inquiries address evaluative questions. Although it sounds paradoxical, it is possible to regard what is called "normative economics" as a positive inquiry into the logical presuppositions and practical means to satisfy preferences efficiently, and some economists have in fact viewed normative economics in this way. Most economists, however, concede that normative economics is a normative inquiry addressing evaluative questions and prescriptions; and, as our discussion in Chapters 2 and 15 shows, arguments such as the Summers memorandum do not address only factual questions.

Hence we shall interpret those economists who would deny that moral philosophy is relevant to economics as distinguishing positive and normative economics, as conceding that moral philosophy is relevant to normative economics, and as denying that moral philosophy is relevant to positive economics. In some cases that denial is linked to a repudiation of normative economics coupled with the view that economists should contribute to policy questions exclusively by providing information concerning the consequences of alternative policies. The first section of this appendix addresses this position and considers the role of positive economics in policy making. The second section turns to the general question of what it means to say that positive economics is "value free" and then sketches what we call "the standard view" of the relation between positive and normative economics. The third section comments briefly on the exaggerated distinctions sometimes drawn between positive and normative inquiries, criticizing in particular the view that evaluative claims cannot be rationally addressed. The fourth section then considers whether positive economists have anything to learn from the study of ethics.

A.1 Objection 1: Economists as Engineers

Economists who deny that ethics is relevant to economics and who repudiate normative economics as the bastard offspring of completely distinct inquiries do not deny that economics is relevant to policy making. It is relevant – vitally relevant – but only in the way that civil engineering is. Owing to a need for electric power, policy makers might consider building a dam. Civil engineering does not say whether this is a worthy objective. Engineers instead provide information about how difficult it is to build dams in different locations, how much electricity the dams can generate, how much land they will flood, and so forth. Engineers thereby provide answers to some of the "What if ...?" questions that policy makers need to answer when they are deciding whether or where to build a dam. In order to accomplish anything, one needs knowledge of cause and effect. Engineering is one source of such knowledge. The first objection to the view that ethics is relevant to economics asserts that the role of economics is exclusively to provide causal information. Ethics helps determine the ends that policy makers pursue, and it constrains the means that may be employed. Economics clarifies the consequences of alternative policies. So-called normative economics confuses matters by amalgamating these completely distinct tasks. Both economics and ethics are crucial to policy making, but neither has anything to do with the other.

The following very simple schema might help clarify this view.

1. Our policy ought to achieve goal G and satisfy constraints $C_1, \ldots, C_n$.
2. X satisfies $C_1, \ldots, C_n$ and achieves G.
3. Thus our policy should be to do X.

In this oversimplified schema, premise 1 comes from ethics or political philosophy while premise 2 comes from economics and other relevant bodies of empirical knowledge. The conclusion is a moral judgment, and it requires both the moral premise 1 and the technical premise 2. On the view of economists as engineers, premises 1 and 2 have nothing to do with each other, and ethics has nothing to contribute to economics.

This schema is too simple. After all, X may have other desirable or undesirable features, and there may be better alternatives. The following schema is more defensible.

1. Our policy should be governed by the complete moral or social ranking R.
2. The consequences of X rank higher in R than do the consequences of any other policy under serious consideration.
3. Thus our policy should be to do X.

The moral part of policy making, represented by premise 1, requires not only a specification of a particular goal and constraint but a complete determination of the relative moral importance of the various consequences that different policies may have. For example, in the case of a proposed law prohibiting arbitrary firings, R would specify the moral importance of arbitrary firings, of intimidation of workers on the job, of unemployment among the lowest strata of workers, of difficulties for firms in firing workers, of overall productivity, and so forth. The economist's job is then to determine all the relevant consequences of the alternative policies. Once all the consequences are known, the policies can be ranked in terms of their consequences. If policy makers do not know for sure what the consequences of policies will be but still can estimate the probabilities of outcomes, then they can rank policies via weighting the moral values of their outcomes by the probabilities that they will obtain. As in the simpler schema, the policy conclusion depends on both moral and technical premises that are independent of one another.

The second schema, unlike the first, is not too simple. Its drawback is that it is too demanding. No existing moral system is refined enough to provide the needed first premise, and economics is not up to the challenge of providing the second. Real policy making relies instead on partial moral rankings of salient features of policies and their consequences. The terms of the partial rankings rely on positive information concerning what sort of consequences the policies under consideration have, while the positive study of their consequences is channeled by normative views of what kinds of consequences matter.

Although the view of economists as pure engineers thus cannot be sustained, there is some truth to this picture of economists as purveyors of technical information. It is a useful caricature. It fits some economic activities – for example, work devoted to estimating how much revenue would result from changing income-tax rates. But economists need to understand the values to which policy makers are committed in order to understand what policy makers want to know and what questions to ask. Economists need not share those values, but they do need to understand them.

As Fritz Machlup (1969) recognized, applying the second schema to the activities economists undertake is often impossible. Machlup made the telling observation that the political process rarely formulates its economic problems clearly. When economists are called on to give "purely technical" advice about how to accomplish certain ends, they are rarely given purely technical problems. Just think about the tasks of economists who are asked to advise governments on how to transform formerly socialist command economies into market economies. Without knowledge of the prevailing system of values and moral constraints in those societies, they will not know how to proceed. Even if wholly without moral views of their own (as if that were possible!), economists concerned to give policy advice will need to understand the structure of the moral theory or theories that are implicit in people's ideals and commitments. Although these may be different from the economist's own commitments, there are likely to be many points in common, and economists who understand their own moral perspective and who have some general grasp of ethics are likely to understand the moral perspectives of others better. It is likely (though not inevitable) that the values to which economists are committed – including their best judgment about what will be good for the people whose government they are advising – will influence what alternatives they consider and what weights they place on the comparative advantages and costs. Could policy makers in Bulgaria or Latvia possibly provide their economic advisors with a full list of all relevant goals and constraints together with precise weights and priorities specified for each? To give advice, economists need to know what other objectives policy makers have and how to weight them. Although in principle all this might be precisely specified, in practice economists will have to rely on a general grasp of ethics to fill in the gaps. Economists may not think systematically about ethics, and they may not want to think about ethics at all. But they will not be able to understand what policy makers want or to translate policy problems into technical problems of economic analysis without some moral understanding.

We are not denying that positive inquiries into technical questions are possible. Of course they are, and a good deal of economics is devoted to them. But economists cannot rely on policy makers to formulate these questions. Though they need not share the values of those whom they advise, economists who seek to weigh in on policy questions need to understand the moral commitments of their advisees. (And if economists find that their values conflict with the values of those whom they advise, then economists will have ethical problems of their own to wrestle with.)

Similar problems arise when economists select problems to investigate. Most economists want to solve problems that matter to people's lives, and they want to solve theoretical puzzles in ways that do not conflict with their moral commitments. So long as economic processes and outcomes remain so important to human interests, evaluative commitments are bound to be crucial to the choice of problems to investigate. In stressing the evaluative roots of economic research, we intend no criticism of this research. We doubt whether there is any alternative, and we are confident that it would be a bad thing if the moral commitments of economists did not inspire their research. Economics is not only devoted to pure research and puzzle solving, it is also relevant to the pressing practical problems we humans face. If economists refuse to deliberate over messy moral matters, then they will not know which questions are important. Economists do not face prespecified technical problems. In deciding what to study and in thinking about how to apply economics to practical problems, economists must think about ethical matters. They need not do so systematically, self-consciously, or well, but they cannot avoid ethics altogether.

In fairness to engineers, we suspect that the complexities we have noted for economists apply often to engineers as well. In designing a superhighway, the civil engineer is unlikely to have available a social ranking of all the various combinations of aesthetic appeal, safety, durability, expense, ease of traffic flow, and so on that combine to make a good highway. As Donald Schon (1983) has observed, much professional practice involves wrestling with situations in which the normative and the positive are deeply entangled.

A.2 Objection 2: Positive Economics Is Value Free

Economists cannot avoid ethical questions if they want to understand the terms of policy debate, to help determine public policies, or to select problems for study. But many economists would argue that these conclusions do not preclude the possibility of a positive science of economics. These economists would grant that ethics has a role in determining what problems to study, pointing out that it is hardly surprising that ethics is relevant to normative economics. But they would insist that the role of ethics ends there: Ethics may pose the economist's questions, but it cannot contribute to their answers. Ethics has nothing to contribute to positive economic analysis. Thus, the second objection maintains that there is a positive economic science and that this science is "value free." Let us call this position *the standard view.* Some further clarification is in order, as follows. (For an insightful alternative view see Mongin 2006.)

A.2.1 Positive and Normative Economics

As mentioned in Chapter 1, most people can readily distinguish questions concerning (nonmoral) facts from questions concerning what is good or bad or what ought or ought not to be done. This is the so-called fact–value distinction. It is difficult to make this notion precise. There are many hard cases. (In stating, "That was a kind thing to do," is one describing or appraising?) And some philosophers argue that the fact–value distinction breaks down altogether. But let us suppose that it is possible roughly to classify questions as either factual or evaluative. Ethics is then taken to be concerned with questions of value, while the sciences (or at least the natural sciences) are taken to be concerned with questions of fact.

The standard view maintains that questions of fact and questions of value are not only distinguishable but, moreover, independent. No question concerning values is supposed to be settled by the facts alone, and no question concerning facts is supposed to be settled by values. On the standard view, it is accordingly possible for there to be value-free inquiries into matters of fact.

To speak of a value-free inquiry may be misleading, because it suggests that the *conduct* of the inquiry is value free. But the conduct of inquiry cannot possibly be value free. Inquiring involves action, and action is driven by values. As we have already seen, values influence choices of what to study. Values also influence choices of what methods to employ and consequently of what hypotheses to discard or to pursue. It is because of their moral identification with the goals of science that economists resist "cooking" their data. It is because of their personal morality that economists rarely shoot those who disagree with them (or try to get them fired). The standard view does not deny that values influence the conduct of inquiry. What is meant by a value-free inquiry is instead (a) an inquiry into a question of fact or a question of logic in which (b) the answers are not influenced by any values except those that are part of scientific activity itself.

Investigations into matters of fact are called positive investigations. The standard view concedes that the results of positive inquiries may be relevant to policy, because those results may show that policies facilitate or frustrate the attainment of valued objectives. But without some prior evaluative commitment, the findings of positive science settle no questions of policy or value. According to the standard view, positive science can be value free, positive science ought to be value free, and – apart from lapses – positive science is in fact value free. Thus the study of ideology and of the values of economists is irrelevant to understanding economics or economic methodology, though it may help one to understand the scientific failings or motivations of particular individuals.

A.2.2 On the Independence of Ethics and Economics

The standard view does not repudiate normative economics, and it does not maintain that economics bears on policy only through its determination of the consequences of alternatives. So our responses in the previous section to these more extreme views should not be taken as a criticism of the standard view. On the standard view, economics consists of many different activities, and ethics is relevant to some of them. Economists are human beings with human interests, and they are accordingly deeply concerned with ethics and economic policy. Parts of economics are thus unsurprisingly concerned with ethics. But it does not follow from this that ethics is relevant to all of economics. In particular it does not follow that ethics is relevant to positive economics, to the part of economics that is concerned to represent, explain, and predict how economic systems function.

As a result, those who hold the standard view can readily concede Machlup's point that economists need to understand some ethics in order to appreciate the objectives of policy makers. Only with reference to such moral understanding will economists be able to formulate clear and relevant technical problems. Economists who hope that their work will be relevant to policy questions need to know some ethics. But once precise technical questions have been formulated, there is no reason to believe that knowing ethics has anything to contribute to answering them. On the contrary, according to the standard view, there is a categorical difference between questions concerning facts and questions concerning values, and ethics will not be relevant to the answers to factual questions.

It may well be impossible to do economics well without knowing some ethics, because ethics is so important in formulating research problems and applying the solutions. But on the standard view there is still a body of knowledge – namely, positive economics – to which ethics has no relevance.

A.3 The Rationality of Normative Inquiry

On the standard view, normative science, which is contrasted to positive science, consists of inquiries into matters of policy or values. Normative economics consists of the application of positive economics to explore questions that are of evaluative relevance. As mentioned in Chapter 1,

Factual Claims	Evaluative Claims
Disagreements can be resolved by evidence	No good way to resolve disagreements
Relatively little disagreement	Relatively little agreement
Descriptive: say how things *are*	Prescriptive: say how things *ought to be*
True or false	Not true or false
Objective	Subjective
Independent of evaluative claims	Dependent on factual claims
Help to achieve goals	Help to determine goals

Exaggerated Contrasts between Facts and Values

the distinction between factual claims and evaluative claims and hence between positive and normative economics is often drawn very sharply. Thus, for example, Milton Friedman maintains with respect to disagreements about fundamental values that "men can ultimately only fight" (1953, p. 5). Reprinted above for the reader's convenience are the stylized distinctions of Figure 1.2.1.

Sharply distinguishing between facts and values and between positive and normative science does not, however, commit one to this set of distinctions. Nothing is settled concerning the status of questions of value when one insists that they are different from questions of fact. If facts and values are completely independent, then all that follows is that truths concerning values (if there are any) are different kinds of truths than truths about facts, and that truths concerning values are known in some way other than via scientific confirmation. But these differences do not imply that there are no truths about what is good or bad, right or wrong, praiseworthy or blameworthy. Moreover, and perhaps even more importantly, the possibility of distinguishing between positive and normative economics does not even imply that they should be pursued separately (Weston 1994).

At issue in this appendix is whether positive economics is value free, not whether rational arguments concerning values are possible; even so, we do want to cast a few words of doubt on the distinctions listed in the figure. A full critique calls for a long story, which we will not give here; and we do not wish to take issue with every one of the distinctions listed. The main point we insist on has, we hope, already been demonstrated in the chapters of this book. It is that *normative questions* – questions concerning evaluative claims – *are subject to rational discussion and resolution.* Consider, for example, our discussion of a preference satisfaction theory of well-being, the adequacy of utilitarianism, or the plausibility of the goal of welfare equality.

One addresses such questions with rational arguments rather than experiments (which are, in any case, only rarely possible even in positive economics). Rational argument does not suffice to resolve all controversies. Nevertheless, principles can be tested by examining their implications, and specific judgments can be criticized on the bases of their factual presuppositions and the moral principles upon which they rely.

So we think it is mistaken to suggest that there is no way to resolve disagreements about values. Substantial disagreements remain, but it is easy to exaggerate the prevalence of disagreement because, of course, people don't argue about what they agree on. What's crucial is not exactly how much agreement and disagreement there is but that rational persuasion is both possible and real.

The other contrasts we want to contest are those between, on the one hand, the objectivity and truth or falsity of factual claims and, on the other hand, the alleged subjectivity and arbitrariness of evaluative claims. Whether evaluative statements are literally true or false is a difficult question that we will not attempt to answer. Prescriptions cannot be literally true or false, but not all evaluative claims are prescriptions (consider again the statement, "that was a kind thing to do"). What is crucial is that not all answers to evaluative or specifically ethical questions are equally good. Some are way off the mark, while others are in some relevant sense correct. If by "objective" one means that claims are correct or incorrect regardless of whether people believe they are correct, then – as our discussion shows – many evaluative claims are objective. Whether a market in pollutants would be beneficial is not determined by whether people think it would be. In other senses of "subjective," evaluative claims are subjective because they often express subjective states. But claims in positive economics concerning preferences or expectations are subjective in much the same way. The bottom line is that recognizing that there is an important difference between facts and values and between normative and positive inquiries does not require one to denigrate normative inquiry or to insist that the inquiries be isolated from one another.

A.4 How Knowing Ethics Contributes to Positive Economics

Before addressing the question of whether studying ethics might have something important to contribute to positive economics, it is important to recognize how much the standard view concedes. It confesses that economists may need to understand the concepts and the criteria that guide the evaluation of economic outcomes and processes, and it concedes that ethics has an important part to play in economics. It concedes that it may be difficult to be a good economist without knowing some ethics. All the standard view maintains is that positive economics, considered as a body of knowledge, is independent of ethics. We can thus grant this objection without abandoning our project of showing how much ethics has to contribute to economics. It is important for economists to know some topology and statistics even though there are important parts of economics to which these fields contribute nothing, and likewise we maintain that it is important for economists to know some ethics even if is not relevant to everything economists do.

We have no wish to deny that there is a good deal of truth to the standard view. Some work in economics is largely independent of all specifically ethical concepts and theories, though less obviously independent from evaluative concepts in general. Consider all the work that goes into estimating demand elasticities. Yet as examples in this book show, positive economics is sometimes penetrated by ethical concerns. We showed that ethical commitments on the part of economists play an important role in discussions of overlapping generations models and of involuntary unemployment. But we do not maintain that such cases are ubiquitous. We do not know how large a part is played in positive economics by the moral appraisals economists make of economic processes, institutions, and outcomes – beyond their role in suggesting questions to investigate or possible answers to be assessed. What we want to argue instead is that studying ethics can make a contribution even to inquiries that are, in the relevant sense, value free. Studying ethics can be useful even in circumstances in which the theorist's own values do not come into play.

Ethical commitments are among the causal factors that influence people's economic behavior, and hence they are among the factors with which economists need to be concerned. If people did not generally tell the truth and keep their promises, then economic life would grind to a halt. As theorists who study labor markets have noted, employees and employers have moral beliefs that affect the wage and employment bargains they make. People's moral dispositions affect economic outcomes.

This fact does not imply that economics cannot be value free. People's beliefs and preferences, including their moral beliefs and preferences, obviously have economic consequences. But so do facts about their physiology or about the terrain. The mere fact that moral beliefs and preferences are important causal factors does not show that economists need to pay more attention to morality than to biology or geology.

However, the importance of knowing some ethics cannot be dismissed this easily, for two related reasons. First, ethical commitments, unlike physiological or geological facts, are not just givens: they depend on economic institutions and outcomes. Second, their content and this dependence on economic institutions and outcomes are very hard to grasp without understanding a good deal about ethics. An example will help clarify these claims.

In 1971 Richard Titmuss published *The Gift Relationship: From Human Blood to Social Policy.* In this book, Titmuss compared different systems by which human blood is collected for the purposes of transfusion. He was particularly concerned to contrast the system in Britain, in which all blood was obtained by voluntary donations, with the system in the United States, in which some blood was donated and some blood was purchased. He found that blood shortages were more severe in the United States and that the incidence of hepatitis and other blood-borne diseases was higher. The monetary costs in the United States were also much higher. The system in the United States appeared to be in some sense much less efficient.

Titmuss went on to offer a striking causal explanation for these data. He argued that the existence of a market in human blood undermines people's willingness to supply blood, which in turn causes the mediocre outcomes of commercial systems. In short, Titmuss argued that the existence of a market *causes* the efficiency loss.

The comparison between Britain and the United States was not Titmuss's only evidence for this striking and (for most economists) paradoxical claim. Titmuss also pointed out that in Japan prior to World War II, there was a voluntary blood donation system with outcomes similar to those in Britain. After World War II a commercial system was instituted, and the outcomes then resembled those in the United States. Blood donations dropped precipitously. The evidence from Japan is particularly impressive because it shows blood shortages developing after the institution of a commercial system for acquiring blood – even though Japan, unlike the United States, is a homogeneous nation with a great deal of social solidarity. In addition to the statistical evidence, Titmuss pointed to statements people make about *why* they donated blood. Donors repeatedly said that they were giving "the gift of life." Implicit in this discussion and explicit in Peter Singer's later (1973) defense of Titmuss from Kenneth Arrow's criticisms was the thought that, when a pint of blood can be purchased for \$50, donors may feel that instead of giving something priceless – the gift of life itself – they are giving the equivalent of \$50. Hence people might be less willing to donate blood when blood can also be obtained commercially. Their moral commitment depends on the institutions and is not simply a given.

The bad consequences – higher costs, shortages, and more hepatitis – can then be explained. Since fresh blood is perishable, supplies must be regular and blood cannot be stockpiled to accommodate fluctuations in demand. Shortages may result under a commercial system because (a) people are less willing to donate blood and (b) the amount of blood available from the small part of the population that is willing to sell blood increases and decreases irregularly. The higher costs are obvious. Finally, in commercial systems, unlike systems involving only voluntary donation, people have an incentive to conceal illnesses such as hepatitis.

This case illustrates how people's moral beliefs and preferences influence economic outcomes and also how economic arrangements (in this case, whether there is a market for human blood) can influence what people feel a moral obligation to do. We do not know whether the explanation offered by Titmuss for the dramatic differences between outcomes in the United States and Britain is correct, but there is no way to understand or assess it without attention to systems of moral beliefs, which (according to Titmuss) explain the choices of individuals.

If there is anything to Titmuss's account, then one cannot understand the contrasts between different systems of blood donation without what might be called a "deep" understanding of moral beliefs. In addition to knowing what people think is permissible, obligatory, or impermissible to do, or merely knowing how they value particular outcomes or states of affairs, one needs to know how these beliefs fit together and how they depend on features of social and economic institutions, processes, and outcomes.

All of this is still arguably sociological knowledge. Economists do not need to make their own moral appraisal of systems of blood donation or of the values of blood donors. What matters are people's behavioral dispositions and their *beliefs* about what is right and charitable, not what is "in fact" right or charitable. Just as economists can study the art market without appreciating art or the wine market without a cultivated palate, so can they study the economic causes and consequences of moral beliefs and preferences without knowing what is right or wrong. Positive economists studying blood distribution systems do not need to make moral judgments or to evaluate moral beliefs. All that matters is whether economic agents make those judgments and have those beliefs – and what their causes and consequences are.

But it does not follow that economists have little to gain from studying ethics. Although they do not need to pass moral judgments on systems of blood donation, they need to understand how people think about blood donation and what the commitments of people depend on. Without some grasp of the moral issues at stake, one will not be able to understand why people give blood or what impact a market in human blood will have on their continued willingness to supply blood.

Some of Titmuss's data, such as the lower rate of disease, make good economic sense. But why should the possibility of selling blood *decrease* its total supply? In a generally sympathetic review, Arrow can make little sense of this. Part of the problem may be in some of Titmuss's specific formulations, as when he says that markets "deprive men of their freedom to choose to give or not to give." The quotation is puzzling because the option to donate blood still exists, and indeed it seems that the existence of markets can only increase freedom by providing an additional alternative. Titmuss also seems to assume that altruism is virtually unlimited, and Arrow is reasonably skeptical about that. Perhaps we are better advised not to expend our limited altruism when it is not needed. As Adam Smith pointed out:

> But man has almost constant occasion for the help of his brethren, and it is in vain for him to expect it from their benevolence only .... It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love and never talk to them of our own necessities but of their advantages. (1776, pp. 26–7)

Why should the blood donor be different from the butcher or the baker? An economist might claim that it makes sense to call upon limited supplies of altruism only when there is no alternative (Robertson 1956). Apart from the specific problem that results from imperfections in the ability to test blood and the consequent incentive that commercial providers have to lie about their health, it would thus appear that markets in blood – like markets in bread, meat, or wine – can only be for the good.

But as Arrow recognizes, Titmuss's case for a voluntary blood donation system cannot be easily dismissed. Although the good will of neighbors, fellow citizens, and fellow humans is limited, the best society does not minimize acts of altruism. Obviously acts of altruism are costly. (They wouldn't be truly altruistic if they involved no sacrifice at all.) It takes time to donate blood, and sometimes one feels sick or dizzy for a while. It's not very pleasant. But altruistic acts carry rewards, too. People take pride in doing something they take to be decent and unselfish. They take pleasure in thinking of the good their blood may do for someone else. They take pleasure in the good opinion others may form of them (though the concern for reputation depends upon the act of giving being independently valued). Having given blood once, an agent may be more rather than less likely to give again. If many of one's neighbors donate, then one still may be more rather than less likely to give, even though the need for one's gift is reduced by the larger supply. Altruism is scarce but it is not in fixed supply, and in some instances its supply increases with its consumption (Hirschman 1985). To appraise and comprehend Titmuss's explanation, one needs to understand factors such as these. And there is no way to do so if one has no understanding of morality.

Arrow is puzzled by the claim that people are on average less willing to supply blood when it can be sold as well as donated, and he is unwilling to accept this proposition because he cannot understand why it should be true. What is needed to convince Arrow is not statistical data but an explanation of the structure of people's moral beliefs and of how moral beliefs change. Arrow is not convinced because Titmuss fails to make clear why people may be more strongly motivated to make more valuable gifts and how commercializing blood diminishes the value of the gift. If one cannot make sense of these claims, then one will find Titmuss's story an inscrutable one of strange irrationality. We see no reason why there should not be true stories of strange human irrationality. But Titmuss's story is much easier to credit, and the pieces easier to put together, if only one can "make sense" of it. And there is no way to make sense of it without entering into the moral universe of the donors, learning its rules and learning how it is affected by factors such as whether blood may be sold.

Those who are not convinced might complain that we have shown only that economists sometimes need to know what ordinary people think about morality, not that economists themselves need to think about morality, let alone the sophisticated theories surveyed in this book. Part of this complaint is surely correct. Much esoteric work on moral philosophy will not help anyone to understand how the moral convictions of ordinary people hold together, what they depend on, or how they influence actions. But there is a great deal in ethics that does bear on everyday moral thought. Thinking about morality contributes enough to the understanding of how ordinary people think about morality that it can be of great value to economists.

The moral commitments of economic agents can reasonably be regarded as sociological and psychological factors, but they differ from other sociological and psychological factors because they are supported by reasons and held to be generally binding. They need not always make sense; but it's puzzling when they don't, and one's first reaction will be that one has not understood them correctly. To understand economic phenomena, economists will sometimes need to understand the moral commitments of agents, for these may be of great economic importance. And doing so requires some ability to enter into the moral universe of the agents. Furthermore, the consequences of economic institutions and policies will often be mediated by their effects on people's moral commitments. Knowing something about morality will sometimes be crucial to predicting and explaining how people's moral beliefs and preferences change in response to economic policies and institutions, and it will thus enable economists to predict economic outcomes more accurately.

There is no general justification for setting aside questions about the quality of the arguments supporting the moral commitments of economic actors. Economists may be able to advance their work by appraising people's moral dispositions as well as by tracing their causal consequences. But whether or not economists make moral judgments when they are attempting to understand the moral commitments of the agents they study, it is natural, illuminating, and virtually unavoidable for economists to inquire whether people's observed (or alleged) moral commitments "make sense" – and it takes some understanding of ethics to make this inquiry.

There is a further complication: the moral convictions of economic agents, unlike causal factors such as rainfall, can be influenced by the way in which they are analyzed and described by economists. In virtually all ethical systems, the question of whether an action or principle is morally defensible will depend at least in part on what its consequences are. Learning economic theory may change people's view of consequences and hence their moral principles and conduct. Knowledge of economics may, for example, have contributed to the change in attitude toward charging interest on loans of money. Furthermore, generalizations about what people do will often influence what people think ought to be done. Even though what ought to be does not follow logically from what is, it often follows psychologically. Saying that human behavior can be modeled as if it were entirely self-interested unavoidably legitimizes and fosters self-interested behavior. Indeed, it seems that learning economics may make people more selfish (Marwell and Ames 1981; Frank, Gilovich, and Regan 1993). As we argued in Section 5.3, the terms of economic explanations – like Willie Sutton's account of why he robbed banks – can easily carry moral implications. Hence the moral commitments of economic agents are not only causes and effects of economic outcomes, institutions, and processes. They are also influenced by the way in which economists and other agents describe them and appraise them.

Learning economics has moral consequences because people's moral commitments are malleable. They are strongly influenced by the moral education provided by parents, churches, and schools, but they also change in response to pressures from peers and in response to the normative expectations of the many groups and institutions in which individuals find themselves. A firm with a well-deserved reputation for honesty and decency will not only have an easier time hiring honest and decent employees, it will also lead the employees it hires to become more honest and decent. The moral culture of the firm, the moral standards of the employees it hires, and the moral commitments of its customers, suppliers, and the community at large all interact to affect the productivity of the employees and the profitability of the firm.

The honesty, trust, and sense of fair play that help economies to function well are not givens that are fortunately abundant or unfortunately scarce. They are not comparable to geological formations or biological necessities. They grow or wither depending on the institutions within which people live and the shared understandings of those institutions. Their content varies widely from individual to individual and from society to society. Economists need to be concerned about nurturing these vital moral resources. Some moral principles may enable people to coordinate their behavior better than others, and some principles may spread more readily than others because they command more respect and emulation.

Economists need to think about the economic role of moral commitments, and they cannot do so intelligently without at least understanding – if not also appraising – the content of those commitments.

A.5 Conclusions

Economists are not only social engineers contributing to policy in the way that civil engineers contribute to policies concerning dams. Normative economics attempts to appraise policies, even if usually from a limited point of view, and evaluative thinking is in practice unavoidable in order to formulate well-defined questions for positive inquiry. The "standard view" of the relations between positive and normative economics does not deny any of this. It does not deny the existence or legitimacy of normative inquiry. Instead it insists on the possibility of purely positive inquiry and thus on the separability of positive and normative inquiry. Positive science is not, however, independent of all values. On the contrary, it is guided by the values of scientific inquiry; moreover, a plethora of specific values, including policy interests, play a large and legitimate role in determining which questions to ask and even possible leads to follow. In insisting that positive economics be value free, what is meant is that the positive economist's own evaluations of economic outcomes, institutions, and processes are not supposed to influence the answers given to the questions under investigation. The defender of the standard view does not deny that evaluations of economic outcomes sometimes do influence the results of positive inquiries. When this happens, the defender of the standard view sees a scientific failing.

Without directly challenging the standard view or attempting to measure the magnitude of such scientific failings in economics, we argued that the study of ethics can contribute to a good deal of positive economics. The reasons for this are that the moral commitments of economic agents are important causes and effects of significant economic phenomena, and that these commitments are influenced by the way that they are described and appraised by economic agents and by economists. Although understanding the moral commitments of agents and the dependencies among them and other phenomena does not require that economists themselves make ethical appraisals, an understanding of moral commitments requires a good grasp of ethics. As this book has shown, reflection on ethics has an important role to play in both positive and normative economics.

Glossary

altruism (Sections 5.1, 5.3, and 5.4) Altruism consists of action that aims at the good of someone else. It is not the same thing as morality.

Arrow's theorem (Sections 13.1 and 13.2) This theorem proves that there is no way of making social choices or social evaluations that satisfies a small number of apparently reasonable conditions.

backwards induction argument (Section 12.3) This argument purports to show that the uniquely rational strategy to follow in a finite iterated prisoners' dilemma is to defect on every move.

bargaining theory (Section 14.5) A branch of game theory in which individuals can improve upon an initial nonagreement point only if they can reach agreement on some distribution of the gains from cooperation.

Bayesian theory (Section 4.2) Bayesians offer an interpretation of probability as degree of belief, and they support the use of expected utility theory in circumstances of uncertainty.

capability (Section 8.5) A feature of a person that enables the person to carry out various activities ("functionings"). Two examples: literacy and the ability to see. Amartya Sen argues that the standard of living should be understood in terms of capabilities.

cardinal utility (Section 4.2) A utility representation in which comparisons of utility differences are not arbitrary.

commitment (Section 6.5) A person *P* has a commitment to perform action *A* if and only if *P* regards doing *A* as a duty and would perform *A* in some circumstances even if doing so conflicted with self-interested and sympathetic preferences.

completeness (Section 4.1) A condition on preferences. A person's preferences are complete if, for all objects of preference x and y, either the person prefers x to y or y to x or the person is indifferent.

consequentialism (Chapter 7) A structure of ethical theories in which the rightness of actions and policies depends upon the goodness of their results.

contractualism (Chapter 12) A range of views concerning justification and motivation for moral theories. The main idea of contractualism is that a moral theory is justified if it would be agreed to by rational individuals under the right circumstances.

cost–benefit analysis (Sections 9.3, 9.4, and 9.5) Cost–benefit analysis consists of techniques for determining how much those who would benefit from policies would be willing to pay to institute them as well as how much those who would be harmed by policies would have to be compensated in order to accept the policies being instituted. Although the comparison of these benefits and costs may not by itself decide policy questions, practitioners of cost–benefit analysis believe that such information helps.

deontological moral principles (Section 7.6) Deontological moral principles are nonconsequentialist principles like the Ten Commandments. In Samuel Scheffler's terminology, deontological ethical theories employ both "agent-centered prerogatives" (they sometimes *permit* agents to act in a way that does not maximize the good) and "agent-centered constraints" (they sometimes *prohibit* agents from acting so as to maximize the good).

efficiency (Chapter 9) In theoretical economics, efficiency is taken to be Pareto efficiency and thus a matter of preference satisfaction. *See* Pareto efficiency.

envy-free allocations (Section 11.2) An allocation is envy free if no agent prefers another's allocation to his or her own.

equity (Chapter 12) Economists use the term "equity" frequently as a synonym for "justice."

ethics Although some people treat ethics as more theoretical than morality, we take ethics and morality to be synonymous. *See* morality.

expected utility (Section 4.2) The expected utility of an action is the utility of the possible outcomes multiplied by their probabilities.

extended sympathy judgments (Section 7.2) A preference ranking of "extended alternatives" such as (Jill, y) – that is, being Jill and obtaining some alternative y. Extended sympathy judgments are supposed to provide a basis for making interpersonal comparisons of the extent to which people's preferences are satisfied.

externality (Sections 2.3 and 9.1) A consequence of an agent's action for which there is no market and hence no price. Those who cause negative externalities – for example, by polluting the air – need not take the costs imposed on others into account when making calculations of private economic benefit. Nor will someone who causes a positive externality (say, by building a lighthouse) be able to collect from all those who benefit from it.

fair allocations (Section 11.2) An allocation that is both Pareto optimal and envy free.

folk psychology (Chapter 4 and Section 5.2) The view that people's actions are to be explained by their beliefs and their desires. The explanations economists offer of individual choice behavior conform to folk psychology.

freedom (Section 10.1) Freedom is a relation among three things: (1) an agent, (2) obstacles of some kind, and (3) some of the agent's objectives. An agent is free in some regard when there are no obstacles of a particular kind preventing the agent from doing something. The main kinds of freedom concern (i) the absence of intentional interference from others, (ii) the range of alternatives open to one, and (iii) the extent to which one is autonomous or self-determining.

game theory (Chapter 14) A mathematical theory concerning interdependencies with applications in several fields. It arose as an attempt to understand strategic interactions among individuals, and its main applications still lie in this domain.

independence condition (Section 4.2) A condition on preferences. An agent's preferences satisfy the independence condition if the agent's preferences between two lotteries that differ only in one prize match the agent's preferences between the differing prizes.

independence of irrelevant alternatives (Section 13.1) Arrow's independence of irrelevant alternatives condition says that the social ranking of x and y should depend on nothing except individual rankings of x and y.

justice (Chapter 12) Justice involves respecting people's rights, giving people what they deserve, and treating people fairly. Most philosophers take justice to be the most important part of morality.

libertarianism (Section 10.6) A kind of moral theory that takes freedom to be the most important social value, emphasizes rights, and takes justice to be a matter of protecting rights.

morality Morality consists of judgments concerning what things are intrinsically good for people, of what actions are permissible, impermissible, and obligatory, and concerning the moral worth of people and their characters.

moral norms (Sections 6.2 and 6.3) Moral norms are distinguished from other norms by their subject matter (interpersonal interaction where significant benefits and harms are at stake), their weight (they typically override other considerations), and the sanctions – both internal (guilt) and external (blame) – attached to their violation.

normative vs. positive theories (Section A.2) Positive theories say what is, while normative theories say what ought to be. Positive theories are concerned with facts, while normative theories are concerned with values. The distinction is not a sharp one.

norms (Sections 6.2 and 6.3) Prescriptive rules regarding behavior that are shared among a group of people and that are partly sustained by the approval and disapproval of others.

ordinal utility (Section 4.1) An ordinal utility function represents the order of an individual's preferences and nothing more.

paradox of the Paretian libertarian (Section 13.4) Amartya Sen's paradox of the Paretian libertarian is the proof of an inconsistency between the Pareto principle and a principle of minimal liberty (given some other technical conditions).

Pareto improvement (Sections 5.2 and 9.1) A state of affairs S is a Pareto improvement over a state of affairs R if nobody prefers R to S and somebody prefers S to R.

Pareto optimality (or Pareto efficiency) (Sections 5.2 and 9.1) A state of affairs S is Pareto optimal (or Pareto efficient) if no states of affairs are Pareto improvements over S – in other words, a state of affairs in which it is not possible to satisfy anyone's preferences better without frustrating someone else's preferences.

Pareto principle (Sections 9.1 and 13.1) The strong Pareto principle states that if nobody prefers y to x and somebody prefers x to y, then x ought to be socially preferred to y. The weak Pareto principles states that if everybody prefers x to y, then x ought to be socially preferred to y.

perfect competition (Section 5.2) Perfect competition obtains when there are markets for all goods and services (and thus no externalities), no barriers to entry or exit from any market, and so many traders in every market that no one can influence prices. Perfect competition is also sometimes taken to imply there are no interdependencies among people's utility functions.

positive affine transformation (Section 4.2) $f'(x)$ is a positive affine transformation of $f(x)$ if and only if $f'(x) = af(x) + b$, where a and b are real numbers and $a > 0$.

positive vs. normative theories *See* normative vs. positive theories.

potential Pareto improvement (Section 9.3) State *S* is a potential Pareto improvement over state *R* if it would be possible for those whose preferences are better satisfied in *S* than in *R* to compensate those whose preferences are less well satisfied and so bring about a genuine Pareto improvement.

primary goods (Section 8.5) Primary goods are all-purpose means, such as income and education, that contribute to everyone's life plans. This is Rawls's notion, and he argues that primary goods are a better basis for the evaluation of institutions than are utilities.

prisoners' dilemma (Section 14.2) A game in which all the players have a single dominant strategy, and the result of each employing that dominant strategy is suboptimal.

probability (Section 4.3) Probabilities are real numbers between 0 and 1 that are attached to states of affairs or propositions and that satisfy the axioms of the mathematical theory of probability. There are currently two main interpretations. "Objective probabilities" may be regarded as the limit of the relative frequency of an occurrence, as 0.5 is the limit of the frequency with which an unbiased coin will land heads. "Subjective probabilities" are degrees of belief.

rationality (Part I) A controversial term. In its loosest sense, rational behavior is behavior that is based on good reasons. Economists define rational preferences as preferences that are complete and transitive, and they define rational choice as choice that it determined by rational preferences.

reasons (Sections 5.1 and 5.3) Propositions or the contents of mental states that justify actions or other mental states.

revealed preference theory (Section 4.1) The theory of revealed preference attempts to construct preferences out of choice behavior that satisfies consistency conditions. The hope is to avoid referring to preference as a subjective state responsible for choice.

rights (Sections 10.2–10.4) Rights involve freedoms for the right-holder and various duties for others. Legal and conventional rights are established by law and convention. To say that agents have a moral or natural right to *X* is to say that the agent ought to have a legal or a conventional right to *X*.

risk (Section 4.2) One is in a circumstance of risk when the outcome of action is not certain but all the possible outcomes and their probabilities are known.

Sen's paradox *See* paradox of the Paretian libertarian.

social choice theory (Chapter 13) The proof and interpretation of theorems concerning the social aggregation of preferences, judgments, and interests.

social welfare function (Section 13.1) A ranking of social states.

transitivity (Section 4.1) A condition on preferences. An agent's preferences are transitive if the agent prefers x to z whenever the agent prefers x to y and y to z for all objects of preference x, y, and z, and similarly for indifference.

uncertainty (Section 4.2) One is in a circumstance of uncertainty when some of the possible outcomes of an action or their probabilities are not known. One is usually in circumstances of uncertainty.

universal domain condition (Section 13.1) The universal domain condition states that a method of social evaluation or social decision making should work for every profile of individual preferences. This condition is common in social choice theories, including Arrow's theorem and Sen's paradox of the Paretian libertarian.

utilitarianism (Chapter 7) The best-known variety of consequentialism. According to utilitarianism, an action or policy is right if it results in no less happiness or no less preference satisfaction than any alternative.

utility (Section 4.1) A mathematical device for representing features of preference orderings. It is an index or indicator, not a subjective state like pleasure. Utility is not itself an object of preference or choice.

welfare (Section 5.2 and Chapter 8) A person's welfare is how good things are for the person. Economists often take welfare to be the satisfaction of preferences.

References

Ackerman, Bruce. *Social Justice in the Liberal State.* New Haven: Yale University Press, 1981.

Ackerman, Bruce and Anne Alstott. *The Stakeholder Society.* New Haven: Yale University Press, 1999.

Akerlof, George. "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics* 84 (1982): 488–500.

"Loyalty Filters." *American Economic Review* 73 (1983): 54–63.

"Gift Exchange and Efficiency Wage Theory: Four Views." *American Economic Review Proceedings* 74 (1984): 79–83.

Akerlof, George and Janet Yellen. "Introduction," in G. Akerlof and J. Yellen (eds.), *Efficiency Wage Models of the Labor Market.* Cambridge University Press, 1986, pp. 1–21.

Alexander, Sidney. "Social Evaluation through Notional Choice." *Quarterly Journal of Economics* 88 (1974): 597–624.

Allais, Maurice. "The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School" (1952), in Allais and Hagen (1979), pp. 27–145.

Allais, Maurice and Otto Hagen (eds.), *Expected Utility Hypotheses and the Allais Paradox.* Dordrecht: Reidel, 1979.

Anderson, Elizabeth. "What Is the Point of Equality?" *Ethics* 109 (1999): 287–337.

Arneson, Richard. "Equality and Equal Opportunity for Welfare." *Philosophical Studies* 56 (1989): 77–93.

"Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare." *Philosophy and Public Affairs* 19 (1990): 158–94.

"Luck Egalitarianism and Prioritarianism." *Ethics* 110 (2000): 339–49.

Arrow, Kenneth. *Social Choice and Individual Values.* New York: Wiley, 1951 (2nd ed., 1963).

"Values and Collective Decision Making" (1967); reprinted in Frank Hahn and Martin Hollis (eds.), *Philosophy and Economic Theory.* Oxford University Press, 1979, pp. 110–26.

"Gifts and Exchanges." *Philosophy and Public Affairs* 1 (1972): 343–62.

"Some Ordinalist-Utilitarian Notes on Rawls' Theory of Justice." *Journal of Philosophy* 70 (1973): 246–63.

The Limits of Organization. New York: Norton, 1974.
"Extended Sympathy and the Possibility of Social Choice." *Philosophia* 7 (1978): 223–37.
"A Note on Freedom and Flexibility," in Kaushik Basu, Prasanta Pattanaik, and Kotaro Suzumura (eds.), *Choice, Welfare and Development: A Festschrift in Honour of Amartya K. Sen*. Oxford: Clarendon Press, 1995, pp. 7–16.
Arrow, Kenneth and Frank Hahn. *General Competitive Analysis.* San Francisco: Holden-Day, 1971.
Arrow, Kenneth, Amartya Sen, and Kotaro Suzumura (eds.). *Social Choice Re-Examined: Proceedings of the IEA Conference Held at Schloss Herstein, Berndorf, Near Vienna, Austria.* New York: St. Martin's Press, vol. 1 (1996) and vol. 2 (1997).
Handbook of Social Choice and Welfare. Amsterdam: Elsevier, 2002.
Atkinson, Anthony B. "On the Measurement of Inequality." *Journal of Economic Theory* 2 (1970): 244–63.
Axelrod, Robert. *The Evolution of Cooperation.* New York: Basic Books, 1984.
Bacharach, Michael. *Beyond Individual Choice: Teams and Frames in Game Theory,* ed. Natalie Gold and Robert Sugden. Princeton University Press, 2005.
Baker, C. "The Ideology of the Economic Analysis of Law." *Philosophy and Public Affairs* 5 (1975): 3–48.
Barbera, Salvador, Peter Hammond, and Christian Seidl (eds.). *Handbook of Utility Theory.* Dordrecht: Kluwer, vol. 1 (1998) and vol. 2 (1999).
Barry, Brian. *The Liberal Theory of Justice.* Oxford: Clarendon Press, 1973.
Theories of Justice (*Treatise on Social Justice,* vol. 1). Berkeley: University of California Press, 1989.
Basu, Kaushik. "The Right to Give Up Rights." *Economica* 51 (1984): 413–22.
Bator, Francis. "The Simple Analytics of Welfare Maximization." *American Economic Review* 47 (1957): 22–59.
Batson, C. Daniel. "Experimental Tests for the Existence of Altruism," in David Hull, Micky Forbes, and Kathleen Okruklik (eds.), *PSA 1992,* vol. 2. East Lansing: Philosophy of Science Association, 1993, pp. 69–78.
Baumol, William. *Superfairness: Applications and Theory.* Cambridge: MIT Press, 1986.
Becker, Gary. *The Economic Approach to Human Behavior.* University of Chicago Press, 1976.
A Treatise on the Family. Cambridge, MA: Harvard University Press, 1981.
Beitz, Charles (ed.). *International Ethics.* Princeton University Press, 1985.
Benn, Stanley. "Egalitarianism and the Equal Consideration of Interests," in Pennock and Chapman (1967), pp. 61–78.
Ben-Ner, Avner and Louis Putterman (eds.). *Economics, Values, and Organization.* Cambridge University Press, 1998.
Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation* (1789), ed. W. Harrison. Oxford: Blackwell, 1967.
Berg, J., J. Dickhaut, and J. O'Brien. "Preference Reversal and Arbitrage," in Vernon Smith (ed.), *Research in Experimental Economics,* vol. 3. Greenwich: JAI Press, 1985, pp. 31–72.
Bergson, Abram. "A Reformulation of Certain Aspects of Welfare Economics." *Quarterly Journal of Economics* 52 (1938): 30–4.

Berlin, Isaiah. "Two Concepts of Liberty," in I. Berlin, *Four Essays on Liberty.* Oxford University Press, 1969, pp. 118–72.

Bicchieri, Cristina. "Self-Refuting Theories of Strategic Interaction: A Paradox of Common Knowledge." *Erkenntnis* 29 (1988): 69–85.

"Paradoxes of Rationality." *Midwest Studies in Philosophy* 15 (1990): 65–79.

"Covenants without Swords: Group Identity, Norms and Communication in Social Dilemmas." *Rationality and Society* 14 (2002): 187–222.

Binmore, Ken. "Modeling Rational Players: Part I." *Economics and Philosophy* 3 (1987): 179–214.

"Modeling Rational Players: Part II." *Economics and Philosophy* 4 (1988): 9–56.

"Evolution and Utilitarianism." *Constitutional Political Economy* 1 (1990): 1–26.

Fun and Games. New York: D.C. Heath, 1992.

Playing Fair: Game Theory and the Social Contract. Cambridge: MIT Press, 1994.

Just Playing. Game Theory and the Social Contract, vol. II, *Economic Learning and Social Evolution.* Cambridge: MIT Press, 1998.

Binmore, Ken, J. Swierzbinski, S. Hsu, and C. Proulx. "Focal Points and Bargaining." *International Journal of Game Theory* 22 (1993): 381–409.

Blackorby, Charles and David Donaldson. "Cash versus Kind, Self-Selection, and Efficient Transfers." *American Economic Review* 78 (1988): 691–700.

Blau, Julian H. "Liberal Values and Independence." *Review of Economic Studies* 42 (1975): 395–401.

Blinder, Alan and Don Choi. "A Shred of Evidence on Theories of Wage Stickiness." *Quarterly Journal of Economics* 105 (1990): 1003–15.

Block, Walter. "A Free Market in Roads," in Machan (1982), pp. 164–83.

Blount, Sally. "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences." *Organizational Behavior and Human Decision Processes* 63 (1995): 131–44.

Bonnano, Giacomo. "The Logic of Rational Play in Games of Perfect Information." *Economics and Philosophy* 7 (1991): 37–65.

Bossert, W., P. K. Pattanaik, and Y. Xu. "Similarity of Options and the Measurement of Diversity." *Journal of Theoretical Politics* 15 (2003): 405–21.

Boulding, Kenneth. *The Economy of Love and Fear; A Preface to Grants Economics.* Belmont: Wadsworth, 1978.

Bowles, Samuel. "The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian and Marxian Models." *American Economic Review* 75 (1985): 16–36.

Bowles, Samuel and Herbert Gintis. "A Political and Economic Case for the Democratic Enterprise." *Economics and Philosophy* 9 (1993): 75–100.

Bowles, Samuel, Herbert Gintis, Erik Olin Wright, and Harry Brighouse (eds.). *Recasting Egalitarianism: New Rules for Communities, States and Markets.* London: Verso, 1999.

Braithwaite, Richard. *The Theory of Games as a Tool for the Moral Philosopher.* Cambridge University Press, 1955.

Brandt, Richard. *A Theory of the Right and the Good.* Oxford University Press, 1979.

Braybrooke, David. *Meeting Needs.* Princeton University Press, 1987.

Brennan, Geoffrey and James Buchanan. *The Reason of Rules: Constitutional Political Economy.* Cambridge University Press, 1985.

Brennan, Geoffrey and Phillip Pettit. *The Economy of Esteem: An Essay on Civil and Political Society.* Oxford University Press, 2004.

Brennan, Timothy. "A Methodological Assessment of Multiple Utility Frameworks." *Economics and Philosophy* 5 (1989): 189–208.

Brighouse, Harry. *School Choice and Social Justice.* Oxford University Press, 2000.

Broome, John. "Should Social Preferences Be Consistent?" *Economics and Philosophy* 5 (1989): 7–18.

"Utility." *Economics and Philosophy* 7 (1991a): 1–12.

Weighing Goods. Oxford: Blackwell, 1991b.

Counting the Cost of Global Warming. Isle of Stroud: White Horse Press, 1992.

Weighing Lives. Oxford University Press, 2004.

Buchanan, Allen E. *Ethics, Efficiency, and the Market.* Totowa: Rowman & Allanheld, 1985.

Buchanan, James. *The Limits of Liberty: Between Anarchy and Leviathan.* University of Chicago Press, 1975.

Calabresi, Guido and Douglas Melamed. "Property Rules, Liability Rules, and Inalienability: One View of the Cathedral." *Harvard Law Review* 85 (1972): 1089–1128.

Camerer, Colin. *Behavioral Game Theory: Experiments on Strategic Interaction.* Princeton University Press, 2003.

Carter, Ian. *A Measure of Freedom.* Oxford University Press, 1999.

Christman, John (ed.). *The Inner Citadel: Essays on Individual Autonomy.* Oxford University Press, 1989.

Chu, Y. and R. Chu. "The Subsidence of Preference Reversals in Simplified and Marketlike Experimental Settings: A Note." *American Economic Review* 80 (1990): 902–11.

Chubb, John E. and Terry M. Moe. *Politics, Markets and America's Schools.* Washington, DC: Brookings, 1990.

Churchland, Paul M. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge: MIT Press, 1989.

Clayton, Matthew and Andrew Williams. *The Ideal of Equality.* New York: St. Martin's Press, 2000.

Coase, Ronald. "The Problem of Social Cost." *Journal of Law and Economics* 3 (1960): 1–30.

Cohen, G. A. "Functional Explanation, Consequence Explanation, and Marxism." *Inquiry* 25 (1982): 27–56.

"On the Currency of Egalitarian Justice." *Ethics* 99 (1989): 906–44.

"Equality of What? On Welfare, Goods, and Capabilities," in Nussbaum and Sen (1993), pp. 9–30.

Coleman, James. *Foundations of Social Theory.* Cambridge, MA: Harvard University Press, 1990.

Coleman, Jules. "Economics and the Law: A Critical Review of the Foundations of the Economic Approach to Law." *Ethics* 94 (1984): 649–79.

Coleman, Jules L. and Jody Kraus. "Rethinking the Theory of Legal Rights." *Yale Law Review* 95 (1986): 1335–71.

Collard, David. *Altruism and Economy: A Study in Non-Selfish Economics.* Oxford University Press, 1978.

Cooter, Robert. "Rawls's Lexical Orderings Are Good Economics." *Economics and Philosophy* 5 (1989): 47–54.

Cowen, Tyler (ed.). *The Theory of Market Failure: A Critical Examination.* Fairfax: George Mason University Press, 1988.

Daniels, Norman. *Reading Rawls.* New York: Basic Books, 1976.

Just Health Care. Cambridge University Press, 1985.

Darwall, Stephen. "Respect and the Second-Person Standpoint." *Proceedings and Addresses of the American Philosophical Association* 78 (2004): 43–55.

d'Aspremont, Claude and Louis Gevers. "Equity and the Informational Basis of Collective Choice." *Review of Economic Studies* 44 (1977): 199–209.

Davidson, Donald. "Actions, Reasons, and Causes" (1963); reprinted in Davidson (1980), pp. 3–20.

Essays on Actions and Events. Oxford University Press, 1980.

"Judging Interpersonal Interests," in Jon Elster and Aanund Hylland (eds.), *Foundations of Social Choice Theory.* Cambridge University Press, 1986, pp. 195–211.

Davies, J. C. *Human Nature in Politics.* New York: Wiley, 1963.

Dawes, Robyn, Alphons van de Kragt, and John Orbell. "Cooperation for the Benefit of Us – Not Me, or My Conscience," in Jane Mansbridge (ed.), *Beyond Self-Interest.* University of Chicago Press, 1990, pp. 97–111.

Debreu, Gerard. *Theory of Value.* New York: Wiley, 1959.

Demsetz, Harold. "The Exchange and Enforcement of Property Rights" (1964); reprinted in Cowen (1988), pp. 127–45.

"Toward a Theory of Property Rights." *American Economic Review* 57 (1967): 347–59.

Dennett, Daniel. *The Intentional Stance.* Cambridge: MIT Press, 1987.

De Vroey, Michel. *Involuntary Unemployment: The Elusive Quest for a Theory.* London: Routledge, 2004.

Dretske, Fred. *Explaining Behavior: Reasons in a World of Causes.* Cambridge: MIT Press, 1988.

Drèze, Jean and Nicholas Stern. "Shadow Prices and Markets: Policy Reform, Shadow Prices, and Market Prices" (1990); reprinted in Layard and Glaister (1994), pp. 59–99.

Dworkin, Gerald. "Paternalism," in Richard Wasserstrom (ed.), *Morality and the Law.* Belmont: Wadsworth, 1971, pp. 107–36.

The Theory and Practice of Autonomy. Cambridge University Press, 1988.

Dworkin, Ronald. *Taking Rights Seriously.* Cambridge, MA: Harvard University Press, 1977.

"What is Equality? Part 1: Equality of Welfare." *Philosophy and Public Affairs* 10 (1981a): 185–246.

"What is Equality? Part 2: Equality of Resources." *Philosophy and Public Affairs* 10 (1981b): 283–345.

"What is Equality? Part 4: Political Equality." *University of San Francisco Law Review* 22 (1988): 1–28.

"Foundations of Liberal Equality," in Grethe Peterson (ed.), *The Tanner Lectures on Human Values, XI, 1990.* Salt Lake City: University of Utah Press, 1990, pp. 1–119.

Sovereign Virtue: The Theory and Practice of Equality. Cambridge, MA: Harvard University Press, 2000.
"*Sovereign Virtue* Revisited." *Ethics* 113 (2002): 106–43.
Dworkin, Ronald, Gordon Bermant, and Peter Brown (eds.). *Markets and Morals.* New York: HarperCollins, 1977.
Eells, Ellery. *Rational Decision and Causality.* Cambridge University Press, 1982.
Ellsberg, Daniel. "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics* 75 (1961): 643–99.
Elster, Jon. *Sour Grapes: Studies in the Subversion of Rationality.* Cambridge University Press, 1983.
The Cement of Society: A Study of Social Order. Cambridge University Press, 1989a.
"Social Norms and Economic Theory." *Journal of Economic Perspectives* 3 (1989b): 99–117.
Elster, Jon and Aanund Hylland (eds.). *Foundations of Social Choice Theory.* Cambridge University Press, 1986.
Elster, Jon and John Roemer (eds.). *Interpersonal Comparisons of Well-Being.* Cambridge University Press, 1991.
Etzioni, Amitai. *The Moral Dimension. Toward a New Economics.* New York: Macmillan, 1988.
Feinberg, Joel. *Social Philosophy.* Englewood Cliffs: Prentice-Hall, 1973.
Harm to Self. Oxford University Press, 1986.
Fleischacker, Samuel. *On Adam Smith's Wealth of Nations: A Philosophical Companion.* Princeton University Press, 2004.
Fleming, Marcus. "A Cardinal Concept of Welfare." *Quarterly Journal of Economics* 66 (1952): 366–84.
Fleurbaey, Marc. "Equal Opportunity or Equal Social Outcome." *Economics and Philosophy* 10 (1995): 25–55.
"Equality of Resources Revisited." *Ethics* 113 (2002): 82–105.
"Equality versus Priority," in Daniel Wikler and Christopher Murray (eds.), *From Fairness to Goodness.* Geneva: World Health Organization, 2006.
Fleurbaey, Marc and Martin van Hees. "On Rights in Game Forms." *Synthese* 123 (2000): 295–326.
Fleurbaey, Marc and François Maniquet. "Fair Social Orderings When Agents Have Unequal Production Skills." *Social Choice and Welfare* 24 (2005): 93–127.
Fleurbaey, Marc, Kotaro Suzumura, and Koichi Tadenuma. "The Informational Basis of the Theory of Fair Allocation." *Social Choice and Welfare* 24 (2005): 311–41.
Foley, Duncan. "Resource Allocation and the Public Sector." *Yale Economic Essays* (Spring 1967): 45–102.
Frank, Robert. *Choosing the Right Pond: Human Behavior and the Quest for Status.* Oxford University Press, 1987.
Passions within Reason: The Strategic Role of the Emotions. New York: Norton, 1988.
"The Strategic Role of the Emotions: Reconciling Over- and Undersocialized Accounts of Behavior." *Rationality and Society* 5 (1993): 160–84.
What Price the Moral High Ground? Princeton University Press, 2003.
Frank, Robert and Philip Cook. *The Winner-Take-All Society.* New York: Free Press, 1995.

Frank, Robert, Thomas Gilovich, and Dennis Regan. "Does Studying Economics Inhibit Cooperation?" *Journal of Economic Perspectives* 7 (1993): 159–72.

Frankfurt, Harry. "Freedom of Will and the Concept of a Person." *Journal of Philosophy* 68 (1971): 5–20.

"Equality as a Moral Ideal." *Ethics* 98 (1987): 21–43.

Friedman, Milton. "The Methodology of Positive Economics," in M. Friedman, *Essays in Positive Economics.* University of Chicago Press, 1953, pp. 3–43.

Capitalism and Freedom. University of Chicago Press, 1962.

Friedman, Milton and Rose Friedman. *Free to Choose.* New York: Harcourt Brace Jovanovich, 1980.

Frohlich, Norman, Thomas Hunt, Joe Oppenheimer, and R. Harrison Wagner. "Individual Contributions for Collective Goods: Alternative Models." *Journal of Conflict Resolution* 19 (1975): 310–29.

Gaertner, Wulf, Prasanta Pattanaik, and Kotaro Suzumura. "Individual Rights Revisited." *Economica* 59 (1992): 161–77.

Gambetta, Diego (ed.). *Trust: Making and Breaking Cooperative Relations.* Oxford: Blackwell, 1988.

Gärdenfors, Peter. "Rights, Games and Social Choice." *Nous* 15 (1981): 341–56.

Gärdenfors, Peter and Nils-Eric Sahlin (eds.). *Decision, Probability and Utility: Selected Readings.* Cambridge University Press, 1988.

Gauthier, David. *Morals by Agreement.* Oxford University Press, 1986.

Moral Dealing: Contract, Ethics, and Reason. Ithaca: Cornell University Press, 1990.

Gibbard, Allan. "A Pareto-Consistent Libertarian Claim." *Journal of Economic Theory* 7 (1974): 388–410.

"Interpersonal Comparisons: Preference, Good, and the Intrinsic Reward of a Life," in Elster and Hylland (1986), pp. 165–94.

Wise Choices, Apt Feelings: A Theory of Normative Judgment. Cambridge, MA: Harvard University Press, 1990.

Gibbard, Allan and William Harper. "Counterfactuals and Two Kinds of Expected Utility," in C. Hook, J. Leach, and E. McClennen (eds.), *Foundations and Applications of Decision Theory,* vol. 1. Dordrecht: Reidel, 1978, pp. 125–62.

Gigerenzer, Gerd. *Adaptive Thinking: Rationality in the Real World.* Oxford University Press, 2000.

Gilbert, Margaret. "Walking Together: A Paradigmatic Social Phenomenon," in Peter French, Theodore Uehling, and Howard Wettstein (eds.), *Midwest Studies in Philosophy,* vol. 15, *Philosophy of the Human Sciences.* University of Notre Dame Press, 1990, pp. 1–14.

Glover, Jonathan. *Causing Death and Saving Lives.* London: Penguin, 1990.

Goldfarb, Robert and William Griffith. "Amending the Economist's Rational Egoist Model to Include Moral Values and Norms: Part 1, The Problem; and Part 2, Alternative Solutions," in K. Koford and J. Miller (eds.), *Social Norms and Economic Institutions.* Ann Arbor: University of Michigan Press, 1991, pp. 39–84.

Goodin, Robert. "Laundering Preferences," in Elster and Hylland (1986), pp. 75–101.

Graaff, Jan de Van. *Theoretical Welfare Economics.* Cambridge University Press, 1957.

Grether, David and Charles Plott. "Economic Theory of Choice and the Preference Reversal Phenomenon." *American Economic Review* 69 (1979): 623–38.

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

"Against the Taste Model," in Elster and Roemer (1991), pp. 45–69.
Value Judgment: Improving our Ethical Beliefs. Oxford University Press, 1996.
Gutmann, Amy. *Democratic Education.* Princeton University Press, 1987.
"Why Should Schools Care about Civic Education?" in Lorraine M. McDonnell, P. Michael Timpane, and Roger Benjamin (eds.), *Rediscovering the Democratic Purposes of Education.* Lawrence: University Press of Kansas, 2000, pp. 73–90.
Hamlin, Alan. *Ethics, Economics, and the State.* New York: St. Martin's Press, 1986.
Hammond, Peter. "Equity, Arrow's Conditions and Rawls' Difference Principle." *Econometrica* 44 (1976): 793–804.
"Ex-Post Optimality as a Dynamically Consistent Objective for Collective Choice under Uncertainty," in Prasanta Pattanaik and Maurice Salles (eds.), *Social Choice and Welfare.* Amsterdam: North-Holland, 1983.
Hampton, Jean. *Hobbes and the Social Contract Tradition.* Cambridge University Press, 1986.
"Free-Rider Problems in the Production of Collective Goods." *Economics and Philosophy* 3 (1987): 245–73.
The Authority of Reason. Cambridge University Press, 1998.
Harberger, Arnold C. "On the Use of Distributional Weights in Social Cost–Benefit Analysis." *Journal of Political Economy* 86 (1978): S87–S120.
Hardin, Russell. *Collective Action.* Baltimore: Johns Hopkins University Press, 1982.
Morality within the Limits of Reason. University of Chicago Press, 1988.
One for All: The Logic of Group Conflict. Princeton University Press, 1995.
Hare, Richard. *Moral Thinking: Its Levels, Method, and Point.* Oxford University Press, 1981.
Harper, William. "Ratifiability and Refinements in Two-Person Non-Cooperative Games," in Michael Bacharach and Susan Hurley (eds.), *Foundations of Game Theory: Issues and Advances.* Oxford: Blackwell, 1991, pp. 263–93.
Harris, Richard and Nancy Olewiler. "The Welfare Economics of Ex Post Optimality." *Economica* 46 (1979): 137–47.
Harsanyi, John. "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63 (1955): 309–21.
"Morality and the Theory of Rational Behavior" (1977a); reprinted in Sen and Williams (1982), pp. 39–62.
Rational Behavior and Bargaining Equilibrium in Games and Social Situations. Cambridge University Press, 1977b.
Hart, Herbert L. A. "Are There Any Natural Rights?" (1955); reprinted in Waldron (1984), pp. 77–90.
Hausman, Daniel. *The Inexact and Separate Science of Economics.* Cambridge University Press, 1992.
"The Impossibility of Interpersonal Utility Comparisons." *Mind* 104 (1995): 473–90.
"Sympathy, Commitment and Preference." *Economics and Philosophy* 21 (2005): 33–50.
Hausman, Daniel and Julian LeGrand. "Incentives and Health Policy: Primary and Secondary Care in the British National Health Service." *Social Science and Medicine* 49 (1999): 1299–1307.

Hausman, Daniel and Michael McPherson. "Taking Ethics Seriously: Economics and Contemporary Moral Philosophy." *Journal of Economic Literature* 31 (1993): 671–731.

"Preference, Belief, and Welfare." *American Economic Review Papers and Proceedings* 84 (1994): 396–400.

Hayek, Friedrich von. *The Constitution of Liberty.* University of Chicago Press, 1960.

"The Moral Element in Free Enterprise," in F. von Hayek, *Studies in Philosophy, Politics and Economics.* University of Chicago Press, 1967, pp. 229–36.

The Mirage of Social Justice. University of Chicago Press, 1976.

Herstein, Israel and John W. Milnor. "An Axiomatic Approach to Measurable Utility." *Econometrica* 21 (1953): 291–7.

Hicks, John. "The Foundations of Welfare Economics." *Economic Journal* 49 (1939): 696–712.

Hirsch, Fred. *The Social Limits to Growth.* Cambridge, MA: Harvard University Press, 1976.

Hirschman, Albert. *Shifting Involvements: Private Interest and Public Action.* Princeton University Press, 1982.

"Against Parsimony: Three Easy Ways of Complicating Some Categories of Economic Discourse." *Economics and Philosophy* 1 (1985): 7–22.

Hobbes, Thomas. *Leviathan, Or the Matter, Forme and Power of a Commonwealth Ecclesiasticall and Civil* (1651). New York: Collier, 1962.

Hochman, Harold and James Rodgers. "Pareto Optimal Redistribution." *American Economic Review* 59 (1969): 524–57.

Hohfeld, Wesley. *Fundamental Legal Conceptions.* New Haven: Yale University Press, 1923.

Hollis, Martin and Robert Sugden. "Rationality in Action." *Mind* 102 (1993): 1–35.

Hooker, Brad. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality.* Oxford University Press, 2001.

Hooker, Brad, Elinor Mason, and Dale Miller (eds.). *Morality, Rules, and Consequences.* Totowa: Rowman & Littlefield, 2000.

Hornstein, Harvey, Elisha Fisch, and Michael Holmes. "Influence of a Model's Feelings about His Behavior and His Relevance as a Comparison Other on Observers' Helping Behavior." *Journal of Personality and Social Psychology* 10 (1968): 220–6.

Hume, David. *A Treatise of Human Nature* (1738). Oxford: Clarendon Press, 1966.

An Inquiry Concerning Human Understanding (1748). Indianapolis: Bobbs-Merrill, 1955.

Hurley, Susan L. *Natural Reasons: Personality and Polity.* Oxford University Press, 1989.

Jones, P. and R. Sugden. "Evaluating Choice." *International Review of Law and Economics* 2 (1982): 47–65.

Kagan, Shelly. *The Limits of Morality.* Oxford University Press, 1989.

Normative Ethics. New York: Perseus, 1997.

Kahneman, Daniel, Paul Slovic, and Amos Tversky (eds.). *Judgement under Uncertainty: Heuristics and Biases.* Cambridge University Press, 1982.

Kalai, Ehud and Meir Smorodinsky. "Other Solutions to Nash's Bargaining Problem." *Econometrica* 43 (1975): 513–18.

Kaldor, Nicholas. "Welfare Propositions of Economics and Interpersonal Comparisons of Utility." *Economic Journal* 49 (1939): 549–52.

Kamm, Frances M. "Non-consequentialism, the Person as an End-in-Itself, and the Significance of Status." *Philosophy and Public Affairs* 21 (1992): 354–89.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals* (1785), tr. H. Paton. New York: Harper & Row, 1948.

Kavka, Gregory. *Hobbesian Moral and Political Theory.* Princeton University Press, 1986.

Kelman, Steven. *What Price Incentives?* Boston: Auburn House, 1981.

"A Case for In-Kind Transfers." *Economics and Philosophy* 2 (1986): 53–74.

Knight, Frank. *Risk, Uncertainty and Profit* (1921). University of Chicago Press, 1971.

"Economics and Human Action," in F. Knight, *The Ethics of Competition, and Other Essays.* New York: Harper, 1935.

Kolm, Serge-Christophe. *Justice et Équité.* Paris: Éditions du Centre National de la Recherche Scientifique, 1972; tr. by Harold See as *Justice and Equity.* Cambridge: MIT Press, 2002.

Modern Theories of Justice. Cambridge: MIT Press, 1997.

Koopmans, Tjalling. *Three Essays on the State of Economic Science.* New York: McGraw-Hill, 1957.

Kraus, Jody and Jules Coleman. "Morality and the Theory of Rational Choice." *Ethics* 97 (1987): 715–49.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory* 27 (1982): 245–52.

Kreps, David and Robert Wilson. "Sequential Equilibria." *Econometrica* 50 (1982): 863–94.

Kukathas, Chandran and Philip Pettit. *A Theory of Justice and Its Critics.* Oxford: Blackwell, 1990.

Kuran, Timur. *Private Truths, Public Lies: The Social Consequences of Preference Falsification.* Cambridge, MA: Harvard University Press, 1995.

"Ethnic Norms and Their Transformation through Reputational Cascades." *Journal of Legal Studies* 27 (1998): 623–59.

Kymlicka, Will. "Rethinking the Family." *Philosophy and Public Affairs* 20 (1991): 77–97.

Laffont, Jean-Jacques. *Fundamentals of Public Economics,* tr. John Bonin and Hélène Bonin. Cambridge: MIT Press, 1988.

Lancaster, Kelvin. "A New Approach to Consumer Theory." *Journal of Political Economy* 74 (1966): 132–57.

Laslier, Jean-François, Marc Fleurbaey, Nicolas Gravel, and Alan Trannoy. *Freedom in Economics: New Perspectives in Normative Analysis.* London: Routledge, 1998.

Layard, Richard and Stephen Glaister (eds.). *Cost–Benefit Analysis,* 2nd ed. Cambridge University Press, 1994.

Lerner, Abba. "Consumption-Loan Interest and Money." *Journal of Political Economy* 67 (1959a): 512–18.

"Rejoinder." *Journal of Political Economy* 67 (1959b): 523–5.

Levi, Isaac. *The Enterprise of Knowledge.* Cambridge: MIT Press, 1980.

"The Paradoxes of Allais and Ellsberg." *Economics and Philosophy* 2 (1986): 23–53.

"Pareto Unanimity and Consensus." *Journal of Philosophy* 89 (1990): 481–92.

Lewis, David. *Convention*. Cambridge, MA: Harvard University Press, 1969.

"Causal Decision Theory." *Australasian Journal of Philosophy* 59 (1981): 5–30.

Lichtenstein, Sarah and Paul Slovic. "Reversals of Preference between Bids and Choices in Gambling Decisions." *Journal of Experimental Psychology* 89 (1971): 46–55.

"Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas." *Journal of Experimental Psychology* 101 (1973): 16–20.

List, Christian and Philip Pettit. "Aggregating Sets of Judgments: An Impossibility Result." *Economics and Philosophy* 18 (2002): 89–110.

"Aggregating Sets of Judgments: Two Impossibility Results Compared." *Synthese* 140 (2004): 207–35.

Little, Ian. *A Critique of Welfare Economics,* 2nd ed. Oxford University Press, 1957.

Locke, John. *Second Treatise on Government* (1690), in Peter Laslett (ed.), *Two Treatises of Government: A Critical Edition,* 2nd ed. Cambridge University Press, 1967.

Lomasky, Loren. *Persons, Rights and the Moral Community.* Oxford University Press, 1987.

Loomes, Graham and Robert Sugden. "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty." *Economic Journal* 92 (1982): 805–24.

Lucas, Robert E. "Unemployment Policy." *American Economic Review* 68 (1978): 353–7.

Luce, Robert Duncan and Howard Raiffa. *Games and Decisions.* New York: Wiley, 1957.

Lukes, Steven. *Essays in Social Theory.* New York: Columbia University Press, 1977.

MacCallum, Gerald. "Negative and Positive Freedom." *Philosophical Review* 76 (1967): 312–34.

McClennen, Edward. *Rationality and Dynamic Choice: Foundational Explorations.* Cambridge University Press, 1990.

McDowell, John. "Are Moral Judgments Hypothetical Imperatives?" *Proceedings of the Aristotlian Society Supplementary Volume* (1978): 13–29.

Machan, Tibor (ed.). *The Libertarian Reader.* Totowa: Rowman & Littlefield, 1982.

Machina, Mark. "Choice under Uncertainty: Problems Solved and Unsolved." *Journal of Economic Perspectives* 1 (1987): 121–54.

Machlup, Fritz. "Positive and Normative Economics," in Robert Heilbroner (ed.), *Economic Means and Social Ends: Essays in Political Economics.* Englewood Cliffs: Prentice-Hall, 1969, pp. 99–124.

MacKay, Alfred. *Arrow's Theorem: The Paradox of Social Choice. A Case Study in the Philosophy of Economics.* New Haven: Yale University Press, 1980.

"Extended Sympathy and Interpersonal Utility Comparisons." *Journal of Philosophy* 83 (1986): 305–22.

McKean, Roland N. "Economics of Trust, Altruism, and Corporate Responsibility," in Edmund Phelps (ed.), *Altruism, Morality and Economic Theory.* New York: Russell Sage Foundation, 1975, pp. 29–44.

McPherson, Michael. "Mill's Moral Theory and the Problem of Preference Change." *Ethics* 92 (1982): 252–73.

"Efficiency and Liberty in the Productive Enterprise: Recent Work in the Economics of Work Organization." *Philosophy and Public Affairs* 12 (1983a): 354–68.

"Want Formation, Morality, and Some 'Interpretive' Aspects of Economic Inquiry," in Norma Haan, Robert N. Bellah, Paul Rabinow, and William M. Sullivan (eds.), *Social Science as Moral Inquiry.* New York: Columbia University Press, 1983b, pp. 96–124.

Mansbridge, Jane (ed.). *Beyond Self-Interest.* University of Chicago Press, 1990.

Margolis, Howard. *Selfishness, Altruism and Rationality.* Cambridge University Press, 1982.

Marmot, Michael. *The Status Syndrome: How Your Social Standing Directly Affects Your Health and Life Expectancy.* London: Bloomsbury, 2004.

Martin, Rex. *A System of Rights.* Oxford: Clarendon Press, 1993.

Marwell, Gerald and Ruth Ames. "Economists Free Ride. Does Anyone Else? Experiments on the Provision of Public Goods. IV." *Journal of Public Economics* 15 (1981): 295–310.

Mason, Andrew (ed.). *Ideals of Equality.* Oxford: Blackwell, 1998.

Maynard Smith, J. *Evolution and the Theory of Games.* Cambridge University Press, 1982.

Meade, James. *Equality, Efficiency and the Ownership of Property.* Cambridge, MA: Harvard University Press, 1965.

Meckling, William. "An Exact Consumption-Loan Model of Interest: A Comment." *Journal of Political Economy* 68 (1960a): 72–6.

"Rejoinder." *Journal of Political Economy* 68 (1960b): 83–4.

Melden, A. I. *Free Action.* London: Routledge, 1961.

Mill, John Stuart. *On Liberty* (1859), ed. Currin V. Shields. New York: Macmillan, 1985.

Utilitarianism (1863); reprinted in Marshall Cohen (ed.), *The Philosophy of John Stuart Mill.* New York: Modern Library, 1961, pp. 321–98.

Miller, David. "Arguments for Equality." *Midwest Studies in Philosophy* 7 (1982): 73–88.

Miller, David and Michael Walzer. *Pluralism, Justice, and Equality.* Oxford University Press, 1995.

Mishan, E. *Cost–Benefit Analysis: An Introduction.* New York: Praeger, 1971.

An Introduction to Normative Economics. Oxford University Press, 1981.

Mongin, Philippe. "Consistent Bayesian Aggregation." *Journal of Economic Theory* 66 (1995): 313–51.

"The Impartial Observer Theorem of Social Ethics." *Economics and Philosophy* 17 (2001): 147–79.

"Value Judgments and Value Neutrality in Economics." *Economica* 72 (2006).

Mongin, Philippe and Claude d'Aspremont. "Utility Theory and Ethics," in Barbera et al. (1998), pp. 371–481.

Moulin, Hervé and William Thomson. "Axiomatic Analysis of Resource Allocation Problems," in Arrow et al. (1997), pp. 101–20.

Mueller, Dennis. "The Corporation and the Economist." *International Journal of Industrial Organization* 10 (1992): 147–70.

Murphy, Liam and Thomas Nagel. *The Myth of Ownership: Taxes and Justice.* Oxford University Press, 2004.

Musgrave, Richard. "Maximin, Uncertainty, and the Leisure Trade-Off." *Quarterly Journal of Economics* 88 (1974): 625–42.

Nagel, Thomas. *The Possibility of Altruism.* Oxford: Clarendon Press, 1970.

"Libertarianism without Foundations" (1981); reprinted in Jeffrey Paul (ed.), *Reading Nozick: Essays on Anarchy, State and Utopia*. Oxford: Blackwell, 1983, pp. 191–205.

The View from Nowhere. Oxford University Press, 1986.

Narveson, Jan. *The Libertarian Idea*. Philadelphia: Temple University Press, 1988.

Nash, John. "The Bargaining Problem." *Econometrica* 18 (1950): 155–62.

"Non-Cooperative Games." *Annals of Mathematics* 54 (1951): 286–95.

Nehring, Klaus and Clemens Puppe. "A Theory of Diversity." *Econometrica* 70 (2002): 1155–98.

Nelson, Alan. "Economic Rationality and Morality." *Philosophy and Public Affairs* 17 (1988): 149–66.

Ng, Yew-Kwang. *Welfare Economics: Introduction and Development of Basic Concepts*, rev. ed. London: Macmillan, 1983.

Nozick, Robert. "Newcomb's Problem and Two Principles of Choice," in Nicholas Rescher (ed.), *Essays in Honor of Carl G. Hempel*. Dordrecht: Kluwer, 1969, pp. 114–46.

Anarchy, State and Utopia. New York: Basic Books, 1974.

Nussbaum, Martha. *Women and Human Development*. Cambridge University Press, 2000.

Upheavals of Thought: The Intelligence of Emotions. Cambridge University Press, 2001.

Nussbaum, Martha and Amartya Sen (eds.). *The Quality of Life*. Oxford: Clarendon Press, 1993.

Okin, Susan Moller. *Justice, Gender and the Family*. New York: Basic Books, 1989.

Okun, Arthur. *Equality and Efficiency: The Big Tradeoff*. Washington, DC: Brookings, 1975.

Osborne, Martin. *An Introduction to Game Theory*. Oxford University Press, 2003.

Parfit, Derek. "Prudence, Morality and the Prisoner's Dilemma." *Proceedings of the British Academy* 65 (1979): 539–64.

Reasons and Persons. Oxford University Press, 1984.

"Equality or Priority." Lindley Lecture, University of Kansas, 1991; reprinted in Clayton and Williams (2000), pp. 81–125.

Pattanaik, Prasanta and Yongsheng Xu. "On Ranking Opportunity Sets in Terms of Freedom of Choice." *Recherches Economiques de Louvain* 56 (1990): 383–90.

Paul, Ellen, Fred Miller, Jr., Jeffrey Paul, and John Ahrens (eds.). *The New Social Contract: Essays on Gauthier*. Oxford: Blackwell, 1988.

Paxton, Will (ed.). *Equal Shares? Building a Progressive and Coherent Asset-Based Welfare Policy*. London: Institute for Public Policy Research, 2002.

Pazner, Elisha A. and David Schmeidler. "A Difficulty in the Concept of Fairness." *Review of Economic Studies* 41 (1974): 441–3.

Pennock, J. Roland and John W. Chapman (eds.). *Equality*. New York: Atherton, 1967.

Peter, Fabienne. "Choice, Consent, and the Legitimacy of Market Transactions." *Economics and Philosophy* 20 (2004): 1–18.

Pettit, Philip. "*Virtus Normativa:* Rational Choice Perspectives." *Ethics* 100 (1990): 725–55.

Pettit, Philip and Robert Sugden. "The Backward Induction Paradox." *Journal of Philosophy* 86 (1989): 169–82.

Phelps, Edmund S. "Taxation of Wage Income for Economic Justice." *Quarterly Journal of Economics* 87 (1973): 332–54.

Pinker, Steven. *The Blank Slate: The Modern Denial of Human Nature.* New York: Viking, 2002.

Plamenatz, John. "Diversity of Rights and Kinds of Equality," in Pennock and Chapman (1967), pp. 79–98.

Pogge, Thomas. *Realizing Rawls.* Ithaca: Cornell University Press, 1989.

World Poverty and Human Rights: Cosmopolitan Responsibilities and Reforms. Cambridge, MA: Polity Press, 2002.

Pojman, Lewis and Robert Westmoreland (eds.). *Equality: Selected Readings.* Oxford University Press, 1997.

Posner, Richard. "Utilitarianism, Economics, and Legal Theory." *Journal of Legal Studies* 8 (1979): 103–40.

Putnam, Hilary. *The Collapse of the Fact/Value Dichotomy and Other Essays.* Cambridge, MA: Harvard University Press, 2002.

Putterman, Louis. "On Some Recent Explanations of Why Capital Hires Labor." *Economic Inquiry* 22 (1984): 171–87.

Radin, Margaret. "Market Inalienability." *Harvard Law Review* 100 (1987): 1849–1937.

Contested Commodities. Cambridge, MA: Harvard University Press, 1996.

Railton, Peter. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy and Public Affairs* 13 (1984): 134–71.

Ramsey, Frank. "Truth and Probability," in R. B. Braithwaite (ed.), *The Foundations of Mathematics and Other Logical Essays.* London: Routledge, 1926, pp. 156–98.

Rand, Ayn. *The Virtue of Selfishness.* New York: Signet, 1964.

Rawls, John. "Two Concepts of Rules." *Philosophical Review* 64 (1955): 3–33.

A Theory of Justice. Cambridge, MA: Harvard University Press, 1971.

"Social Unity and Primary Goods," in Sen and Williams (1982), pp. 159–86.

Political Liberalism. New York: Columbia University Press, 1993.

Justice As Fairness: A Restatement. Cambridge, MA: Harvard University Press, 2001a.

Law of Peoples. Cambridge, MA: Harvard University Press, 2001b.

Raz, Joseph. *The Morality of Freedom.* Oxford University Press, 1986.

Reder, Melvin. "The Place of Ethics in the Theory of Production," in Michael Boskin (ed.), *Economics and Human Welfare: Essays in Honor of Tibor Scitovsky.* New York: Academic Press, 1979, pp. 133–46.

Regan, Donald. *Utilitarianism and Cooperation.* Oxford: Clarendon Press, 1980.

Reich, Michael and James Devine. "The Microeconomics of Conflict and Hierarchy in Capitalist Production." *Review of Radical Political Economy* 12 (1981): 27–45.

Riley, Jonathan. "Rights to Liberty in Purely Private Matters, Part I." *Economics and Philosophy* 5 (1989): 121–66.

"Rights to Liberty in Purely Private Matters, Part II." *Economics and Philosophy* 6 (1990): 27–64.

Robbins, Lionel. *An Essay on the Nature and Significance of Economic Science,* 2nd ed. London: Macmillan, 1935.

Robertson, Dennis. "What Does the Economist Economize?" in *Economic Commentaries.* London: Staples, 1956, pp. 147–55.

Roemer, John. "Equality of Talent." *Economics and Philosophy* 1 (1985): 151–88.

"The Mismarriage of Bargaining Theory and Distributive Justice." *Ethics* 97 (1986a): 88–110.

"Equality of Resources Implies Equality of Welfare." *Quarterly Journal of Economics* 101 (1986b): 751–84.

"Egalitarianism, Responsibility, and Information." *Economics and Philosophy* 3 (1987): 215–44.

Free to Lose. Cambridge, MA: Harvard University Press, 1988.

Theories of Distributive Justice. Cambridge, MA: Harvard University Press, 1996.

Equality of Opportunity. Cambridge, MA: Harvard University Press, 1998.

"Equality of Opportunity: A Progress Report." *Social Choice and Welfare* 19 (2002): 455–72.

Rosenberg, Alexander. *Microeconomic Laws: A Philosophical Analysis.* University of Pittsburgh Press, 1976.

Philosophy of Social Science. Boulder: Westview, 1995.

Roth, Alvin and Michael Malouf. "Game Theoretical Models and the Role of Information in Bargaining." *Psychological Review* 86 (1979): 574–94.

Roth, Alvin, Michael Malouf, and J. Keith Murnighan. "Sociological versus Strategic Factors in Bargaining." *Journal of Economic Behavior and Organizations* 2 (1981): 153–77.

Rothstein, Richard. *Class and Schools: Using Social, Economic and Educational Reform to Close the Black–White Achievement Gap.* Washington, DC: Economic Policy Institute, 2004.

Rousseau, Jean-Jacques. *The Social Contract* (1762), tr. Maurice Cranston. Harmondsworth: Penguin, 1968.

Rubinstein, Ariel. *Modeling Bounded Rationality.* Cambridge: MIT Press, 1997.

Ryan, Alan (ed.). *The Idea of Freedom: Essays in Honor of Isaiah Berlin.* Oxford University Press, 1979.

Sagoff, Mark. "Values and Preferences." *Ethics* 96 (1986): 301–16.

Price, Principle, and the Environment. Cambridge University Press, 2004.

Samuelson, Paul. *Foundations of Economic Analysis.* Cambridge, MA: Harvard University Press, 1947.

"Evaluation of Real National Income." *Oxford Economic Papers* (N.S.) 2 (1950): 1–29.

"An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money." *Journal of Political Economy* 66 (1958): 467–82.

"Reply." *Journal of Political Economy* 67 (1959): 518–22.

"Infinity, Unanimity and Singularity: A Reply." *Journal of Political Economy* 68 (1960): 76–83.

Sandel, Michael. *Liberalism and the Limits of Justice.* Cambridge University Press, 1982.

Savage, Leonard. *The Foundations of Statistics.* New York: Dover, 1972.

Scanlon, Thomas. "Preference and Urgency." *Journal of Philosophy* 72 (1975): 655–70.

"Contractualism and Utilitarianism," in Sen and Williams (1982), pp. 103–28.

"Equality of Resources and Equality of Welfare: A Forced Marriage?" *Ethics* 97 (1986): 111–18.

"The Moral Basis of Interpersonal Comparisons," in Elster and Roemer (1991), pp. 17–44.

What We Owe to Each Other. Cambridge, MA: Harvard University Press, 1998.
"The Diversity of Objections to Inequality," in T. Scanlon, *The Difficulty of Tolerance: Essays in Political Philosophy*. Cambridge University Press, 2003, pp. 202–18.
Scheffler, Samuel. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations underlying Rival Moral Conceptions*. Oxford: Clarendon Press, 1982.
(ed.). *Consequentialism and Its Critics*. Oxford University Press, 1988.
Schelling, Thomas. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press, 1960.
Micromotives and Macrobehavior. New York: Norton, 1978.
Choice and Consequence. Cambridge, MA: Harvard University Press, 1984.
Schick, Frederic. "Money Pumps and Dutch Bookies." *Journal of Philosophy* 83 (1986): 112–19.
Schmidtz, David. "The Institution of Property." *Social Philosophy and Policy* 11 (1994): 42–62.
Schon, Donald. *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books, 1983.
Schotter, Andrew. *The Economic Theory of Social Institutions*. Cambridge University Press, 1981.
Schultze, Charles L. *The Public Use of Private Interest*. Washington, DC: Brookings, 1977.
Seidenfeld, Teddy, Joseph Kadane, and Mark Schervish. "On the Shared Preferences of Two Bayesian Decision Makers." *Journal of Philosophy* 86 (1989): 225–44.
Selten, Reinhart. "Re-examination of the Perfectness Concept for Equilibrium in Extensive Games." *International Journal of Game Theory* 4 (1975): 22–5.
Sen, Amartya. "Isolation, Assurance, and the Social Rate of Discount." *Quarterly Journal of Economics* 81 (1967): 112–24.
Collective Welfare and Social Choice. San Francisco: Holden-Day, 1970a.
"The Impossibility of a Paretian Liberal." *Journal of Political Economy* 78 (1970b): 152–7.
"Choice Functions and Revealed Preference." *Review of Economic Studies* 38 (1971): 307–17.
"Behaviour and the Concept of Preference." *Economica* 40 (1973): 241–59.
"Welfare Inequalities and Rawlsian Axiomatics." *Theory and Decision* 7 (1976): 243–62.
"Rational Fools." *Philosophy and Public Affairs* 6 (1977): 317–44.
"Utilitarianism and Welfarism." *Journal of Philosophy* 76 (1979): 463–88.
Choice, Welfare and Measurement. Cambridge: MIT Press, 1982a.
"Rights and Agency." *Philosophy and Public Affairs* 11 (1982b): 3–39.
"Liberty and Social Choice." *Journal of Philosophy* 80 (1983): 5–28.
Commodities and Capabilities. Hennipman Lectures in Economics, vol. 7, *Theory, Institutions, Policy*. Amsterdam: North-Holland, 1985a.
"Rationality and Uncertainty." *Theory and Decision* 18 (1985b): 109–27.
"Foundations of Social Choice Theory: An Epilogue," in Elster and Hylland (1986), pp. 213–48.
On Ethics and Economics. Oxford: Blackwell, 1987a.

"The Standard of Living: Lecture I, Concepts and Critiques" (1987b), in Sen et al. (1987), pp. 1–19.

"The Standard of Living: Lecture II, Lives and Capabilities" (1987c), in Sen et al. (1987), pp. 20–38.

"Gender and Cooperative Conflicts," in Irene Tinker (ed.), *Persistent Inequalities.* Oxford University Press, 1990, pp. 123–49.

Inequality Reexamined. Cambridge, MA: Harvard University Press, 1992a.

"Minimal Liberty." *Economica* 59 (1992b): 139–59.

"Maximization and the Act of Choice." *Econometrica* 65 (1997): 745–79.

"The Possibility of Social Choice." *American Economic Review* 89 (1999): 349–78.

Rationality and Freedom. Cambridge, MA: Harvard University Press, 2002.

Sen, Amartya, John Muellbauer, Ravi Kanbur, Keith Hart, and Bernard Williams. *The Standard of Living,* ed. Geoffrey Hawthorn. Cambridge University Press, 1987.

Sen, Amartya and Bernard Williams (eds.). *Utilitarianism and Beyond.* Cambridge University Press, 1982.

Shapiro, Carl and Joseph E. Stiglitz. "Can Unemployment Be Involuntary? Reply." *American Economic Review* 75 (1985): 1215–17.

Sidgwick, Henry. *The Methods of Ethics,* 6th ed. London: Macmillan, 1901.

Simon, Herbert. *Models of Bounded Rationality.* Cambridge: MIT Press, vol. 1 (1982) and vol. 3 (1997).

Singer, Peter. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1 (1972): 229–43.

"Altruism and Commerce: A Defense of Titmuss against Arrow." *Philosophy and Public Affairs* 2 (1973): 312–20.

Animal Liberation: A New Ethics for Our Treatment of Animals. New York: Random House, 1975.

Practical Ethics. Cambridge University Press, 1979.

(ed.). *Applied Ethics.* Oxford University Press, 1986.

One World: The Ethics of Globalization. New Haven: Yale University Press, 2002.

Skidelsky, Robert. *John Maynard Keynes: The Economist as Savior, 1920–1937.* New York: Penguin, 1992.

Skyrms, Brian. *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws.* New Haven: Yale University Press, 1980.

Evolution of the Social Contract. Cambridge University Press, 1996.

The Stag Hunt and the Evolution of Social Structure. Cambridge University Press, 2003.

Smart, J. J. C. and Bernard Williams. *Utilitarianism: For and Against.* Cambridge University Press, 1973.

Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776). New York: Random House, 1937.

Snowdon, Brian and Howard Vane. *Conversations with Leading Economists: Interpreting Modern Macroeconomics.* Cheltenham: Edward Elgar, 1999.

Solow, Robert. "Wage Bargaining and Unemployment." *American Economic Review* 71 (1981): 896–908.

The Labor Market as a Social Institution. Oxford: Blackwell, 1990.

Steiner, Hillel. *An Essay on Rights.* Oxford: Blackwell, 1994.

Stich, Steven. *From Folk Psychology to Cognitive Science: The Case against Belief.* Cambridge: MIT Press, 1983.

Stone, Christopher. *Should Trees Have Standing, and Other Essays on Law, Morals and the Environment.* Dobbs Ferry: Oceana, 1996.

Strasnick, Steven. "Social Choice and the Derivation of Rawls's Difference Principle." *Journal of Philosophy* 73 (1976): 85–99.

Sugden, Robert. "Liberty, Preference and Choice." *Economics and Philosophy* 2 (1985): 213–31.

The Economics of Rights, Co-operation and Welfare. New York: Blackwell, 1986.

"Spontaneous Order." *Journal of Economic Perspectives* 3 (1989): 85–97.

"Contractarianism and Norms." *Ethics* 100 (1990): 768–86.

"The Metric of Opportunity." *Economics and Philosophy* 14 (1998): 307–37.

"The Logic of Team Reasoning." *Philosophical Explorations* 6 (2003): 165–81.

Sugden, Robert and Alan Williams. *The Principles of Practical Cost–Benefit Analysis.* Oxford University Press, 1978.

Sumner, L. W. *The Moral Foundation of Rights.* Oxford: Clarendon Press, 1987.

Sunstein, Cass. *Republic.com.* Princeton University Press, 2002.

Tawney, R. H. *Equality.* New York: Harcourt, Brace, 1931.

Taylor, Michael C. *The Possibility of Cooperation.* Cambridge University Press, 1987.

Taylor, Michael and Hugh Ward. "Chickens, Whales and Lumpy Goods: Alternative Models of Public Goods Provision." *Political Science* 30 (1982): 350–70.

Temkin, Larry. *Inequality.* Oxford University Press, 1993.

"Egalitarianism Defended." *Ethics* 113 (2003): 764–82.

Thomson, Garrett. *Needs.* London: Routledge, 1987.

Thomson, Judith. *The Realm of Rights.* Cambridge, MA: Harvard University Press, 1990.

Thomson, William and Terje Lensberg. *Axiomatic Theory of Bargaining with a Variable Number of Agents.* Cambridge University Press, 1989.

Thurow, Lester C. "Cash vs. In-Kind Redistribution," in Dworkin et al. (1977), pp. 85–106.

Titmuss, Richard. *The Gift Relationship: From Human Blood to Social Policy.* New York: Random House, 1971.

Tuomela, Raimo. *Cooperation: A Philosophical Study* (Philosophical Studies Series, 82). Dordrecht: Kluwer, 2000.

"The We-mode and the I-mode," in F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality.* Lanham: Rowman & Littlefield, 2003, pp. 93–128.

Tversky, Amos and Daniel Kahneman. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (1981): 453–8.

Tversky, Amos, Paul Slovic, and Daniel Kahneman. "The Causes of Preference Reversal." *American Economic Review* 80 (1990): 204–17.

Tversky, Amos and Richard Thaler. "Preference Reversals." *Journal of Economic Perspectives* 4 (1990): 201–11.

Unger, Peter. *Living High and Letting Die.* Oxford University Press, 1996.

Vanderschraaf, Peter. *Learning and Coordination: Inductive Deliberation, Equilibrium and Convention.* New York: Routledge, 2001.

van Parijs, Philippe. "On the Ethical Foundations of Basic Income." Working Paper no. CMID 32, Institut Supérieur de Philosophie, Université Catholique de Louvain, 1989.

"The Second Marriage of Justice and Efficiency." *Journal of Social Policy* 19 (1990): 1–25.

"Why Surfers Should Be Fed: The Liberal Case for an Unconditional Basic Income." *Philosophy and Public Affairs* 20 (1991): 101–31.

Varian, Hal. "Equity, Envy and Efficiency." *Journal of Economic Theory* 9 (1974): 63–91.

"Distributive Justice, Welfare Economics and the Theory of Fairness." *Philosophy and Public Affairs* 4 (1975): 223–47.

"Dworkin on Equality of Resources." *Economics and Philosophy* 1 (1985): 110–27.

Vickrey, William. "Measuring Marginal Utility by Reactions to Risk." *Econometrica* 13 (1945): 319–33.

von Mises, Ludwig. *Human Actions: A Treatise on Economics.* New Haven: Yale University Press, 1941.

von Neumann, John and Oskar Morgenstern. *Theory of Games and Economic Behavior,* 2nd ed. Princeton University Press, 1947.

von Wright, Georg. *Explanation and Understanding.* Ithaca: Cornell University Press, 1971.

Waldron, Jeremy (ed.). *Theories of Rights.* Oxford University Press, 1984.

The Right to Private Property. Oxford University Press, 1990.

Walsh, Vivian. "Rationality as Self-Interest versus Rationality as Present Aim." *American Economic Review Papers and Proceedings* 84 (1994): 401–5.

Walzer, Michael. *Spheres of Justice.* New York: Basic Books, 1983.

Weibull, Jörgen. "Persistent Unemployment as Subgame Perfect Equilibrium." Seminar Paper no. 381, Institute for International Economic Studies, Stockholm, May 1987.

Weitzman, Lenore. *The Divorce Revolution: The Unexpected Social and Economic Consequences for Women and Children in America.* New York: Free Press, 1985.

Wellman, Carl. *A Theory of Rights: Persons under Laws, Institutions and Morals.* Totowa: Rowman & Allenheld, 1985.

Weston, Samuel. "Toward a Better Understanding of the Positive/Normative Distinction in Economics." *Economics and Philosophy* 10 (1994): 1–18.

Weymark, John. "A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism," in Elster and Roemer (1991), pp. 255–320.

Williams, Bernard. "The Idea of Equality," in Peter Laslett and William Runciman (eds.), *Philosophy, Politics and Society.* New York: Barnes & Noble, 1962.

"A Critique of Utilitarianism," in Smart and Williams (1973), pp. 77–150.

Winch, Peter. *The Idea of a Social Science.* London: Routledge, 1958.

Wolff, Robert. *Understanding Rawls: A Reconstruction and Critique of a Theory of Justice.* Princeton University Press, 1977.

World Bank. *Breaking the Conflict Trap: Civil War and Development Policy.* Oxford University Press, 2003.

Yaari, Menachem and Maja Bar-Hillel. "On Dividing Justly." *Social Choice and Welfare* 1 (1984): 1–24.

Young, H. Peyton. *Equity.* Princeton University Press, 1994.

Index