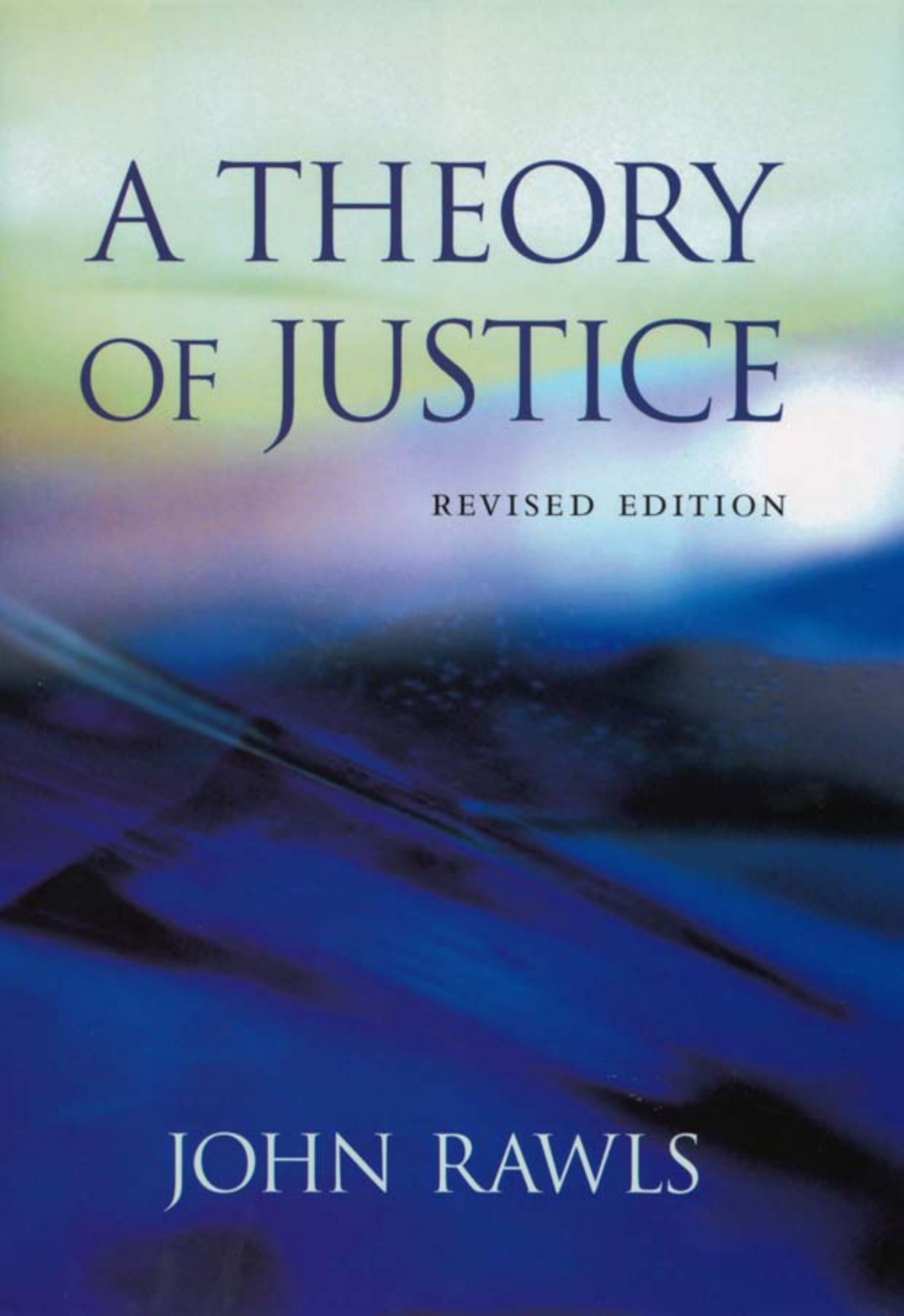
A THEORY OF JUSTICE

REVISED EDITION



JOHN RAWLS

A THEORY OF JUSTICE

A THEORY OF JUSTICE

Revised Edition

JOHN RAWLS

THE BELKNAP PRESS OF
HARVARD UNIVERSITY PRESS
CAMBRIDGE, MASSACHUSETTS

This book is a revised edition of *A Theory of Justice,*
published in 1971 by Harvard University Press.

Library of Congress Cataloging-in-Publication Data
Rawls, John, 1921–
A theory of justice / John Rawls. — Rev. ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-674-00077-3 (cloth : alk. paper). —
ISBN 0-674-00078-1 (paper : alk. paper)
1. Justice. I. Title.
JC578.R38 1999
320′.01′1—dc21 99-29110

For Mard

CONTENTS

PREFACE FOR THE REVISED EDITION

It gives me great pleasure to provide this preface to the revised edition of *A Theory of Justice.* Despite many criticisms of the original work, I still accept its main outlines and defend its central doctrines. Of course, I wish, as one might expect, that I had done certain things differently, and I would now make a number of important revisions. But if I were writing *A Theory of Justice* over again, I would not write, as authors sometimes say, a completely different book.

In February and March of 1975 the original English text was considerably revised for the German edition of that year. To the best of my knowledge these revisions have been included in all subsequent translations and no further ones have been added since that time. All translations have, therefore, been made from the same revised text. Since this revised text includes what I believe are significant improvements, the translated editions (provided accuracy is preserved) until now have been superior to the original. This revised edition incorporates these improvements.

Before commenting on the more important revisions and why they were made, I will comment on the conception of justice presented in *A Theory of Justice,* a conception I call "justice as fairness." The central ideas and aims of this conception I see as those of a philosophical conception for a constitutional democracy. My hope is that justice as fairness will seem reasonable and useful, even if not fully convincing, to a wide range of thoughtful political opinions and thereby express an essential part of the common core of the democratic tradition.

The central aims and ideas of that conception I refer to in the preface to the first edition. As I explain in the second and third paragraphs of that preface, I wanted to work out a conception of justice that provides a reasonably systematic alternative to utilitarianism, which in one form or another has long dominated the Anglo-Saxon tradition of political thought. The primary reason for wanting to find such an alternative is the weakness, so I think, of utilitarian doctrine as a basis for the institutions

of constitutional democracy. In particular, I do not believe that utilitarianism can provide a satisfactory account of the basic rights and liberties of citizens as free and equal persons, a requirement of absolutely first importance for an account of democratic institutions. I used a more general and abstract rendering of the idea of the social contract by means of the idea of the original position as a way to do that. A convincing account of basic rights and liberties, and of their priority, was the first objective of justice as fairness. A second objective was to integrate that account with an understanding of democratic equality, which led to the principle of fair equality of opportunity and the difference principle.[1]

In the revisions I made in 1975 I removed certain weaknesses in the original edition. These I shall now try to indicate, although I am afraid much of what I say will not be intelligible without some prior knowledge of the text. Leaving this concern aside, one of the most serious weaknesses was in the account of liberty, the defects of which were pointed out by H. L. A. Hart in his critical discussion of 1973.[2] Beginning with §11, I made revisions to clear up several of the difficulties Hart noted. It must be said, however, that the account in the revised text, although considerably improved, is still not fully satisfactory. A better version is found in a later essay of 1982 entitled "The Basic Liberties and Their Priority."[3] This essay attempts to answer what I came to regard as Hart's most important objections. The basic rights and liberties and their priority are there said to guarantee equally for all citizens the social conditions essential for the adequate development and the full and informed exercise of their two moral powers—their capacity for a sense of justice and their capacity for a conception of the good—in what I call the two fundamental cases. Very briefly, the first fundamental case is the application of the principles of justice to the basic structure of society by the exercise of citizens' sense of justice. The second fundamental case is the application of citizens' powers of practical reason and thought in forming, revising, and rationally pursuing their conception of the good. The equal political liberties, including their fair value (an idea introduced in §36), and freedom of thought, liberty of conscience, and freedom of association, are to insure

1. For these two principles see §§12–14 of Chapter II. It is these two principles, and particularly the difference principle, which give justice as fairness its liberal, or social democratic, character.

2. See his "Rawls on Liberty and Its Priority," *University of Chicago Law Review,* 40 (1973), pp. 534–555.

3. See *Tanner Lectures on Human Values* (Salt Lake City: University of Utah Press, 1982), vol. III, pp. 3–87, republished as Lecture VIII in John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993).

that the exercise of the moral powers can be free, informed, and effective in these two cases. These changes in the account of liberty can, I think, fit comfortably within the framework of justice as fairness as found in the revised text.

A second serious weakness of the original edition was its account of primary goods. These were said to be things that rational persons want whatever else they want, and what these were and why was to be explained by the account of goodness in Chapter VII. Unhappily that account left it ambiguous whether something's being a primary good depends solely on the natural facts of human psychology or whether it also depends on a moral conception of the person that embodies a certain ideal. This ambiguity is to be resolved in favor of the latter: persons are to be viewed as having two moral powers (those mentioned above) and as having higher-order interests in developing and exercising those powers. Primary goods are now characterized as what persons need in their status as free and equal citizens, and as normal and fully cooperating members of society over a complete life. Interpersonal comparisons for purposes of political justice are to be made in terms of citizens' index of primary goods and these goods are seen as answering to their needs as citizens as opposed to their preferences and desires. Beginning with §15, I made revisions to convey this change of view, but these revisions fall short of the fuller statement I have given since in an essay, published in 1982, entitled "Social Unity and Primary Goods."[4] As with the changes in the account of the basic liberties, I think the changes required by that statement can be incorporated within the framework of the revised text.

Many other revisions were made, especially in Chapter III and again, though fewer, in Chapter IV. In Chapter III I simply tried to make the reasoning clearer and less open to misunderstanding. The revisions are too numerous to note here, but they do not, I think, depart in any important way from the view of the original edition. After Chapter IV there are few changes. I revised §44 in Chapter V on just savings, again trying to make it clearer; and I rewrote the first six paragraphs of §82 of Chapter IX to correct a serious mistake in the argument for the priority of liberty;[5] and there are further changes in the rest of that section. Perhaps having identified what I regard as the two important changes, those in the ac-

4. This essay appears in *Utilitarianism and Beyond,* edited by Amartya Sen and Bernard Williams (Cambridge: Cambridge University Press, 1982), pp. 159–185; also in John Rawls, *Collected Papers,* edited by Samuel Freeman (Cambridge, Mass.: Harvard University Press, 1999, chap. 17, pp. 359–387.

5. For this mistake see "Basic Liberties and Their Priority," ibid., n. 83, p. 87, or *Political Liberalism,* n. 84, p. 371.

counts of the basic liberties and of primary goods, these indications suffice to convey the nature and extent of the revisions.

If I were writing *A Theory of Justice* now, there are two things I would handle differently. One concerns how to present the argument from the original position (see Chapter III) for the two principles of justice (see Chapter II). It would have been better to present it in terms of two comparisons. In the first parties would decide between the two principles of justice, taken as a unit, and the principle of (average) utility as the sole principle of justice. In the second comparison, the parties would decide between the two principles of justice and those same principles but for one important change: the principle of (average) utility is substituted for the difference principle. (The two principles after this substitution I called a mixed conception, and here it is understood that the principle of utility is to be applied subject to the constraints of the prior principles: the principle of the equal liberties and the principle of fair equality of opportunity.) Using these two comparisons has the merit of separating the arguments for the equal basic liberties and their priority from the arguments for the difference principle itself. The arguments for the equal basic liberties are at first glance much stronger, as those for the difference principle involve a more delicate balance of considerations. The primary aim of justice as fairness is achieved once it is clear that the two principles would be adopted in the first comparison, or even in a third comparison in which the mixed conception of the second comparison is adopted rather than the principle of utility. I continue to think the difference principle important and would still make the case for it, taking for granted (as in the second comparison) an institutional background that satisfies the two preceding principles. But it is better to recognize that this case is less evident and is unlikely ever to have the force of the argument for the two prior principles.

Another revision I would now make is to distinguish more sharply the idea of a property-owning democracy (introduced in Chapter V) from the idea of a welfare state.[6] These ideas are quite different, but since they both allow private property in productive assets, we may be misled into thinking them essentially the same. One major difference is that the background institutions of property-owning democracy, with its system of (workably) competitive markets, tries to disperse the ownership of wealth and capital, and thus to prevent a small part of society from

6. The term "property-owning democracy," as well as some features of the idea, I borrowed from J. E. Meade, *Efficiency, Equality, and the Ownership of Property* (London: G. Allen & Unwin, 1964); see esp. Chapter V.

controlling the economy and indirectly political life itself. Property-owning democracy avoids this, not by redistributing income to those with less at the end of each period, so to speak, but rather by ensuring the widespread ownership of productive assets and human capital (educated abilities and trained skills) at the beginning of each period; all this against a background of equal basic liberties and fair equality of opportunity. The idea is not simply to assist those who lose out through accident or misfortune (although this must be done), but instead to put all citizens in a position to manage their own affairs and to take part in social cooperation on a footing of mutual respect under appropriately equal conditions.

Note here two different conceptions of the aim of political institutions over time. In a welfare state the aim is that none should fall below a decent standard of life, and that all should receive certain protections against accident and misfortune—for example, unemployment compensation and medical care. The redistribution of income serves this purpose when, at the end of each period, those who need assistance can be identified. Such a system may allow large and inheritable inequities of wealth incompatible with the fair value of the political liberties (introduced in §36), as well as large disparities of income that violate the difference principle. While some effort is made to secure fair equality of opportunity, it is either insufficient or else ineffective given the disparities of wealth and the political influence they permit.

By contrast, in a property-owning democracy the aim is to carry out the idea of society as a fair system of cooperation over time among citizens as free and equal persons. Thus, basic institutions must from the outset put in the hands of citizens generally, and not only of a few, the productive means to be fully cooperating members of a society. The emphasis falls on the steady dispersal over time of the ownership of capital and resources by the laws of inheritance and bequest, on fair equality of opportunity secured by provisions for education and training, and the like, as well as on institutions that support the fair value of the political liberties. To see the full force of the difference principle it should be taken in the context of property-owning democracy (or of a liberal socialist regime) and not a welfare state: it is a principle of reciprocity, or mutuality, for society seen as a fair system of cooperation among free and equal citizens from one generation to the next.

The mention (a few lines back) of a liberal socialist regime prompts me to add that justice as fairness leaves open the question whether its principles are best realized by some form of property-owning democracy or by a liberal socialist regime. This question is left to be settled by

historical conditions and the traditions, institutions, and social forces of each country.[7] As a political conception, then, justice, as fairness includes no natural right of private property in the means of production (although it does include a right to personal property as necessary for citizens' independence and integrity), nor a natural right to worker-owned and -managed firms. It offers instead a conception of justice in the light of which, given the particular circumstances of a country, those questions can be reasonably decided.

John Rawls
November 1990

7. See the last two paragraphs of §42, Chapter V.

PREFACE

In presenting a theory of justice I have tried to bring together into one coherent view the ideas expressed in the papers I have written over the past dozen years or so. All of the central topics of these essays are taken up again, usually in considerably more detail. The further questions required to round out the theory are also discussed. The exposition falls into three parts. The first part covers with much greater elaboration the same ground as "Justice as Fairness" (1958) and "Distributive Justice: Some Addenda" (1968), while the three chapters of the second part correspond respectively, but with many additions, to the topics of "Constitutional Liberty" (1963), "Distributive Justice" (1967), and "Civil Disobedience" (1966). The second chapter of the last part covers the subjects of "The Sense of Justice" (1963). Except in a few places, the other chapters of this part do not parallel the published essays. Although the main ideas are much the same, I have tried to eliminate inconsistencies and to fill out and strengthen the argument at many points.

Perhaps I can best explain my aim in this book as follows. During much of modern moral philosophy the predominant systematic theory has been some form of utilitarianism. One reason for this is that it has been espoused by a long line of brilliant writers who have built up a body of thought truly impressive in its scope and refinement. We sometimes forget that the great utilitarians, Hume and Adam Smith, Bentham and Mill, were social theorists and economists of the first rank; and the moral doctrine they worked out was framed to meet the needs of their wider interests and to fit into a comprehensive scheme. Those who criticized them often did so on a much narrower front. They pointed out the obscurities of the principle of utility and noted the apparent incongruities between many of its implications and our moral sentiments. But they failed, I believe, to construct a workable and systematic moral conception to oppose it. The outcome is that we often seem forced to choose between utilitarianism and intuitionism. Most likely we finally settle upon a vari-

ant of the utility principle circumscribed and restricted in certain ad hoc ways by intuitionistic constraints. Such a view is not irrational; and there is no assurance that we can do better. But this is no reason not to try.

What I have attempted to do is to generalize and carry to a higher order of abstraction the traditional theory of the social contract as represented by Locke, Rousseau, and Kant. In this way I hope that the theory can be developed so that it is no longer open to the more obvious objections often thought fatal to it. Moreover, this theory seems to offer an alternative systematic account of justice that is superior, or so I argue, to the dominant utilitarianism of the tradition. The theory that results is highly Kantian in nature. Indeed, I must disclaim any originality for the views I put forward. The leading ideas are classical and well known. My intention has been to organize them into a general framework by using certain simplifying devices so that their full force can be appreciated. My ambitions for the book will be completely realized if it enables one to see more clearly the chief structural features of the alternative conception of justice that is implicit in the contract tradition and points the way to its further elaboration. Of the traditional views, it is this conception, I believe, which best approximates our considered judgments of justice and constitutes the most appropriate moral basis for a democratic society.

This is a long book, not only in pages. Therefore, to make things easier for the reader, a few remarks by way of guidance. The fundamental intuitive ideas of the theory of justice are presented in §§1–4 of Chapter I. From here it is possible to go directly to the discussion of the two principles of justice for institutions in §§11–17 of Chapter II, and then to the account of the original position in Chapter III, the whole chapter. A glance at §8 on the priority problem may prove necessary if this notion is unfamiliar. Next, parts of Chapter IV, §§33–35 on equal liberty and §§39–40 on the meaning of the priority of liberty and the Kantian interpretation, give the best picture of the doctrine. So far this is about a third of the whole and comprises most of the essentials of the theory.

There is a danger, however, that without consideration of the argument of the last part, the theory of justice will be misunderstood. In particular, the following sections should be emphasized: §§66–67 of Chapter VII on moral worth and self-respect and related notions; §77 of Chapter VIII on the basis of equality; and §§78–79 on autonomy and social union, §82 on the priority of liberty, and §§85–86 on the unity of the self and congruence, all in Chapter IX. Adding these sections to the others still comes to considerably less than half the text.

The section headings, the remarks that preface each chapter, and the index will guide the reader to the contents of the book. It seems superfluous to comment on this except to say that I have avoided extensive methodological discussions. There is a brief consideration of the nature of moral theory in §9, and of justification in §4 and §87. A short digression on the meaning of "good" is found in §62. Occasionally there are methodological comments and asides, but for the most part I try to work out a substantive theory of justice. Comparisons and contrasts with other theories, and criticisms thereof now and then, especially of utilitarianism, are viewed as means to this end.

By not including most of Chapters IV–VIII in the more basic parts of the book, I do not mean to suggest that these chapters are peripheral, or merely applications. Rather, I believe that an important test of a theory of justice is how well it introduces order and system into our considered judgments over a wide range of questions. Therefore the topics of these chapters need to be taken up, and the conclusions reached modify in turn the view proposed. But in this regard the reader is more free to follow his preferences and to look at the problems which most concern him.

In writing this book I have acquired many debts in addition to those indicated in the text. Some of these I should like to acknowledge here. Three different versions of the manuscript have passed among students and colleagues, and I have benefited beyond estimation from the innumerable suggestions and criticisms that I have received. I am grateful to Allan Gibbard for his criticism of the first version (1964–1965). To meet his objections to the veil of ignorance as then presented, it seemed necessary to include a theory of the good. The notion of primary goods based on the conception discussed in Chapter VII is the result. I also owe him thanks, along with Norman Daniels, for pointing out difficulties with my account of utilitarianism as a basis for individual duties and obligations. Their objections led me to eliminate much of this topic and to simplify the treatment of this part of the theory. David Diamond objected forcefully to my discussion of equality, particularly to its failure to consider the relevance of status. I eventually included an account of self-respect as a primary good to try to deal with this and other questions, including those of society as a social union of social unions and the priority of liberty. I had profitable discussions with David Richards on the problems of political duty and obligation. Although supererogation is not a central topic of the book, I have been helped in my comments on it by Barry

Curtis and John Troyer; even so they may still object to what I say. Thanks should also go to Michael Gardner and Jane English for several corrections which I managed to make in the final text.

I have been fortunate in receiving valuable criticisms from persons who have discussed the essays in print.[1] I am indebted to Brian Barry, Michael Lessnoff, and R. P. Wolff for their discussions of the formulation of and the argument for the two principles of justice.[2] Where I have not accepted their conclusions I have had to amplify the argument to meet their objections. I hope the theory as now presented is no longer open to the difficulties they raised, nor to those urged by John Chapman.[3] The relation between the two principles of justice and what I call the general conception of justice is similar to that proposed by S. I. Benn.[4] I am grateful to him, and to Lawrence Stern and Scott Boorman, for suggestions in this direction. The substance of Norman Care's criticisms of the conception of moral theory found in the essays seems sound to me, and I have tried to develop the theory of justice so that it avoids his objections.[5] In doing this, I have learned from Burton Dreben, who made W. V. Quine's view clear to me and persuaded me that the notions of meaning and analyticity play no essential role in moral theory as I conceive of it. Their relevance for other philosophical questions need not be disputed here one way or the other; but I have tried to make the theory of justice

1. In the order mentioned in the first paragraph, the references for the six essays are as follows: "Justice as Fairness," *The Philosophical Review,* vol. 57 (1958); "Distributive Justice: Some Addenda," *Natural Law Forum,* vol. 13 (1968); "Constitutional Liberty and the Concept of Justice," *Nomos VI: Justice,* ed. C. J. Friedrich and John Chapman (New York, Atherton Press, 1963); "Distributive Justice," *Philosophy, Politics, and Society,* Third Series, ed. Peter Laslett and W. G. Runciman (Oxford, Basil Blackwell, 1967); "The Justification of Civil Disobedience," *Civil Disobedience,* ed. H. A. Bedau (New York, Pegasus, 1969); "The Sense of Justice," *The Philosophical Review,* vol. 62 (1963).

2. See Brian Barry, "On Social Justice," *The Oxford Review* (Trinity Term, 1967), pp. 29–52; Michael Lessnoff, "John Rawls' Theory of Justice," *Political Studies,* vol. 19 (1971), pp. 65–80; and R. P. Wolff, "A Refutation of Rawls' Theorem on Justice," *Journal of Philosophy,* vol. 63 (1966), pp. 179–190. While "Distributive Justice" (1967) was completed and sent to the publisher before Wolff's article appeared, I regret that from oversight I failed to add a reference to it in proof.

3. See John Chapman, "Justice and Fairness," in *Nomos VI: Justice.*

4. See S. I. Benn, "Egalitarianism and the Equal Consideration of Interests," *Nomos IX: Equality,* ed. J. R. Pennock and John Chapman (New York, Atherton Press, 1967), pp. 72–78.

5. See Norman Care, "Contractualism and Moral Criticism," *The Review of Metaphysics,* vol. 23 (1969), pp. 85–101. I should also like to acknowledge here the criticisms of my work by R. L. Cunningham, "Justice: Efficiency or Fairness," *The Personalist,* vol. 52 (1971); Dorothy Emmett, "Justice," *Proceedings of the Aristotelian Society,* supp. vol. (1969); Charles Frankel, "Justice and Rationality," in *Philosophy, Science, and Method,* ed. Sidney Morgenbesser, Patrick Suppes, and Morton White (New York, St. Martin's Press, 1969); and Ch. Perelman, *Justice* (New York, Random House, 1967), esp. pp. 39–51.

independent of them. Thus I have followed with some modifications the point of view of my "Outline for Ethics."[6] I should also like to thank A. K. Sen for his searching discussion and criticisms of the theory of justice.[7] These have enabled me to improve the presentation at various places. His book will prove indispensable to philosophers who wish to study the more formal theory of social choice as economists think of it. At the same time, the philosophical problems receive careful treatment.

Many persons have volunteered written comments on the several versions of the manuscript. Gilbert Harman's on the earliest one were fundamental and forced me to abandon a number of views and to make basic changes at many points. I received others while at the Philosophical Institute at Boulder (summer 1966), from Leonard Krimerman, Richard Lee, and Huntington Terrell; and from Terrell again later. I have tried to accommodate to these, and to the very extensive and instructive comments of Charles Fried, Robert Nozick, and J. N. Shklar, each of whom has been of great help throughout. In developing the account of the good, I have gained much from J. M. Cooper, T. M. Scanlon, and A. T. Tymoczko, and from discussions over many years with Thomas Nagel, to whom I am also indebted for clarification about the relation between the theory of justice and utilitarianism. I must also thank R. B. Brandt and Joshua Rabinowitz for their many useful ideas for improvements in the second manuscript (1967–1968), and B. J. Diggs, J. C. Harsanyi, and W. G. Runciman for illuminating correspondence.

During the writing of the third version (1969–1970), Brandt, Tracy Kendler, E. S. Phelps, and Amélie Rorty were a constant source of advice, and their criticisms were of great assistance. On this manuscript I received many valuable comments and suggestions for changes from Herbert Morris, and from Lessnoff and Nozick; these have saved me from a number of lapses and have made the book much better. I am particularly grateful to Nozick for his unfailing help and encouragement during the last stages. Regrettably I have not been able to deal with all criticisms received, and I am well aware of the faults that remain; but the measure of my debt is not the shortfall from what might be but the distance traveled from the beginnings.

The Center for Advanced Study at Stanford provided the ideal place for me to complete my work. I should like to express my deep apprecia-

6. *The Philosophical Review,* vol. 50 (1951).

7. See *Collective Choice and Social Welfare* (San Francisco, Holden-Day, 1970), esp. pp. 136–141, 156–160.

tion for its support in 1969–1970, and for that of the Guggenheim and Kendall foundations in 1964–1965. I am grateful to Anna Tower and to Margaret Griffin for helping me with the final manuscript.

Without the good will of all these good people I never could have finished this book.

John Rawls

Cambridge, Massachusetts
August 1971

PART ONE. THEORY

CHAPTER I. JUSTICE AS FAIRNESS

In this introductory chapter I sketch some of the main ideas of the theory of justice I wish to develop. The exposition is informal and intended to prepare the way for the more detailed arguments that follow. Unavoidably there is some overlap between this and later discussions. I begin by describing the role of justice in social cooperation and with a brief account of the primary subject of justice, the basic structure of society. I then present the main idea of justice as fairness, a theory of justice that generalizes and carries to a higher level of abstraction the traditional conception of the social contract. The compact of society is replaced by an initial situation that incorporates certain procedural constraints on arguments designed to lead to an original agreement on principles of justice. I also take up, for purposes of clarification and contrast, the classical utilitarian and intuitionist conceptions of justice and consider some of the differences between these views and justice as fairness. My guiding aim is to work out a theory of justice that is a viable alternative to these doctrines which have long dominated our philosophical tradition.

1. THE ROLE OF JUSTICE

Justice is the first virtue of social institutions, as truth is of systems of thought. A theory however elegant and economical must be rejected or revised if it is untrue; likewise laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust. Each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override. For this reason justice denies that the loss of freedom for some is made right by a greater good shared by others. It does not allow that the sacrifices imposed on a few are outweighed by the larger sum of advantages enjoyed by many. Therefore in a just society the liberties of equal citizenship are taken as settled;

the rights secured by justice are not subject to political bargaining or to the calculus of social interests. The only thing that permits us to acquiesce in an erroneous theory is the lack of a better one; analogously, an injustice is tolerable only when it is necessary to avoid an even greater injustice. Being first virtues of human activities, truth and justice are uncompromising.

These propositions seem to express our intuitive conviction of the primacy of justice. No doubt they are expressed too strongly. In any event I wish to inquire whether these contentions or others similar to them are sound, and if so how they can be accounted for. To this end it is necessary to work out a theory of justice in the light of which these assertions can be interpreted and assessed. I shall begin by considering the role of the principles of justice. Let us assume, to fix ideas, that a society is a more or less self-sufficient association of persons who in their relations to one another recognize certain rules of conduct as binding and who for the most part act in accordance with them. Suppose further that these rules specify a system of cooperation designed to advance the good of those taking part in it. Then, although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts. There is a conflict of interests since persons are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share. A set of principles is required for choosing among the various social arrangements which determine this division of advantages and for underwriting an agreement on the proper distributive shares. These principles are the principles of social justice: they provide a way of assigning rights and duties in the basic institutions of society and they define the appropriate distribution of the benefits and burdens of social cooperation.

Now let us say that a society is well-ordered when it is not only designed to advance the good of its members but when it is also effectively regulated by a public conception of justice. That is, it is a society in which (1) everyone accepts and knows that the others accept the same principles of justice, and (2) the basic social institutions generally satisfy and are generally known to satisfy these principles. In this case while men may put forth excessive demands on one another, they nevertheless acknowledge a common point of view from which their claims may be adjudicated. If men's inclination to self-interest makes their vigilance

against one another necessary, their public sense of justice makes their secure association together possible. Among individuals with disparate aims and purposes a shared conception of justice establishes the bonds of civic friendship; the general desire for justice limits the pursuit of other ends. One may think of a public conception of justice as constituting the fundamental charter of a well-ordered human association.

Existing societies are of course seldom well-ordered in this sense, for what is just and unjust is usually in dispute. Men disagree about which principles should define the basic terms of their association. Yet we may still say, despite this disagreement, that they each have a conception of justice. That is, they understand the need for, and they are prepared to affirm, a characteristic set of principles for assigning basic rights and duties and for determining what they take to be the proper distribution of the benefits and burdens of social cooperation. Thus it seems natural to think of the concept of justice as distinct from the various conceptions of justice and as being specified by the role which these different sets of principles, these different conceptions, have in common.[1] Those who hold different conceptions of justice can, then, still agree that institutions are just when no arbitrary distinctions are made between persons in the assigning of basic rights and duties and when the rules determine a proper balance between competing claims to the advantages of social life. Men can agree to this description of just institutions since the notions of an arbitrary distinction and of a proper balance, which are included in the concept of justice, are left open for each to interpret according to the principles of justice that he accepts. These principles single out which similarities and differences among persons are relevant in determining rights and duties and they specify which division of advantages is appropriate. Clearly this distinction between the concept and the various conceptions of justice settles no important questions. It simply helps to identify the role of the principles of social justice.

Some measure of agreement in conceptions of justice is, however, not the only prerequisite for a viable human community. There are other fundamental social problems, in particular those of coordination, efficiency, and stability. Thus the plans of individuals need to be fitted together so that their activities are compatible with one another and they can all be carried through without anyone's legitimate expectations being severely disappointed. Moreover, the execution of these plans should lead to the

1. Here I follow H. L. A. Hart, *The Concept of Law* (Oxford, The Clarendon Press, 1961), pp. 155–159.

achievement of social ends in ways that are efficient and consistent with justice. And finally, the scheme of social cooperation must be stable: it must be more or less regularly complied with and its basic rules willingly acted upon; and when infractions occur, stabilizing forces should exist that prevent further violations and tend to restore the arrangement. Now it is evident that these three problems are connected with that of justice. In the absence of a certain measure of agreement on what is just and unjust, it is clearly more difficult for individuals to coordinate their plans efficiently in order to insure that mutually beneficial arrangements are maintained. Distrust and resentment corrode the ties of civility, and suspicion and hostility tempt men to act in ways they would otherwise avoid. So while the distinctive role of conceptions of justice is to specify basic rights and duties and to determine the appropriate distributive shares, the way in which a conception does this is bound to affect the problems of efficiency, coordination, and stability. We cannot, in general, assess a conception of justice by its distributive role alone, however useful this role may be in identifying the concept of justice. We must take into account its wider connections; for even though justice has a certain priority, being the most important virtue of institutions, it is still true that, other things equal, one conception of justice is preferable to another when its broader consequences are more desirable.

2. THE SUBJECT OF JUSTICE

Many different kinds of things are said to be just and unjust: not only laws, institutions, and social systems, but also particular actions of many kinds, including decisions, judgments, and imputations. We also call the attitudes and dispositions of persons, and persons themselves, just and unjust. Our topic, however, is that of social justice. For us the primary subject of justice is the basic structure of society, or more exactly, the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation. By major institutions I understand the political constitution and the principal economic and social arrangements. Thus the legal protection of freedom of thought and liberty of conscience, competitive markets, private property in the means of production, and the monogamous family are examples of major social institutions. Taken together as one scheme, the major institutions define men's rights and duties and influence their life prospects, what they can expect to be and how well they can hope to

do. The basic structure is the primary subject of justice because its effects are so profound and present from the start. The intuitive notion here is that this structure contains various social positions and that men born into different positions have different expectations of life determined, in part, by the political system as well as by economic and social circumstances. In this way the institutions of society favor certain starting places over others. These are especially deep inequalities. Not only are they pervasive, but they affect men's initial chances in life; yet they cannot possibly be justified by an appeal to the notions of merit or desert. It is these inequalities, presumably inevitable in the basic structure of any society, to which the principles of social justice must in the first instance apply. These principles, then, regulate the choice of a political constitution and the main elements of the economic and social system. The justice of a social scheme depends essentially on how fundamental rights and duties are assigned and on the economic opportunities and social conditions in the various sectors of society.

The scope of our inquiry is limited in two ways. First of all, I am concerned with a special case of the problem of justice. I shall not consider the justice of institutions and social practices generally, nor except in passing the justice of the law of nations and of relations between states (§58). Therefore, if one supposes that the concept of justice applies whenever there is an allotment of something rationally regarded as advantageous or disadvantageous, then we are interested in only one instance of its application. There is no reason to suppose ahead of time that the principles satisfactory for the basic structure hold for all cases. These principles may not work for the rules and practices of private associations or for those of less comprehensive social groups. They may be irrelevant for the various informal conventions and customs of everyday life; they may not elucidate the justice, or perhaps better, the fairness of voluntary cooperative arrangements or procedures for making contractual agreements. The conditions for the law of nations may require different principles arrived at in a somewhat different way. I shall be satisfied if it is possible to formulate a reasonable conception of justice for the basic structure of society conceived for the time being as a closed system isolated from other societies. The significance of this special case is obvious and needs no explanation. It is natural to conjecture that once we have a sound theory for this case, the remaining problems of justice will prove more tractable in the light of it. With suitable modifications such a theory should provide the key for some of these other questions.

The other limitation on our discussion is that for the most part I

examine the principles of justice that would regulate a well-ordered society. Everyone is presumed to act justly and to do his part in upholding just institutions. Though justice may be, as Hume remarked, the cautious, jealous virtue, we can still ask what a perfectly just society would be like.[2] Thus I consider primarily what I call strict compliance as opposed to partial compliance theory (§§25, 39). The latter studies the principles that govern how we are to deal with injustice. It comprises such topics as the theory of punishment, the doctrine of just war, and the justification of the various ways of opposing unjust regimes, ranging from civil disobedience and conscientious objection to militant resistance and revolution. Also included here are questions of compensatory justice and of weighing one form of institutional injustice against another. Obviously the problems of partial compliance theory are the pressing and urgent matters. These are the things that we are faced with in everyday life. The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems. The discussion of civil disobedience, for example, depends upon it (§§55–59). At least, I shall assume that a deeper understanding can be gained in no other way, and that the nature and aims of a perfectly just society is the fundamental part of the theory of justice.

Now admittedly the concept of the basic structure is somewhat vague. It is not always clear which institutions or features thereof should be included. But it would be premature to worry about this matter here. I shall proceed by discussing principles which do apply to what is certainly a part of the basic structure as intuitively understood; I shall then try to extend the application of these principles so that they cover what would appear to be the main elements of this structure. Perhaps these principles will turn out to be perfectly general, although this is unlikely. It is sufficient that they apply to the most important cases of social justice. The point to keep in mind is that a conception of justice for the basic structure is worth having for its own sake. It should not be dismissed because its principles are not everywhere satisfactory.

A conception of social justice, then, is to be regarded as providing in the first instance a standard whereby the distributive aspects of the basic structure of society are to be assessed. This standard, however, is not to be confused with the principles defining the other virtues, for the basic structure, and social arrangements generally, may be efficient or ineffi-

2. *An Enquiry Concerning the Principles of Morals,* sec. III, pt. I, par. 3, ed. L. A. Selby-Bigge, 2nd edition (Oxford, 1902), p. 184.

cient, liberal or illiberal, and many other things, as well as just or unjust. A complete conception defining principles for all the virtues of the basic structure, together with their respective weights when they conflict, is more than a conception of justice; it is a social ideal. The principles of justice are but a part, although perhaps the most important part, of such a conception. A social ideal in turn is connected with a conception of society, a vision of the way in which the aims and purposes of social cooperation are to be understood. The various conceptions of justice are the outgrowth of different notions of society against the background of opposing views of the natural necessities and opportunities of human life. Fully to understand a conception of justice we must make explicit the conception of social cooperation from which it derives. But in doing this we should not lose sight of the special role of the principles of justice or of the primary subject to which they apply.

In these preliminary remarks I have distinguished the concept of justice as meaning a proper balance between competing claims from a conception of justice as a set of related principles for identifying the relevant considerations which determine this balance. I have also characterized justice as but one part of a social ideal, although the theory I shall propose no doubt extends its everyday sense. This theory is not offered as a description of ordinary meanings but as an account of certain distributive principles for the basic structure of society. I assume that any reasonably complete ethical theory must include principles for this fundamental problem and that these principles, whatever they are, constitute its doctrine of justice. The concept of justice I take to be defined, then, by the role of its principles in assigning rights and duties and in defining the appropriate division of social advantages. A conception of justice is an interpretation of this role.

Now this approach may not seem to tally with tradition. I believe, though, that it does. The more specific sense that Aristotle gives to justice, and from which the most familiar formulations derive, is that of refraining from *pleonexia,* that is, from gaining some advantage for oneself by seizing what belongs to another, his property, his reward, his office, and the like, or by denying a person that which is due to him, the fulfillment of a promise, the repayment of a debt, the showing of proper respect, and so on.[3] It is evident that this definition is framed to apply to actions, and

3. *Nicomachean Ethics,* 1129b–1130b5. I have followed the interpretation of Gregory Vlastos, "Justice and Happiness in *The Republic,*" in *Plato: A Collection of Critical Essays,* edited by Vlastos (Garden City, N.Y., Doubleday and Company, 1971), vol. 2, pp. 70f. For a discussion of Aristotle on justice, see W. F. R. Hardie, *Aristotle's Ethical Theory* (Oxford, The Clarendon Press, 1968), ch. X.

persons are thought to be just insofar as they have, as one of the permanent elements of their character, a steady and effective desire to act justly. Aristotle's definition clearly presupposes, however, an account of what properly belongs to a person and of what is due to him. Now such entitlements are, I believe, very often derived from social institutions and the legitimate expectations to which they give rise. There is no reason to think that Aristotle would disagree with this, and certainly he has a conception of social justice to account for these claims. The definition I adopt is designed to apply directly to the most important case, the justice of the basic structure. There is no conflict with the traditional notion.

3. THE MAIN IDEA OF THE THEORY OF JUSTICE

My aim is to present a conception of justice which generalizes and carries to a higher level of abstraction the familiar theory of the social contract as found, say, in Locke, Rousseau, and Kant.[4] In order to do this we are not to think of the original contract as one to enter a particular society or to set up a particular form of government. Rather, the guiding idea is that the principles of justice for the basic structure of society are the object of the original agreement. They are the principles that free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association. These principles are to regulate all further agreements; they specify the kinds of social cooperation that can be entered into and the forms of government that can be established. This way of regarding the principles of justice I shall call justice as fairness.

Thus we are to imagine that those who engage in social cooperation choose together, in one joint act, the principles which are to assign basic rights and duties and to determine the division of social benefits. Men are to decide in advance how they are to regulate their claims against one another and what is to be the foundation charter of their society. Just as each person must decide by rational reflection what constitutes his good,

4. As the text suggests, I shall regard Locke's *Second Treatise of Government,* Rousseau's *The Social Contract,* and Kant's ethical works beginning with *The Foundations of the Metaphysics of Morals* as definitive of the contract tradition. For all of its greatness, Hobbes's *Leviathan* raises special problems. A general historical survey is provided by J. W. Gough, *The Social Contract,* 2nd ed. (Oxford, The Clarendon Press, 1957), and Otto Gierke, *Natural Law and the Theory of Society,* trans. with an introduction by Ernest Barker (Cambridge, The University Press, 1934). A presentation of the contract view as primarily an ethical theory is to be found in G. R. Grice, *The Grounds of Moral Judgment* (Cambridge, The University Press, 1967). See also §19, note 30.

that is, the system of ends which it is rational for him to pursue, so a group of persons must decide once and for all what is to count among them as just and unjust. The choice which rational men would make in this hypothetical situation of equal liberty, assuming for the present that this choice problem has a solution, determines the principles of justice.

In justice as fairness the original position of equality corresponds to the state of nature in the traditional theory of the social contract. This original position is not, of course, thought of as an actual historical state of affairs, much less as a primitive condition of culture. It is understood as a purely hypothetical situation characterized so as to lead to a certain conception of justice.[5] Among the essential features of this situation is that no one knows his place in society, his class position or social status, nor does any one know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like. I shall even assume that the parties do not know their conceptions of the good or their special psychological propensities. The principles of justice are chosen behind a veil of ignorance. This ensures that no one is advantaged or disadvantaged in the choice of principles by the outcome of natural chance or the contingency of social circumstances. Since all are similarly situated and no one is able to design principles to favor his particular condition, the principles of justice are the result of a fair agreement or bargain. For given the circumstances of the original position, the symmetry of everyone's relations to each other, this initial situation is fair between individuals as moral persons, that is, as rational beings with their own ends and capable, I shall assume, of a sense of justice. The original position is, one might say, the appropriate initial status quo, and thus the fundamental agreements reached in it are fair. This explains the propriety of the name "justice as fairness": it conveys the idea that the principles of justice are agreed to in an initial situation that is fair. The name does not mean that the concepts of justice and fairness are the same, any more than the phrase "poetry as metaphor" means that the concepts of poetry and metaphor are the same.

Justice as fairness begins, as I have said, with one of the most general of all choices which persons might make together, namely, with the

5. Kant is clear that the original agreement is hypothetical. See *The Metaphysics of Morals,* pt. I *(Rechtslehre),* especially §§47, 52; and pt. II of the essay "Concerning the Common Saying: This May Be True in Theory but It Does Not Apply in Practice," in *Kant's Political Writings,* ed. Hans Reiss and trans. by H. B. Nisbet (Cambridge, The University Press, 1970), pp. 73–87. See Georges Vlachos, *La Pensée politique de Kant* (Paris, Presses Universitaires de France, 1962), pp. 326–335; and J. G. Murphy, *Kant: The Philosophy of Right* (London, Macmillan, 1970), pp. 109–112, 133–136, for a further discussion.

choice of the first principles of a conception of justice which is to regulate all subsequent criticism and reform of institutions. Then, having chosen a conception of justice, we can suppose that they are to choose a constitution and a legislature to enact laws, and so on, all in accordance with the principles of justice initially agreed upon. Our social situation is just if it is such that by this sequence of hypothetical agreements we would have contracted into the general system of rules which defines it. Moreover, assuming that the original position does determine a set of principles (that is, that a particular conception of justice would be chosen), it will then be true that whenever social institutions satisfy these principles those engaged in them can say to one another that they are cooperating on terms to which they would agree if they were free and equal persons whose relations with respect to one another were fair. They could all view their arrangements as meeting the stipulations which they would acknowledge in an initial situation that embodies widely accepted and reasonable constraints on the choice of principles. The general recognition of this fact would provide the basis for a public acceptance of the corresponding principles of justice. No society can, of course, be a scheme of cooperation which men enter voluntarily in a literal sense; each person finds himself placed at birth in some particular position in some particular society, and the nature of this position materially affects his life prospects. Yet a society satisfying the principles of justice as fairness comes as close as a society can to being a voluntary scheme, for it meets the principles which free and equal persons would assent to under circumstances that are fair. In this sense its members are autonomous and the obligations they recognize self-imposed.

One feature of justice as fairness is to think of the parties in the initial situation as rational and mutually disinterested. This does not mean that the parties are egoists, that is, individuals with only certain kinds of interests, say in wealth, prestige, and domination. But they are conceived as not taking an interest in one another's interests. They are to presume that even their spiritual aims may be opposed, in the way that the aims of those of different religions may be opposed. Moreover, the concept of rationality must be interpreted as far as possible in the narrow sense, standard in economic theory, of taking the most effective means to given ends. I shall modify this concept to some extent, as explained later (§25), but one must try to avoid introducing into it any controversial ethical elements. The initial situation must be characterized by stipulations that are widely accepted.

In working out the conception of justice as fairness one main task

clearly is to determine which principles of justice would be chosen in the original position. To do this we must describe this situation in some detail and formulate with care the problem of choice which it presents. These matters I shall take up in the immediately succeeding chapters. It may be observed, however, that once the principles of justice are thought of as arising from an original agreement in a situation of equality, it is an open question whether the principle of utility would be acknowledged. Offhand it hardly seems likely that persons who view themselves as equals, entitled to press their claims upon one another, would agree to a principle which may require lesser life prospects for some simply for the sake of a greater sum of advantages enjoyed by others. Since each desires to protect his interests, his capacity to advance his conception of the good, no one has a reason to acquiesce in an enduring loss for himself in order to bring about a greater net balance of satisfaction. In the absence of strong and lasting benevolent impulses, a rational man would not accept a basic structure merely because it maximized the algebraic sum of advantages irrespective of its permanent effects on his own basic rights and interests. Thus it seems that the principle of utility is incompatible with the conception of social cooperation among equals for mutual advantage. It appears to be inconsistent with the idea of reciprocity implicit in the notion of a well-ordered society. Or, at any rate, so I shall argue.

I shall maintain instead that the persons in the initial situation would choose two rather different principles: the first requires equality in the assignment of basic rights and duties, while the second holds that social and economic inequalities, for example inequalities of wealth and authority, are just only if they result in compensating benefits for everyone, and in particular for the least advantaged members of society. These principles rule out justifying institutions on the grounds that the hardships of some are offset by a greater good in the aggregate. It may be expedient but it is not just that some should have less in order that others may prosper. But there is no injustice in the greater benefits earned by a few provided that the situation of persons not so fortunate is thereby improved. The intuitive idea is that since everyone's well-being depends upon a scheme of cooperation without which no one could have a satisfactory life, the division of advantages should be such as to draw forth the willing cooperation of everyone taking part in it, including those less well situated. The two principles mentioned seem to be a fair basis on which those better endowed, or more fortunate in their social position, neither of which we can be said to deserve, could expect the willing cooperation of others when some workable scheme is a necessary condition of the wel-

fare of all.[6] Once we decide to look for a conception of justice that prevents the use of the accidents of natural endowment and the contingencies of social circumstance as counters in a quest for political and economic advantage, we are led to these principles. They express the result of leaving aside those aspects of the social world that seem arbitrary from a moral point of view.

The problem of the choice of principles, however, is extremely difficult. I do not expect the answer I shall suggest to be convincing to everyone. It is, therefore, worth noting from the outset that justice as fairness, like other contract views, consists of two parts: (1) an interpretation of the initial situation and of the problem of choice posed there, and (2) a set of principles which, it is argued, would be agreed to. One may accept the first part of the theory (or some variant thereof), but not the other, and conversely. The concept of the initial contractual situation may seem reasonable although the particular principles proposed are rejected. To be sure, I want to maintain that the most appropriate conception of this situation does lead to principles of justice contrary to utilitarianism and perfectionism, and therefore that the contract doctrine provides an alternative to these views. Still, one may dispute this contention even though one grants that the contractarian method is a useful way of studying ethical theories and of setting forth their underlying assumptions.

Justice as fairness is an example of what I have called a contract theory. Now there may be an objection to the term "contract" and related expressions, but I think it will serve reasonably well. Many words have misleading connotations which at first are likely to confuse. The terms "utility" and "utilitarianism" are surely no exception. They too have unfortunate suggestions which hostile critics have been willing to exploit; yet they are clear enough for those prepared to study utilitarian doctrine. The same should be true of the term "contract" applied to moral theories. As I have mentioned, to understand it one has to keep in mind that it implies a certain level of abstraction. In particular, the content of the relevant agreement is not to enter a given society or to adopt a given form of government, but to accept certain moral principles. Moreover, the undertakings referred to are purely hypothetical: a contract view holds that certain principles would be accepted in a well-defined initial situation.

The merit of the contract terminology is that it conveys the idea that principles of justice may be conceived as principles that would be chosen by rational persons, and that in this way conceptions of justice may be

6. For the formulation of this intuitive idea I am indebted to Allan Gibbard.

explained and justified. The theory of justice is a part, perhaps the most significant part, of the theory of rational choice. Furthermore, principles of justice deal with conflicting claims upon the advantages won by social cooperation; they apply to the relations among several persons or groups. The word "contract" suggests this plurality as well as the condition that the appropriate division of advantages must be in accordance with principles acceptable to all parties. The condition of publicity for principles of justice is also connoted by the contract phraseology. Thus, if these principles are the outcome of an agreement, citizens have a knowledge of the principles that others follow. It is characteristic of contract theories to stress the public nature of political principles. Finally there is the long tradition of the contract doctrine. Expressing the tie with this line of thought helps to define ideas and accords with natural piety. There are then several advantages in the use of the term "contract." With due precautions taken, it should not be misleading.

A final remark. Justice as fairness is not a complete contract theory. For it is clear that the contractarian idea can be extended to the choice of more or less an entire ethical system, that is, to a system including principles for all the virtues and not only for justice. Now for the most part I shall consider only principles of justice and others closely related to them; I make no attempt to discuss the virtues in a systematic way. Obviously if justice as fairness succeeds reasonably well, a next step would be to study the more general view suggested by the name "rightness as fairness." But even this wider theory fails to embrace all moral relationships, since it would seem to include only our relations with other persons and to leave out of account how we are to conduct ourselves toward animals and the rest of nature. I do not contend that the contract notion offers a way to approach these questions which are certainly of the first importance; and I shall have to put them aside. We must recognize the limited scope of justice as fairness and of the general type of view that it exemplifies. How far its conclusions must be revised once these other matters are understood cannot be decided in advance.

4. THE ORIGINAL POSITION AND JUSTIFICATION

I have said that the original position is the appropriate initial status quo which insures that the fundamental agreements reached in it are fair. This fact yields the name "justice as fairness." It is clear, then, that I want to say that one conception of justice is more reasonable than another, or

justifiable with respect to it, if rational persons in the initial situation would choose its principles over those of the other for the role of justice. Conceptions of justice are to be ranked by their acceptability to persons so circumstanced. Understood in this way the question of justification is settled by working out a problem of deliberation: we have to ascertain which principles it would be rational to adopt given the contractual situation. This connects the theory of justice with the theory of rational choice.

If this view of the problem of justification is to succeed, we must, of course, describe in some detail the nature of this choice problem. A problem of rational decision has a definite answer only if we know the beliefs and interests of the parties, their relations with respect to one another, the alternatives between which they are to choose, the procedure whereby they make up their minds, and so on. As the circumstances are presented in different ways, correspondingly different principles are accepted. The concept of the original position, as I shall refer to it, is that of the most philosophically favored interpretation of this initial choice situation for the purposes of a theory of justice.

But how are we to decide what is the most favored interpretation? I assume, for one thing, that there is a broad measure of agreement that principles of justice should be chosen under certain conditions. To justify a particular description of the initial situation one shows that it incorporates these commonly shared presumptions. One argues from widely accepted but weak premises to more specific conclusions. Each of the presumptions should by itself be natural and plausible; some of them may seem innocuous or even trivial. The aim of the contract approach is to establish that taken together they impose significant bounds on acceptable principles of justice. The ideal outcome would be that these conditions determine a unique set of principles; but I shall be satisfied if they suffice to rank the main traditional conceptions of social justice.

One should not be misled, then, by the somewhat unusual conditions which characterize the original position. The idea here is simply to make vivid to ourselves the restrictions that it seems reasonable to impose on arguments for principles of justice, and therefore on these principles themselves. Thus it seems reasonable and generally acceptable that no one should be advantaged or disadvantaged by natural fortune or social circumstances in the choice of principles. It also seems widely agreed that it should be impossible to tailor principles to the circumstances of one's own case. We should insure further that particular inclinations and aspirations, and persons' conceptions of their good do not affect the prin-

ciples adopted. The aim is to rule out those principles that it would be rational to propose for acceptance, however little the chance of success, only if one knew certain things that are irrelevant from the standpoint of justice. For example, if a man knew that he was wealthy, he might find it rational to advance the principle that various taxes for welfare measures be counted unjust; if he knew that he was poor, he would most likely propose the contrary principle. To represent the desired restrictions one imagines a situation in which everyone is deprived of this sort of information. One excludes the knowledge of those contingencies which sets men at odds and allows them to be guided by their prejudices. In this manner the veil of ignorance is arrived at in a natural way. This concept should cause no difficulty if we keep in mind the constraints on arguments that it is meant to express. At any time we can enter the original position, so to speak, simply by following a certain procedure, namely, by arguing for principles of justice in accordance with these restrictions.

It seems reasonable to suppose that the parties in the original position are equal. That is, all have the same rights in the procedure for choosing principles; each can make proposals, submit reasons for their acceptance, and so on. Obviously the purpose of these conditions is to represent equality between human beings as moral persons, as creatures having a conception of their good and capable of a sense of justice. The basis of equality is taken to be similarity in these two respects. Systems of ends are not ranked in value; and each man is presumed to have the requisite ability to understand and to act upon whatever principles are adopted. Together with the veil of ignorance, these conditions define the principles of justice as those which rational persons concerned to advance their interests would consent to as equals when none are known to be advantaged or disadvantaged by social and natural contingencies.

There is, however, another side to justifying a particular description of the original position. This is to see if the principles which would be chosen match our considered convictions of justice or extend them in an acceptable way. We can note whether applying these principles would lead us to make the same judgments about the basic structure of society which we now make intuitively and in which we have the greatest confidence; or whether, in cases where our present judgments are in doubt and given with hesitation, these principles offer a resolution which we can affirm on reflection. There are questions which we feel sure must be answered in a certain way. For example, we are confident that religious intolerance and racial discrimination are unjust. We think that we have examined these things with care and have reached what we believe is an

impartial judgment not likely to be distorted by an excessive attention to our own interests. These convictions are provisional fixed points which we presume any conception of justice must fit. But we have much less assurance as to what is the correct distribution of wealth and authority. Here we may be looking for a way to remove our doubts. We can check an interpretation of the initial situation, then, by the capacity of its principles to accommodate our firmest convictions and to provide guidance where guidance is needed.

In searching for the most favored description of this situation we work from both ends. We begin by describing it so that it represents generally shared and preferably weak conditions. We then see if these conditions are strong enough to yield a significant set of principles. If not, we look for further premises equally reasonable. But if so, and these principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium.[7] It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation. At the moment everything is in order. But this equilibrium is not necessarily stable. It is liable to be upset by further examination of the conditions which should be imposed on the contractual situation and by particular cases which may lead us to revise our judgments. Yet for the time being we have done what we can to render coherent and to justify our convictions of social justice. We have reached a conception of the original position.

I shall not, of course, actually work through this process. Still, we may think of the interpretation of the original position that I shall present as the result of such a hypothetical course of reflection. It represents the

7. The process of mutual adjustment of principles and considered judgments is not peculiar to moral philosophy. See Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge, Mass., Harvard University Press, 1955), pp. 65–68, for parallel remarks concerning the justification of the principles of deductive and inductive inference.

attempt to accommodate within one scheme both reasonable philosophical conditions on principles as well as our considered judgments of justice. In arriving at the favored interpretation of the initial situation there is no point at which an appeal is made to self-evidence in the traditional sense either of general conceptions or particular convictions. I do not claim for the principles of justice proposed that they are necessary truths or derivable from such truths. A conception of justice cannot be deduced from self-evident premises or conditions on principles; instead, its justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view.

A final comment. We shall want to say that certain principles of justice are justified because they would be agreed to in an initial situation of equality. I have emphasized that this original position is purely hypothetical. It is natural to ask why, if this agreement is never actually entered into, we should take any interest in these principles, moral or otherwise. The answer is that the conditions embodied in the description of the original position are ones that we do in fact accept. Or if we do not, then perhaps we can be persuaded to do so by philosophical reflection. Each aspect of the contractual situation can be given supporting grounds. Thus what we shall do is to collect together into one conception a number of conditions on principles that we are ready upon due consideration to recognize as reasonable. These constraints express what we are prepared to regard as limits on fair terms of social cooperation. One way to look at the idea of the original position, therefore, is to see it as an expository device which sums up the meaning of these conditions and helps us to extract their consequences. On the other hand, this conception is also an intuitive notion that suggests its own elaboration, so that led on by it we are drawn to define more clearly the standpoint from which we can best interpret moral relationships. We need a conception that enables us to envision our objective from afar: the intuitive notion of the original position is to do this for us.[8]

5. CLASSICAL UTILITARIANISM

There are many forms of utilitarianism, and the development of the theory has continued in recent years. I shall not survey these forms here, nor

8. Henri Poincaré remarks: "Il nous faut une faculté qui nous fasse voir le but de loin, et, cette faculté, c'est l'intuition." *La Valeur de la science* (Paris, Flammarion, 1909), p. 27.

take account of the numerous refinements found in contemporary discussions. My aim is to work out a theory of justice that represents an alternative to utilitarian thought generally and so to all of these different versions of it. I believe that the contrast between the contract view and utilitarianism remains essentially the same in all these cases. Therefore I shall compare justice as fairness with familiar variants of intuitionism, perfectionism, and utilitarianism in order to bring out the underlying differences in the simplest way. With this end in mind, the kind of utilitarianism I shall describe here is the strict classical doctrine which receives perhaps its clearest and most accessible formulation in Sidgwick. The main idea is that society is rightly ordered, and therefore just, when its major institutions are arranged so as to achieve the greatest net balance of satisfaction summed over all the individuals belonging to it.[9]

We may note first that there is, indeed, a way of thinking of society which makes it easy to suppose that the most rational conception of jus-

9. I shall take Henry Sidgwick's *The Methods of Ethics,* 7th ed. (London, 1907), as summarizing the development of utilitarian moral theory. Book III of his *Principles of Political Economy* (London, 1883) applies this doctrine to questions of economic and social justice, and is a precursor of A. C. Pigou, *The Economics of Welfare* (London, Macmillan, 1920). Sidgwick's *Outlines of the History of Ethics,* 5th ed. (London, 1902), contains a brief history of the utilitarian tradition. We may follow him in assuming, somewhat arbitrarily, that it begins with Shaftesbury's *An Inquiry Concerning Virtue and Merit* (1711) and Hutcheson's *An Inquiry Concerning Moral Good and Evil* (1725). Hutcheson seems to have been the first to state clearly the principle of utility. He says in *Inquiry,* sec. 111, §8, that "that action is best, which procures the greatest happiness for the greatest numbers; and that, worst, which, in like manner, occasions misery." Other major eighteenth century works are Hume's *A Treatise of Human Nature* (1739), and *An Enquiry Concerning the Principles of Morals* (1751); Adam Smith's *A Theory of the Moral Sentiments* (1759); and Bentham's *The Principles of Morals and Legislation* (1789). To these we must add the writings of J. S. Mill represented by *Utilitarianism* (1863) and F. Y. Edgeworth's *Mathematical Psychics* (London, 1888).

The discussion of utilitarianism has taken a different turn in recent years by focusing on what we may call the coordination problem and related questions of publicity. This development stems from the essays of R. F. Harrod, "Utilitarianism Revised," *Mind,* vol. 45 (1936); J. D. Mabbott, "Punishment," *Mind,* vol. 48 (1939); Jonathan Harrison, "Utilitarianism, Universalisation, and Our Duty to Be Just," *Proceedings of the Aristotelian Society,* vol. 53 (1952–53); and J. O. Urmson, "The Interpretation of the Philosophy of J. S. Mill," *Philosophical Quarterly,* vol. 3 (1953). See also J. J. C. Smart, "Extreme and Restricted Utilitarianism," *Philosophical Quarterly,* vol. 6 (1956), and his *An Outline of a System of Utilitarian Ethics* (Cambridge, The University Press, 1961). For an account of these matters, see David Lyons, *Forms and Limits of Utilitarianism* (Oxford, The Clarendon Press, 1965); and Allan Gibbard, "Utilitarianisms and Coordination" (dissertation, Harvard University, 1971). The problems raised by these works, as important as they are, I shall leave aside as not bearing directly on the more elementary question of distribution which I wish to discuss.

Finally, we should note here the essays of J. C. Harsanyi, in particular, "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," *Journal of Political Economy,* 1953, and "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy,* 1955; and R. B. Brandt, "Some Merits of One Form of Rule-Utilitarianism," *University of Colorado Studies* (Boulder, Colorado, 1967). See below §§27–28.

tice is utilitarian. For consider: each man in realizing his own interests is certainly free to balance his own losses against his own gains. We may impose a sacrifice on ourselves now for the sake of a greater advantage later. A person quite properly acts, at least when others are not affected, to achieve his own greatest good, to advance his rational ends as far as possible. Now why should not a society act on precisely the same principle applied to the group and therefore regard that which is rational for one man as right for an association of men? Just as the well-being of a person is constructed from the series of satisfactions that are experienced at different moments in the course of his life, so in very much the same way the well-being of society is to be constructed from the fulfillment of the systems of desires of the many individuals who belong to it. Since the principle for an individual is to advance as far as possible his own welfare, his own system of desires, the principle for society is to advance as far as possible the welfare of the group, to realize to the greatest extent the comprehensive system of desire arrived at from the desires of its members. Just as an individual balances present and future gains against present and future losses, so a society may balance satisfactions and dissatisfactions between different individuals. And so by these reflections one reaches the principle of utility in a natural way: a society is properly arranged when its institutions maximize the net balance of satisfaction. The principle of choice for an association of men is interpreted as an extension of the principle of choice for one man. Social justice is the principle of rational prudence applied to an aggregative conception of the welfare of the group (§30).[10]

This idea is made all the more attractive by a further consideration. The two main concepts of ethics are those of the right and the good; the concept of a morally worthy person is, I believe, derived from them. The structure of an ethical theory is, then, largely determined by how it defines and connects these two basic notions. Now it seems that the simplest way of relating them is taken by teleological theories: the good is defined

10. On this point see also D. P. Gauthier, *Practical Reasoning* (Oxford, Clarendon Press, 1963), pp. 126f. The text elaborates the suggestion found in "Constitutional Liberty and the Concept of Justice," *Nomos VI: Justice,* ed. C. J. Friedrich and J. W. Chapman (New York, Atherton Press, 1963), pp. 124f, which in turn is related to the idea of justice as a higher-order administrative decision. See "Justice as Fairness," *Philosophical Review,* 1958, pp. 185–187. For references to utilitarians who explicitly affirm this extension, see §30, note 37. That the principle of social integration is distinct from the principle of personal integration is stated by R. B. Perry, *General Theory of Value* (New York, Longmans, Green, and Company, 1926), pp. 674–677. He attributes the error of overlooking this fact to Emile Durkheim and others with similar views. Perry's conception of social integration is that brought about by a shared and dominant benevolent purpose. See below, §24.

independently from the right, and then the right is defined as that which maximizes the good.[11] More precisely, those institutions and acts are right which of the available alternatives produce the most good, or at least as much good as any of the other institutions and acts open as real possibilities (a rider needed when the maximal class is not a singleton). Teleological theories have a deep intuitive appeal since they seem to embody the idea of rationality. It is natural to think that rationality is maximizing something and that in morals it must be maximizing the good. Indeed, it is tempting to suppose that it is self-evident that things should be arranged so as to lead to the most good.

It is essential to keep in mind that in a teleological theory the good is defined independently from the right. This means two things. First, the theory accounts for our considered judgments as to which things are good (our judgments of value) as a separate class of judgments intuitively distinguishable by common sense, and then proposes the hypothesis that the right is maximizing the good as already specified. Second, the theory enables one to judge the goodness of things without referring to what is right. For example, if pleasure is said to be the sole good, then presumably pleasures can be recognized and ranked in value by criteria that do not presuppose any standards of right, or what we would normally think of as such. Whereas if the distribution of goods is also counted as a good, perhaps a higher order one, and the theory directs us to produce the most good (including the good of distribution among others), we no longer have a teleological view in the classical sense. The problem of distribution falls under the concept of right as one intuitively understands it, and so the theory lacks an independent definition of the good. The clarity and simplicity of classical teleological theories derives largely from the fact that they factor our moral judgments into two classes, the one being characterized separately while the other is then connected with it by a maximizing principle.

Teleological doctrines differ, pretty clearly, according to how the conception of the good is specified. If it is taken as the realization of human excellence in the various forms of culture, we have what may be called perfectionism. This notion is found in Aristotle and Nietzsche, among others. If the good is defined as pleasure, we have hedonism; if as happiness, eudaimonism, and so on. I shall understand the principle of utility in its classical form as defining the good as the satisfaction of desire, or

11. Here I adopt W. K. Frankena's definition of teleological theories in *Ethics* (Englewood Cliffs, N.J., Prentice Hall, Inc., 1963), p. 13.

perhaps better, as the satisfaction of rational desire. This accords with the view in all essentials and provides, I believe, a fair interpretation of it. The appropriate terms of social cooperation are settled by whatever in the circumstances will achieve the greatest sum of satisfaction of the rational desires of individuals. It is impossible to deny the initial plausibility and attractiveness of this conception.

The striking feature of the utilitarian view of justice is that it does not matter, except indirectly, how this sum of satisfactions is distributed among individuals any more than it matters, except indirectly, how one man distributes his satisfactions over time. The correct distribution in either case is that which yields the maximum fulfillment. Society must allocate its means of satisfaction whatever these are, rights and duties, opportunities and privileges, and various forms of wealth, so as to achieve this maximum if it can. But in itself no distribution of satisfaction is better than another except that the more equal distribution is to be preferred to break ties.[12] It is true that certain common sense precepts of justice, particularly those which concern the protection of liberties and rights, or which express the claims of desert, seem to contradict this contention. But from a utilitarian standpoint the explanation of these precepts and of their seemingly stringent character is that they are those precepts which experience shows should be strictly respected and departed from only under exceptional circumstances if the sum of advantages is to be maximized.[13] Yet, as with all other precepts, those of justice are derivative from the one end of attaining the greatest balance of satisfaction. Thus there is no reason in principle why the greater gains of some should not compensate for the lesser losses of others; or more importantly, why the violation of the liberty of a few might not be made right by the greater good shared by many. It simply happens that under most conditions, at least in a reasonably advanced stage of civilization, the greatest sum of advantages is not attained in this way. No doubt the strictness of common sense precepts of justice has a certain usefulness in limiting men's propensities to injustice and to socially injurious actions, but the utilitarian believes that to affirm this strictness as a first principle of morals is a mistake. For just as it is rational for one man to maximize the fulfillment of his system of desires, it is right for a society to maximize the net balance of satisfaction taken over all of its members.

The most natural way, then, of arriving at utilitarianism (although not,

12. On this point see Sidgwick, *The Methods of Ethics,* pp. 416f.

13. See J. S. Mill, *Utilitarianism,* ch. V, last two pars.

of course, the only way of doing so) is to adopt for society as a whole the principle of rational choice for one man. Once this is recognized, the place of the impartial spectator and the emphasis on sympathy in the history of utilitarian thought is readily understood. For it is by the conception of the impartial spectator and the use of sympathetic identification in guiding our imagination that the principle for one man is applied to society. It is this spectator who is conceived as carrying out the required organization of the desires of all persons into one coherent system of desire; it is by this construction that many persons are fused into one. Endowed with ideal powers of sympathy and imagination, the impartial spectator is the perfectly rational individual who identifies with and experiences the desires of others as if these desires were his own. In this way he ascertains the intensity of these desires and assigns them their appropriate weight in the one system of desire the satisfaction of which the ideal legislator then tries to maximize by adjusting the rules of the social system. On this conception of society separate individuals are thought of as so many different lines along which rights and duties are to be assigned and scarce means of satisfaction allocated in accordance with rules so as to give the greatest fulfillment of wants. The nature of the decision made by the ideal legislator is not, therefore, materially different from that of an entrepreneur deciding how to maximize his profit by producing this or that commodity, or that of a consumer deciding how to maximize his satisfaction by the purchase of this or that collection of goods. In each case there is a single person whose system of desires determines the best allocation of limited means. The correct decision is essentially a question of efficient administration. This view of social cooperation is the consequence of extending to society the principle of choice for one man, and then, to make this extension work, conflating all persons into one through the imaginative acts of the impartial sympathetic spectator. Utilitarianism does not take seriously the distinction between persons.

6. SOME RELATED CONTRASTS

It has seemed to many philosophers, and it appears to be supported by the convictions of common sense, that we distinguish as a matter of principle between the claims of liberty and right on the one hand and the desirability of increasing aggregate social welfare on the other; and that we give a certain priority, if not absolute weight, to the former. Each member of society is thought to have an inviolability founded on justice or, as some

say, on natural right, which even the welfare of every one else cannot override. Justice denies that the loss of freedom for some is made right by a greater good shared by others. The reasoning which balances the gains and losses of different persons as if they were one person is excluded. Therefore in a just society the basic liberties are taken for granted and the rights secured by justice are not subject to political bargaining or to the calculus of social interests.

Justice as fairness attempts to account for these common sense convictions concerning the priority of justice by showing that they are the consequence of principles which would be chosen in the original position. These judgments reflect the rational preferences and the initial equality of the contracting parties. Although the utilitarian recognizes that, strictly speaking, his doctrine conflicts with these sentiments of justice, he maintains that common sense precepts of justice and notions of natural right have but a subordinate validity as secondary rules; they arise from the fact that under the conditions of civilized society there is great social utility in following them for the most part and in permitting violations only under exceptional circumstances. Even the excessive zeal with which we are apt to affirm these precepts and to appeal to these rights is itself granted a certain usefulness, since it counterbalances a natural human tendency to violate them in ways not sanctioned by utility. Once we understand this, the apparent disparity between the utilitarian principle and the strength of these persuasions of justice is no longer a philosophical difficulty. Thus while the contract doctrine accepts our convictions about the priority of justice as on the whole sound, utilitarianism seeks to account for them as a socially useful illusion.

A second contrast is that whereas the utilitarian extends to society the principle of choice for one man, justice as fairness, being a contract view, assumes that the principles of social choice, and so the principles of justice, are themselves the object of an original agreement. There is no reason to suppose that the principles which should regulate an association of men is simply an extension of the principle of choice for one man. On the contrary: if we assume that the correct regulative principle for anything depends on the nature of that thing, and that the plurality of distinct persons with separate systems of ends is an essential feature of human societies, we should not expect the principles of social choice to be utilitarian. To be sure, it has not been shown by anything said so far that the parties in the original position would not choose the principle of utility to define the terms of social cooperation. This is a difficult question which I shall examine later on. It is perfectly possible, from all that

one knows at this point, that some form of the principle of utility would be adopted, and therefore that contract theory leads eventually to a deeper and more roundabout justification of utilitarianism. In fact a derivation of this kind is sometimes suggested by Bentham and Edgeworth, although it is not developed by them in any systematic way and to my knowledge it is not found in Sidgwick.[14] For the present I shall simply assume that the persons in the original position would reject the utility principle and that they would adopt instead, for the kinds of reasons previously sketched, the two principles of justice already mentioned. In any case, from the standpoint of contract theory one cannot arrive at a principle of social choice merely by extending the principle of rational prudence to the system of desires constructed by the impartial spectator. To do this is not to take seriously the plurality and distinctness of individuals, nor to recognize as the basis of justice that to which men would consent. Here we may note a curious anomaly. It is customary to think of utilitarianism as individualistic, and certainly there are good reasons for this. The utilitarians were strong defenders of liberty and freedom of thought, and they held that the good of society is constituted by the advantages enjoyed by individuals. Yet utilitarianism is not individualistic, at least when arrived at by the more natural course of reflection, in that, by conflating all systems of desires, it applies to society the principle of choice for one man. And thus we see that the second contrast is related to the first, since it is this conflation, and the principle based upon it, which subjects the rights secured by justice to the calculus of social interests.

The last contrast that I shall mention now is that utilitarianism is a teleological theory whereas justice as fairness is not. By definition, then, the latter is a deontological theory, one that either does not specify the good independently from the right, or does not interpret the right as maximizing the good. (It should be noted that deontological theories are defined as non-teleological ones, not as views that characterize the rightness of institutions and acts independently from their consequences. All ethical doctrines worth our attention take consequences into account in judging rightness. One which did not would simply be irrational, crazy.) Justice as fairness is a deontological theory in the second way. For if it is assumed that the persons in the original position would choose a principle of equal liberty and restrict economic and social inequalities to those in

14. For Bentham see *The Principles of International Law,* Essay I, in *The Works of Jeremy Bentham,* ed. John Bowring (Edinburgh, 1838–1843), vol. 11, p. 537; for Edgeworth see *Mathematical Psychics,* pp. 52–56, and also the first pages of "The Pure Theory of Taxation," *Economic Journal,* vol. 7 (1897), where the same argument is presented more briefly. See below, §28.

everyone's interests, there is no reason to think that just institutions will maximize the good. (Here I suppose with utilitarianism that the good is defined as the satisfaction of rational desire.) Of course, it is not impossible that the most good is produced but it would be a coincidence. The question of attaining the greatest net balance of satisfaction never arises in justice as fairness; this maximum principle is not used at all.

There is a further point in this connection. In utilitarianism the satisfaction of any desire has some value in itself which must be taken into account in deciding what is right. In calculating the greatest balance of satisfaction it does not matter, except indirectly, what the desires are for.[15] We are to arrange institutions so as to obtain the greatest sum of satisfactions; we ask no questions about their source or quality but only how their satisfaction would affect the total of well-being. Social welfare depends directly and solely upon the levels of satisfaction or dissatisfaction of individuals. Thus if men take a certain pleasure in discriminating against one another, in subjecting others to a lesser liberty as a means of enhancing their self-respect, then the satisfaction of these desires must be weighed in our deliberations according to their intensity, or whatever, along with other desires. If society decides to deny them fulfillment, or to suppress them, it is because they tend to be socially destructive and a greater welfare can be achieved in other ways.

In justice as fairness, on the other hand, persons accept in advance a principle of equal liberty and they do this without a knowledge of their more particular ends. They implicitly agree, therefore, to conform their conceptions of their good to what the principles of justice require, or at least not to press claims which directly violate them. An individual who finds that he enjoys seeing others in positions of lesser liberty understands that he has no claim whatever to this enjoyment. The pleasure he takes in others' deprivations is wrong in itself: it is a satisfaction which requires the violation of a principle to which he would agree in the original position. The principles of right, and so of justice, put limits on which satisfactions have value; they impose restrictions on what are reasonable conceptions of one's good. In drawing up plans and in deciding on aspirations men are to take these constraints into account. Hence in justice as fairness one does not take men's propensities and inclinations as given, whatever they are, and then seek the best way to fulfill them. Rather, their desires and aspirations are restricted from the outset by the principles of justice which specify the boundaries that men's systems of

15. Bentham, *The Principles of Morals and Legislation,* ch. I, sec. IV.

ends must respect. We can express this by saying that in justice as fairness the concept of right is prior to that of the good. A just social system defines the scope within which individuals must develop their aims, and it provides a framework of rights and opportunities and the means of satisfaction within and by the use of which these ends may be equitably pursued. The priority of justice is accounted for, in part, by holding that the interests requiring the violation of justice have no value. Having no merit in the first place, they cannot override its claims.[16]

This priority of the right over the good in justice as fairness turns out to be a central feature of the conception. It imposes certain criteria on the design of the basic structure as a whole; these arrangements must not tend to generate propensities and attitudes contrary to the two principles of justice (that is, to certain principles which are given from the first a definite content) and they must insure that just institutions are stable. Thus certain initial bounds are placed upon what is good and what forms of character are morally worthy, and so upon what kinds of persons men should be. Now any theory of justice will set up some limits of this kind, namely, those that are required if its first principles are to be satisfied given the circumstances. Utilitarianism excludes those desires and propensities which if encouraged or permitted would, in view of the situation, lead to a lesser net balance of satisfaction. But this restriction is largely formal, and in the absence of fairly detailed knowledge of the circumstances it does not give much indication of what these desires and propensities are. This is not, by itself, an objection to utilitarianism. It is simply a feature of utilitarian doctrine that it relies very heavily upon the natural facts and contingencies of human life in determining what forms of moral character are to be encouraged in a just society. The moral ideal of justice as fairness is more deeply embedded in the first principles of the ethical theory. This is characteristic of natural rights views (the contractarian tradition) in comparison with the theory of utility.

In setting forth these contrasts between justice as fairness and utilitarianism, I have had in mind only the classical doctrine. This is the view of Bentham and Sidgwick and of the utilitarian economists Edgeworth and Pigou. The kind of utilitarianism espoused by Hume would not serve my purpose; indeed, it is not strictly speaking utilitarian. In his well-known arguments against Locke's contract theory, for example, Hume maintains

16. The priority of right is a central feature of Kant's ethics. See, for example, *The Critique of Practical Reason,* ch. II, bk. I of *pt. I,* esp. pp. 62–65 of vol. 5 of *Kants Gesammelte Schriften, Preussische Akademie der Wissenschaften* (Berlin, 1913). A clear statement is to be found in "Theory and Practice" (to abbreviate the title), *Political Writings,* pp. 67f.

that the principles of fidelity and allegiance both have the same foundation in utility, and therefore that nothing is gained from basing political obligation on an original contract. Locke's doctrine represents, for Hume, an unnecessary shuffle: one might as well appeal directly to utility.[17] But all Hume seems to mean by utility is the general interests and necessities of society. The principles of fidelity and allegiance derive from utility in the sense that the maintenance of the social order is impossible unless these principles are generally respected. But then Hume assumes that each man stands to gain, as judged by his long-term advantage, when law and government conform to the precepts founded on utility. No mention is made of the gains of some outweighing the disadvantages of others. For Hume, then, utility seems to be identical with some form of the common good; institutions satisfy its demands when they are to everyone's interests, at least in the long run. Now if this interpretation of Hume is correct, there is offhand no conflict with the priority of justice and no incompatibility with Locke's contract doctrine. For the role of equal rights in Locke is precisely to ensure that the only permissible departures from the state of nature are those which respect these rights and serve the common interest. It is clear that all the transformations from the state of nature which Locke approves of satisfy this condition and are such that rational men concerned to advance their ends could consent to them in a state of equality. Hume nowhere disputes the propriety of these constraints. His critique of Locke's contract doctrine never denies, or even seems to recognize, its fundamental contention.

The merit of the classical view as formulated by Bentham, Edgeworth, and Sidgwick is that it clearly recognizes what is at stake, namely, the relative priority of the principles of justice and of the rights derived from these principles. The question is whether the imposition of disadvantages on a few can be outweighed by a greater sum of advantages enjoyed by others; or whether the weight of justice requires an equal liberty for all and permits only those economic and social inequalities which are to each person's interests. Implicit in the contrasts between classical utilitarianism and justice as fairness is a difference in the underlying conceptions of society. In the one we think of a well-ordered society as a scheme of cooperation for reciprocal advantage regulated by principles which persons would choose in an initial situation that is fair, in the other as the efficient administration of social resources to maximize the satisfaction

17. "Of the Original Contract," *Essays: Moral, Political, and Literary,* ed. T. H. Green and T. H. Grose, vol. 1 (London, 1875), pp. 454f.

of the system of desire constructed by the impartial spectator from the many individual systems of desires accepted as given. The comparison with classical utilitarianism in its more natural derivation brings out this contrast.

7. INTUITIONISM

I shall think of intuitionism in a more general way than is customary: namely, as the doctrine that there is an irreducible family of first principles which have to be weighed against one another by asking ourselves which balance, in our considered judgment, is the most just. Once we reach a certain level of generality, the intuitionist maintains that there exist no higher-order constructive criteria for determining the proper emphasis for the competing principles of justice. While the complexity of the moral facts requires a number of distinct principles, there is no single standard that accounts for them or assigns them their weights. Intuitionist theories, then, have two features: first, they consist of a plurality of first principles which may conflict to give contrary directives in particular types of cases; and second, they include no explicit method, no priority rules, for weighing these principles against one another: we are simply to strike a balance by intuition, by what seems to us most nearly right. Or if there are priority rules, these are thought to be more or less trivial and of no substantial assistance in reaching a judgment.[18]

Various other contentions are commonly associated with intuitionism, for example, that the concepts of the right and the good are unanalyz-

18. Intuitionist theories of this type are found in Brian Barry, *Political Argument* (London, Routledge and Kegan Paul, 1965), see esp. pp. 4–8, 286f; R. B. Brandt, *Ethical Theory* (Englewood Cliffs, N.J., Prentice-Hall, Inc. 1959), pp. 404, 426, 429f, where the principle of utility is combined with a principle of equality; and Nicholas Rescher, *Distributive Justice* (New York, Bobbs-Merrill, 1966), pp. 35–41, 115–121, where analogous restrictions are introduced by the concept of the effective average. Robert Nozick discusses some of the problems in developing this kind of intuitionism in "Moral Complications and Moral Structures," *Natural Law Forum,* vol. 13 (1968).

Intuitionism in the traditional sense includes certain epistemological theses, for example, those concerning the self-evidence and necessity of moral principles. Here representative works are G. E. Moore, *Principia Ethica* (Cambridge, The University Press, 1903), esp. chs. I and VI; H. A. Prichard's essays and lectures in *Moral Obligation* (Oxford, The Clarendon Press, 1949), especially the first essay, "Does Moral Philosophy Rest on a Mistake?" (1912); W. D. Ross, *The Right and the Good* (Oxford, The Clarendon Press, 1930), especially chs. I and II, and *The Foundations of Ethics* (Oxford, The Clarendon Press, 1939). See also the eighteenth century treatise by Richard Price, *A Review of the Principal Questions of Morals,* 3rd ed., 1787, ed. D. D. Raphael (Oxford, The Clarendon Press, 1948). For a recent discussion of this classical form of intuitionism, see H. J. McCloskey, *Meta-Ethics and Normative Ethics* (The Hague, Martinus Nijhoff, 1969).

able, that moral principles when suitably formulated express self-evident propositions about legitimate moral claims, and so on. But I shall leave these matters aside. These characteristic epistemological doctrines are not a necessary part of intuitionism as I understand it. Perhaps it would be better if we were to speak of intuitionism in this broad sense as pluralism. Still, a conception of justice can be pluralistic without requiring us to weigh its principles by intuition. It may contain the requisite priority rules. To emphasize the direct appeal to our considered judgment in the balancing of principles, it seems appropriate to think of intuitionism in this more general fashion. How far such a view is committed to certain epistemological theories is a separate question.

Now so understood, there are many kinds of intuitionism. Not only are our everyday notions of this type but so perhaps are most philosophical doctrines. One way of distinguishing between intuitionist views is by the level of generality of their principles. Common sense intuitionism takes the form of groups of rather specific precepts, each group applying to a particular problem of justice. There is a group of precepts which applies to the question of fair wages, another to that of taxation, still another to punishment, and so on. In arriving at the notion of a fair wage, say, we are to balance somehow various competing criteria, for example, the claims of skill, training, effort, responsibility, and the hazards of the job, as well as to make some allowance for need. No one presumably would decide by any one of these precepts alone, and some compromise between them must be struck. The determination of wages by existing institutions also represents, in effect, a particular weighting of these claims. This weighting, however, is normally influenced by the demands of different social interests and so by relative positions of power and influence. It may not, therefore, conform to any one's conception of a fair wage. This is particularly likely to be true since persons with different interests are likely to stress the criteria which advance their ends. Those with more ability and education are prone to emphasize the claims of skill and training, whereas those lacking these advantages urge the claim of need. But not only are our everyday ideas of justice influenced by our own situation, they are also strongly colored by custom and current expectations. And by what criteria are we to judge the justice of custom itself and the legitimacy of these expectations? To reach some measure of understanding and agreement which goes beyond a mere de facto resolution of competing interests and a reliance on existing conventions and established expectations, it is necessary to move to a more general scheme for determining the balance of precepts, or at least for confining it within narrower limits.

Thus we can consider the problems of justice by reference to certain ends of social policy. Yet this approach also is likely to rely on intuition, since it normally takes the form of balancing various economic and social objectives. For example, suppose that allocative efficiency, full employment, a larger national income, and its more equal distribution are accepted as social ends. Then, given the desired weighting of these aims, and the existing institutional setup, the precepts of fair wages, just taxation, and so on will receive their due emphasis. In order to achieve greater efficiency and equity, one may follow a policy which has the effect of stressing skill and effort in the payment of wages, leaving the precept of need to be handled in some other fashion, perhaps by welfare transfers. An intuitionism of social ends provides a basis for deciding whether the determination of fair wages makes sense in view of the taxes to be imposed. How we weigh the precepts in one group is adjusted to how we weigh them in another. In this way we have managed to introduce a certain coherence into our judgments of justice; we have moved beyond the narrow de facto compromise of interests to a wider view. Of course we are still left with an appeal to intuition in the balancing of the higher-order ends of policy themselves. Different weightings for these are not by any means trivial variations but often correspond to profoundly opposed political convictions.

The principles of philosophical conceptions are of the most general kind. Not only are they intended to account for the ends of social policy, but the emphasis assigned to these principles should correspondingly determine the balance of these ends. For purposes of illustration, let us discuss a rather simple yet familiar conception based on the aggregative-distributive dichotomy. It has two principles: the basic structure of society is to be designed first to produce the most good in the sense of the greatest net balance of satisfaction, and second to distribute satisfactions equally. Both principles have, of course, *ceteris paribus* clauses. The first principle, the principle of utility, acts in this case as a standard of efficiency, urging us to produce as large a total as we can, other things equal; whereas the second principle serves as a standard of justice constraining the pursuit of aggregate well-being and evening out of the distribution of advantages.

This conception is intuitionist because no priority rule is provided for determining how these two principles are to be balanced against each other. Widely different weights are consistent with accepting these principles. No doubt it is natural to make certain assumptions about how most people would in fact balance them. For one thing, at different combina-

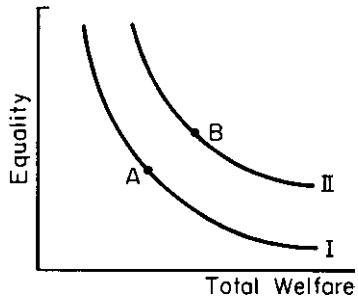


FIGURE 1

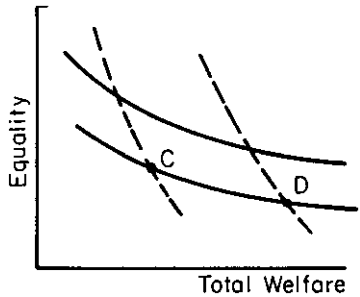


FIGURE 2

tions of total satisfaction and degrees of equality, we presumably would give these principles different weights. For example, if there is a large total satisfaction but it is unequally distributed, we would probably think it more urgent to increase equality than if the large aggregate well-being were already rather evenly shared. This can be put more formally by using the economist's device of indifference curves.[19] Assume that we can measure the extent to which particular arrangements of the basic structure satisfy these principles; and represent total satisfaction on the positive X-axis and equality on the positive Y-axis. (The latter may be supposed to have an upper bound at perfect equality.) The extent to which an arrangement of the basic structure fulfills these principles can now be represented by a point in the plane.

Now clearly a point which is northeast of another is a better arrangement: it is superior on both counts. For example, the point B is better than the point A in figure 1. Indifference curves are formed by connecting points judged equally just. Thus curve I in figure 1 consists of the points rated equally with point A which lies on that curve; curve II consists of the points ranked along with point B, and so on. We may assume that these curves slope downward to the right; and also that they do not intersect, otherwise the judgments they represent would be inconsistent. The slope of the curve at any point expresses the relative weights of equality and total satisfaction at the combination the point represents; the changing slope along an indifference curve shows how the relative urgency of the principles shifts as they are more or less satisfied. Thus, moving along either of the indifference curves in figure 1, we see that as equality de-

19. For the use of this device to illustrate intuitionist conceptions, see Barry, *Political Argument,* pp. 3–8. Most any book on demand theory or welfare economics will contain an exposition. W. J. Baumol, *Economic Theory and Operations Analysis,* 2nd ed. (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1965), ch. IX is an accessible account.

creases a larger and larger increase in the sum of satisfactions is required to compensate for a further decrease in equality.

Moreover, very different weightings are consistent with these principles. Let figure 2 represent the judgments of two different persons. The solid lines depict the judgments of the one who gives a relatively strong weight to equality, while the dashed lines depict the judgments of the other who gives a relatively strong weight to total welfare. Thus while the first person ranks arrangement D equal with C, the second judges D superior. This conception of justice imposes no limitations on what are the correct weightings; and therefore it allows different persons to arrive at a different balance of principles. Nevertheless such an intuitionist conception, if it were to fit our considered judgments on reflection, would be by no means without importance. At least it would single out the criteria which are significant, the apparent axes, so to speak, of our considered judgments of social justice. The intuitionist hopes that once these axes, or principles, are identified, men will in fact balance them more or less similarly, at least when they are impartial and not moved by an excessive attention to their own interests. Or if this is not so, then at least they can agree to some scheme whereby their assignment of weights can be compromised.

It is essential to observe that the intuitionist does not deny that we can describe how we balance competing principles, or how any one man does so, supposing that we weigh them differently. The intuitionist grants the possibility that these weights can be depicted by indifference curves. Knowing the description of these weights, the judgments which will be made can be foreseen. In this sense these judgments have a consistent and definite structure. Of course, it may be claimed that in the assignment of weights we are guided, without being aware of it, by certain further standards or by how best to realize a certain end. Perhaps the weights we assign are those which would result if we were to apply these standards or to pursue this end. Admittedly any given balancing of principles is subject to interpretation in this way. But the intuitionist claims that, in fact, there is no such interpretation. He contends that there exists no expressible ethical conception which underlies these weights. A geometrical figure or a mathematical function may describe them, but there are no constructive moral criteria that establish their reasonableness. Intuitionism holds that in our judgments of social justice we must eventually reach a plurality of first principles in regard to which we can only say that it seems to us more correct to balance them this way rather than that.

Now there is nothing intrinsically irrational about this intuitionist doc-

trine. Indeed, it may be true. We cannot take for granted that there must be a complete derivation of our judgments of social justice from recognizably ethical principles. The intuitionist believes to the contrary that the complexity of the moral facts defies our efforts to give a full account of our judgments and necessitates a plurality of competing principles. He contends that attempts to go beyond these principles either reduce to triviality, as when it is said that social justice is to give every man his due, or else lead to falsehood and oversimplification, as when one settles everything by the principle of utility. The only way therefore to dispute intuitionism is to set forth the recognizably ethical criteria that account for the weights which, in our considered judgments, we think appropriate to give to the plurality of principles. A refutation of intuitionism consists in presenting the sort of constructive criteria that are said not to exist. To be sure, the notion of a recognizably ethical principle is vague, although it is easy to give many examples drawn from tradition and common sense. But it is pointless to discuss this matter in the abstract. The intuitionist and his critic will have to settle this question once the latter has put forward his more systematic account.

It may be asked whether intuitionistic theories are teleological or deontological. They may be of either kind, and any ethical view is bound to rely on intuition to some degree at many points. For example, one could maintain, as Moore did, that personal affection and human understanding, the creation and the contemplation of beauty, and the gaining and appreciation of knowledge are the chief good things, along with pleasure.[20] And one might also maintain (as Moore did not) that these are the sole intrinsic goods. Since these values are specified independently from the right, we have a teleological theory of a perfectionist type if the right is defined as maximizing the good. Yet in estimating what yields the most good, the theory may hold that these values have to be balanced against each other by intuition: it may say that there are no substantive criteria for guidance here. Often, however, intuitionist theories are deontological. In the definitive presentation of Ross, the distribution of good things according to moral worth (distributive justice) is included among the goods to be advanced; and while the principle to produce the most good ranks as a first principle, it is but one such principle which must be balanced by intuition against the claims of the other prima facie principles.[21] The dis-

20. See *Principia Ethica,* ch. VI. The intuitionist nature of Moore's doctrine is assured by his principle of organic unity, pp. 27–31.

21. See W. D. Ross, *The Right and the Good,* pp. 21–27.

tinctive feature, then, of intuitionistic views is not their being teleological or deontological, but the especially prominent place that they give to the appeal to our intuitive capacities unguided by constructive and recognizably ethical criteria. Intuitionism denies that there exists any useful and explicit solution to the priority problem. I now turn to a brief discussion of this topic.

8. THE PRIORITY PROBLEM

We have seen that intuitionism raises the question of the extent to which it is possible to give a systematic account of our considered judgments of the just and the unjust. In particular, it holds that no constructive answer can be given to the problem of assigning weights to competing principles of justice. Here at least we must rely on our intuitive capacities. Classical utilitarianism tries, of course, to avoid the appeal to intuition altogether. It is a single-principle conception with one ultimate standard; the adjustment of weights is, in theory anyway, settled by reference to the principle of utility. Mill thought that there must be but one such standard, otherwise there would be no umpire between competing criteria, and Sidgwick argues at length that the utilitarian principle is the only one which can assume this role. They maintain that our moral judgments are implicitly utilitarian in the sense that when confronted with a clash of precepts, or with notions which are vague and imprecise, we have no alternative except to adopt utilitarianism. Mill and Sidgwick believe that at some point we must have a single principle to straighten out and to systematize our judgments.[22] Undeniably one of the great attractions of the classical doctrine is the way it faces the priority problem and tries to avoid relying on intuition.

As I have already remarked, there is nothing necessarily irrational in the appeal to intuition to settle questions of priority. We must recognize the possibility that there is no way to get beyond a plurality of principles. No doubt any conception of justice will have to rely on intuition to some degree. Nevertheless, we should do what we can to reduce the direct appeal to our considered judgments. For if men balance final principles

22. For Mill, see *A System of Logic,* bk. VI, ch. XII, §7; and *Utilitarianism,* ch. V, pars. 26–31, where this argument is made in connection with common sense precepts of justice. For Sidgwick, see *The Methods of Ethics,* for example, bk. IV, chs. II and III, which summarize much of the argument of bk. III.

differently, as presumably they often do, then their conceptions of justice are different. The assignment of weights is an essential and not a minor part of a conception of justice. If we cannot explain how these weights are to be determined by reasonable ethical criteria, the means of rational discussion have come to an end. An intuitionist conception of justice is, one might say, but half a conception. We should do what we can to formulate explicit principles for the priority problem, even though the dependence on intuition cannot be eliminated entirely.

In justice as fairness the role of intuition is limited in several ways. Since the whole question is rather difficult, I shall only make a few comments here the full sense of which will not be clear until later on. The first point is connected with the fact that the principles of justice are those which would be chosen in the original position. They are the outcome of a certain choice situation. Now being rational, the persons in the original position recognize that they should consider the priority of these principles. For if they wish to establish agreed standards for adjudicating their claims on one another, they will need principles for assigning weights. They cannot assume that their intuitive judgments of priority will in general be the same; given their different positions in society they surely will not. Thus I suppose that in the original position the parties try to reach some agreement as to how the principles of justice are to be balanced. Now part of the value of the notion of choosing principles is that the reasons which underlie their adoption in the first place may also support giving them certain weights. Since in justice as fairness the principles of justice are not thought of as self-evident, but have their justification in the fact that they would be chosen, we may find in the grounds for their acceptance some guidance or limitation as to how they are to be balanced. Given the situation of the original position, it may be clear that certain priority rules are preferable to others for much the same reasons that principles are initially assented to. By emphasizing the role of justice and the special features of the initial choice situation, the priority problem may prove more tractable.

A second possibility is that we may be able to find principles which can be put in what I shall call a serial or lexical order.[23] (The correct term

23. The term "lexicographical" derives from the fact that the most familiar example of such an ordering is that of words in a dictionary. To see this, substitute numerals for letters, putting "1" for "a" "2" for "b" and so on, and then rank the resulting strings of numerals from left to right, moving to the right only when necessary to break ties. In general, a lexical ordering cannot be represented by a continuous real-valued utility function; such a ranking violates the assumption of continuity. See I. F.

is "lexicographical," but it is too cumbersome.) This is an order which requires us to satisfy the first principle in the ordering before we can move on to the second, the second before we consider the third, and so on. A principle does not come into play until those previous to it are either fully met or do not apply. A serial ordering avoids, then, having to balance principles at all; those earlier in the ordering have an absolute weight, so to speak, with respect to later ones, and hold without exception. We can regard such a ranking as analogous to a sequence of constrained maximum principles. For we can suppose that any principle in the order is to be maximized subject to the condition that the preceding principles are fully satisfied. As an important special case I shall, in fact, propose an ordering of this kind by ranking the principle of equal liberty prior to the principle regulating economic and social inequalities. This means, in effect, that the basic structure of society is to arrange the inequalities of wealth and authority in ways consistent with the equal liberties required by the preceding principle. Certainly the concept of a lexical, or serial, order does not offhand seem very promising. Indeed, it appears to offend our sense of moderation and good judgment. Moreover, it presupposes that the principles in the order be of a rather special kind. For example, unless the earlier principles have but a limited application and establish definite requirements which can be fulfilled, later principles will never come into play. Thus the principle of equal liberty can assume a prior position since it may, let us suppose, be satisfied. Whereas if the

---

Pearce, *A Contribution to Demand Analysis* (Oxford, The Clarendon Press, 1946), pp. 22–27; and A. K. Sen, *Collective Choice and Social Welfare* (San Francisco, Holden-Day, 1970), pp. 34f. For further references, see H. S. Houthakker, "The Present State of Consumption Theory," *Econometrica,* vol. 29 (1961), pp. 710f.

In the history of moral philosophy the conception of a lexical order occasionally appears though it is not explicitly discussed. A clear example may be found in Hutcheson, *A System of Moral Philosophy* (1755). He proposes that in comparing pleasures of the same kind, we use their intensity and duration; in comparing pleasures of different kinds, we must consider their duration and dignity jointly. Pleasures of higher kinds may have a worth greater than those of lower kinds however great the latter's intensity and duration. See L. A. Selby-Bigge, *British Moralists,* vol. I (Oxford, 1897), pp. 421–423. J. S. Mill's well-known view in *Utilitarianism,* ch. II, pars. 6–8, is similar to Hutcheson's. It also is natural to rank moral worth as lexically prior to non-moral values. See for example Ross, *The Right and the Good,* pp. 149–154. And of course the primacy of justice noted in §1, as well as the priority of right as found in Kant, are further cases of such an ordering.

The theory of utility in economics began with an implicit recognition of the hierarchical structure of wants and the priority of moral considerations. This is clear in W. S. Jevons, *The Theory of Political Economy* (London, 1871), pp. 27–32. Jevons states a conception analogous to Hutcheson's and confines the economist's use of the utility calculus to the lowest rank of feelings. For a discussion of the hierarchy of wants and its relation to utility theory, see Nicholas Georgescu-Roegen, "Choice, Expectations, and Measurability," *Quarterly Journal of Economics,* vol. 68 (1954), esp. pp. 510–520.

principle of utility were first, it would render otiose all subsequent criteria. I shall try to show that at least in certain social circumstances a serial ordering of the principles of justice offers an approximate solution to the priority problem.

Finally, the dependence on intuition can be reduced by posing more limited questions and by substituting prudential for moral judgment. Thus someone faced with the principles of an intuitionist conception may reply that without some guidelines for deliberation he does not know what to say. He might maintain, for example, that he could not balance total utility against equality in the distribution of satisfaction. Not only are the notions involved here too abstract and comprehensive for him to have any confidence in his judgment, but there are enormous complications in interpreting what they mean. The aggregative-distributive dichotomy is no doubt an attractive idea, but in this instance it seems unmanageable. It does not factor the problem of social justice into small enough parts. In justice as fairness the appeal to intuition is focused in two ways. First we single out a certain position in the social system from which the system is to be judged, and then we ask whether, from the standpoint of a representative man in this position, it would be rational to prefer this arrangement of the basic structure rather than that. Given certain assumptions, economic and social inequalities are to be judged in terms of the long-run expectations of the least advantaged social group. Of course, the specification of this group is not very exact, and certainly our prudential judgments likewise give considerable scope to intuition, since we may not be able to formulate the principle which determines them. Nevertheless, we have asked a much more limited question and have substituted for an ethical judgment a judgment of rational prudence. Often it is quite clear how we should decide. The reliance on intuition is of a different nature and much less than in the aggregative-distributive dichotomy of the intuitionist conception.

In addressing the priority problem the task is that of reducing and not of eliminating entirely the reliance on intuitive judgments. There is no reason to suppose that we can avoid all appeals to intuition, of whatever kind, or that we should try to. The practical aim is to reach a reasonably reliable agreement in judgment in order to provide a common conception of justice. If men's intuitive priority judgments are similar, it does not matter, practically speaking, that they cannot formulate the principles which account for these convictions, or even whether such principles, exist. Contrary judgments, however, raise a difficulty, since the basis for adjudicating claims is to that extent obscure. Thus our object should be to

formulate a conception of justice which, however much it may call upon intuition, ethical or prudential, tends to make our considered judgments of justice converge. If such a conception does exist, then, from the standpoint of the original position, there would be strong reasons for accepting it, since it is rational to introduce further coherence into our common convictions of justice. Indeed, once we look at things from the standpoint of the initial situation, the priority problem is not that of how to cope with the complexity of already given moral facts which cannot be altered. Instead, it is the problem of formulating reasonable and generally acceptable proposals for bringing about the desired agreement in judgments. On a contract doctrine the moral facts are determined by the principles which would be chosen in the original position. These principles specify which considerations are relevant from the standpoint of social justice. Since it is up to the persons in the original position to choose these principles, it is for them to decide how simple or complex they want the moral facts to be. The original agreement settles how far they are prepared to compromise and to simplify in order to establish the priority rules necessary for a common conception of justice.

I have reviewed two obvious and simple ways of dealing constructively with the priority problem: namely, either by a single overall principle, or by a plurality of principles in lexical order. Other ways no doubt exist, but I shall not consider what they might be. The traditional moral theories are for the most part single-principled or intuitionistic, so that the working out of a serial ordering is novelty enough for a first step. While it seems clear that, in general, a lexical order cannot be strictly correct, it may be an illuminating approximation under certain special though significant conditions (§82). In this way it may indicate the larger structure of conceptions of justice and suggest the directions along which a closer fit can be found.

9. SOME REMARKS ABOUT MORAL THEORY

It seems desirable at this point, in order to prevent misunderstanding, to discuss briefly the nature of moral theory. I shall do this by explaining in more detail the concept of a considered judgment in reflective equilibrium and the reasons for introducing it.[24]

24. In this section I follow the general point of view of "Outline of a Procedure for Ethics," *Philosophical Review,* vol. 60 (1951).

Let us assume that each person beyond a certain age and possessed of the requisite intellectual capacity develops a sense of justice under normal social circumstances. We acquire a skill in judging things to be just and unjust, and in supporting these judgments by reasons. Moreover, we ordinarily have some desire to act in accord with these pronouncements and expect a similar desire on the part of others. Clearly this moral capacity is extraordinarily complex. To see this it suffices to note the potentially infinite number and variety of judgments that we are prepared to make. The fact that we often do not know what to say, and sometimes find our minds unsettled, does not detract from the complexity of the capacity we have.

Now one may think of moral theory at first (and I stress the provisional nature of this view) as the attempt to describe our moral capacity; or, in the present case, one may regard a theory of justice as describing our sense of justice. By such a description is not meant simply a list of the judgments on institutions and actions that we are prepared to render, accompanied with supporting reasons when these are offered. Rather, what is required is a formulation of a set of principles which, when conjoined to our beliefs and knowledge of the circumstances, would lead us to make these judgments with their supporting reasons were we to apply these principles conscientiously and intelligently. A conception of justice characterizes our moral sensibility when the everyday judgments we do make are in accordance with its principles. These principles can serve as part of the premises of an argument which arrives at the matching judgments. We do not understand our sense of justice until we know in some systematic way covering a wide range of cases what these principles are.

A useful comparison here is with the problem of describing the sense of grammaticalness that we have for the sentences of our native language.[25] In this case the aim is to characterize the ability to recognize well-formed sentences by formulating clearly expressed principles which make the same discriminations as the native speaker. This undertaking is known to require theoretical constructions that far outrun the ad hoc precepts of our explicit grammatical knowledge. A similar situation presumably holds in moral theory. There is no reason to assume that our sense of justice can be adequately characterized by familiar common sense precepts, or derived from the more obvious learning principles. A

25. See Noam Chomsky, *Aspects of the Theory of Syntax* (Cambridge, Mass., The M.I.T. Press, 1965), pp. 3–9.

correct account of moral capacities will certainly involve principles and theoretical constructions which go much beyond the norms and standards cited in everyday life; it may eventually require fairly sophisticated mathematics as well. Thus the idea of the original position and of an agreement on principles there does not seem too complicated or unnecessary. Indeed, these notions are rather simple and can serve only as a beginning.

So far, though, I have not said anything about considered judgments. Now, as already suggested, they enter as those judgments in which our moral capacities are most likely to be displayed without distortion. Thus in deciding which of our judgments to take into account we may reasonably select some and exclude others. For example, we can discard those judgments made with hesitation, or in which we have little confidence. Similarly, those given when we are upset or frightened, or when we stand to gain one way or the other can be left aside. All these judgments are likely to be erroneous or to be influenced by an excessive attention to our own interests. Considered judgments are simply those rendered under conditions favorable to the exercise of the sense of justice, and therefore in circumstances where the more common excuses and explanations for making a mistake do not obtain. The person making the judgment is presumed, then, to have the ability, the opportunity, and the desire to reach a correct decision (or at least, not the desire not to). Moreover, the criteria that identify these judgments are not arbitrary. They are, in fact, similar to those that single out considered judgments of any kind. And once we regard the sense of justice as a mental capacity, as involving the exercise of thought, the relevant judgments are those given under conditions favorable for deliberation and judgment in general.

I now turn to the notion of reflective equilibrium. The need for this idea arises as follows. According to the provisional aim of moral philosophy, one might say that justice as fairness is the hypothesis that the principles which would be chosen in the original position are identical with those that match our considered judgments and so these principles describe our sense of justice. But this interpretation is clearly oversimplified. In describing our sense of justice an allowance must be made for the likelihood that considered judgments are no doubt subject to certain irregularities and distortions despite the fact that they are rendered under favorable circumstances. When a person is presented with an intuitively appealing account of his sense of justice (one, say, which embodies various reasonable and natural presumptions), he may well revise his judgments to conform to its principles even though the theory does not fit

his existing judgments exactly. He is especially likely to do this if he can find an explanation for the deviations which undermines his confidence in his original judgments and if the conception presented yields a judgment which he finds he can now accept. From the standpoint of moral theory, the best account of a person's sense of justice is not the one which fits his judgments prior to his examining any conception of justice, but rather the one which matches his judgments in reflective equilibrium. As we have seen, this state is one reached after a person has weighed various proposed conceptions and he has either revised his judgments to accord with one of them or held fast to his initial convictions (and the corresponding conception).

There are, however, several interpretations of reflective equilibrium. For the notion varies depending upon whether one is to be presented with only those descriptions which more or less match one's existing judgments except for minor discrepancies, or whether one is to be presented with all possible descriptions to which one might plausibly conform one's judgments together with all relevant philosophical arguments for them. In the first case we would be describing a person's sense of justice more or less as it is although allowing for the smoothing out of certain irregularities; in the second case a person's sense of justice may or may not undergo a radical shift. Clearly it is the second kind of reflective equilibrium that one is concerned with in moral philosophy. To be sure, it is doubtful whether one can ever reach this state. For even if the idea of all possible descriptions and of all philosophically relevant arguments is well-defined (which is questionable), we cannot examine each of them. The most we can do is to study the conceptions of justice known to us through the tradition of moral philosophy and any further ones that occur to us, and then to consider these. This is pretty much what I shall do, since in presenting justice as fairness I shall compare its principles and arguments with a few other familiar views. In light of these remarks, justice as fairness can be understood as saying that the two principles previously mentioned would be chosen in the original position in preference to other traditional conceptions of justice, for example, those of utility and perfection; and that these principles give a better match with our considered judgments on reflection than these recognized alternatives. Thus justice as fairness moves us closer to the philosophical ideal; it does not, of course, achieve it.

This explanation of reflective equilibrium suggests straightway a number of further questions. For example, does a reflective equilibrium (in the

sense of the philosophical ideal) exist? If so, is it unique? Even if it is unique, can it be reached? Perhaps the judgments from which we begin, or the course of reflection itself (or both), affect the resting point, if any, that we eventually achieve. It would be useless, however, to speculate about these matters here. They are far beyond our reach. I shall not even ask whether the principles that characterize one person's considered judgments are the same as those that characterize another's. I shall take for granted that these principles are either approximately the same for persons whose judgments are in reflective equilibrium, or if not, that their judgments divide along a few main lines represented by the family of traditional doctrines that I shall discuss. (Indeed, one person may find himself torn between opposing conceptions at the same time.) If men's conceptions of justice finally turn out to differ, the ways in which they do so is a matter of first importance. Of course we cannot know how these conceptions vary, or even whether they do, until we have a better account of their structure. And this we now lack, even in the case of one man, or homogeneous group of men. If we can characterize one (educated) person's sense of justice, we might have a good beginning toward a theory of justice. We may suppose that everyone has in himself the whole form of a moral conception. So for the purposes of this book, the views of the reader and the author are the only ones that count. The opinions of others are used only to clear our own heads.

I wish to stress that in its initial stages at least a theory of justice is precisely that, namely, a theory. It is a theory of the moral sentiments (to recall an eighteenth century title) setting out the principles governing our moral powers, or, more specifically, our sense of justice. There is a definite if limited class of facts against which conjectured principles can be checked, namely, our considered judgments in reflective equilibrium. A theory of justice is subject to the same rules of method as other theories. Definitions and analyses of meaning do not have a special place: definition is but one device used in setting up the general structure of theory. Once the whole framework is worked out, definitions have no distinct status and stand or fall with the theory itself. In any case, it is obviously impossible to develop a substantive theory of justice founded solely on truths of logic and definition. The analysis of moral concepts and the a priori, however traditionally understood, is too slender a basis. Moral theory must be free to use contingent assumptions and general facts as it pleases. There is no other way to give an account of our considered judgments in reflective equilibrium. This is the conception of the subject

adopted by most classical British writers through Sidgwick. I see no reason to depart from it.[26]

Moreover, if we can find an accurate account of our moral conceptions, then questions of meaning and justification may prove much easier to answer. Indeed some of them may no longer be real questions at all. Note, for example, the extraordinary deepening of our understanding of the meaning and justification of statements in logic and mathematics made possible by developments since Frege and Cantor. A knowledge of the fundamental structures of logic and set theory and their relation to mathematics has transformed the philosophy of these subjects in a way that conceptual analysis and linguistic investigations never could. One has only to observe the effect of the division of theories into those which are decidable and complete, undecidable yet complete, and neither complete nor decidable. The problem of meaning and truth in logic and mathematics is profoundly altered by the discovery of logical systems illustrating these concepts. Once the substantive content of moral conceptions is better understood, a similar transformation may occur. It is possible that convincing answers to questions of the meaning and justification of moral judgments can be found in no other way.

I wish, then, to stress the central place of the study of our substantive moral conceptions. But the corollary to recognizing their complexity is accepting the fact that our present theories are primitive and have grave defects. We need to be tolerant of simplifications if they reveal and approximate the general outlines of our judgments. Objections by way of counterexamples are to be made with care, since these may tell us only what we know already, namely that our theory is wrong somewhere. The important thing is to find out how often and how far it is wrong. All theories are presumably mistaken in places. The real question at any given time is which of the views already proposed is the best approximation overall. To ascertain this some grasp of the structure of rival theories

26. I believe that this view goes back in its essentials to Aristotle's procedure in the *Nicomachean Ethics.* See W. F. R. Hardie, *Aristotle's Ethical Theory,* ch. III, esp. pp. 37–45. And Sidgwick thought of the history of moral philosophy as a series of attempts to state "in full breadth and clearness those primary intuitions of Reason, by the scientific application of which the common moral thought of mankind may be at once systematized and corrected." *The Methods of Ethics,* pp. 373f. He takes for granted that philosophical reflection will lead to revisions in our considered judgments, and although there are elements of epistemological intuitionism in his doctrine, these are not given much weight when unsupported by systematic considerations. For an account of Sidgwick's methodology, see J. B. Schneewind, "First Principles and Common Sense Morality in Sidgwick's Ethics," *Archiv für Geschichte der Philosophie,* Bd. 45 (1963).

is surely necessary. It is for this reason that I have tried to classify and to discuss conceptions of justice by reference to their basic intuitive ideas, since these disclose the main differences between them.

In presenting justice as fairness I shall contrast it with utilitarianism. I do this for various reasons, partly as an expository device, partly because the several variants of the utilitarian view have long dominated our philosophical tradition and continue to do so. And this dominance has been maintained despite the persistent misgivings that utilitarianism so easily arouses. The explanation for this peculiar state of affairs lies, I believe, in the fact that no constructive alternative theory has been advanced which has the comparable virtues of clarity and system and which at the same time allays these doubts. Intuitionism is not constructive, perfectionism is unacceptable. My conjecture is that the contract doctrine properly worked out can fill this gap. I think justice as fairness an endeavor in this direction.

Of course the contract theory as I shall present it is subject to the strictures that we have just noted. It is no exception to the primitiveness that marks existing moral theories. It is disheartening, for example, how little can now be said about priority rules; and while a lexical ordering may serve fairly well for some important cases, I assume that it will not be completely satisfactory. Nevertheless, we are free to use simplifying devices, and this I have often done. We should view a theory of justice as a guiding framework designed to focus our moral sensibilities and to put before our intuitive capacities more limited and manageable questions for judgment. The principles of justice identify certain considerations as morally relevant and the priority rules indicate the appropriate precedence when these conflict, while the conception of the original position defines the underlying idea which is to inform our deliberations. If the scheme as a whole seems on reflection to clarify and to order our thoughts, and if it tends to reduce disagreements and to bring divergent convictions more in line, then it has done all that one may reasonably ask. Understood as parts of a framework that does indeed seem to help, the numerous simplifications may be regarded as provisionally justified.

CHAPTER II. THE PRINCIPLES OF JUSTICE

The theory of justice may be divided into two main parts: (1) an interpretation of the initial situation and a formulation of the various principles available for choice there, and (2) an argument establishing which of these principles would in fact be adopted. In this chapter two principles of justice for institutions and several principles for individuals are discussed and their meaning explained. Thus I am concerned for the present with only one aspect of the first part of the theory. Not until the next chapter do I take up the interpretation of the initial situation and begin the argument to show that the principles considered here would indeed be acknowledged. A variety of topics are discussed: institutions as subjects of justice and the concept of formal justice; three kinds of procedural justice; the place of the theory of the good; and the sense in which the principles of justice are egalitarian, among others. In each case the aim is to explain the meaning and application of the principles.

10. INSTITUTIONS AND FORMAL JUSTICE

The primary subject of the principles of social justice is the basic structure of society, the arrangement of major social institutions into one scheme of cooperation. We have seen that these principles are to govern the assignment of rights and duties in these institutions and they are to determine the appropriate distribution of the benefits and burdens of social life. The principles of justice for institutions must not be confused with the principles which apply to individuals and their actions in particular circumstances. These two kinds of principles apply to different subjects and must be discussed separately.

Now by an institution I shall understand a public system of rules which defines offices and positions with their rights and duties, powers and immunities, and the like. These rules specify certain forms of action as

permissible, others as forbidden; and they provide for certain penalties and defenses, and so on, when violations occur. As examples of institutions, or more generally social practices, we may think of games and rituals, trials and parliaments, markets and systems of property. An institution may be thought of in two ways: first as an abstract object, that is, as a possible form of conduct expressed by a system of rules; and second, as the realization in the thought and conduct of certain persons at a certain time and place of the actions specified by these rules. There is an ambiguity, then, as to which is just or unjust, the institution as realized or the institution as an abstract object. It seems best to say that it is the institution as realized and effectively and impartially administered which is just or unjust. The institution as an abstract object is just or unjust in the sense that any realization of it would be just or unjust.

An institution exists at a certain time and place when the actions specified by it are regularly carried out in accordance with a public understanding that the system of rules defining the institution is to be followed. Thus parliamentary institutions are defined by a certain system of rules (or family of such systems to allow for variations). These rules enumerate certain forms of action ranging from holding a session of parliament to taking a vote on a bill to raising a point of order. Various kinds of general norms are organized into a coherent scheme. A parliamentary institution exists at a certain time and place when certain people perform the appropriate actions, engage in these activities in the required way, with a reciprocal recognition of one another's understanding that their conduct accords with the rules they are to comply with.[1]

In saying that an institution, and therefore the basic structure of society, is a public system of rules, I mean then that everyone engaged in it knows what he would know if these rules and his participation in the activity they define were the result of an agreement. A person taking part in an institution knows what the rules demand of him and of the others. He also knows that the others know this and that they know that he knows this, and so on. To be sure, this condition is not always fulfilled in the case of actual institutions, but it is a reasonable simplifying assumption. The principles of justice are to apply to social arrangements understood to be public in this sense. Where the rules of a certain subpart of an institution are known only to those belonging to it, we may assume that there is an understanding that those in this part can make rules for themselves as

1. See H. L. A. Hart, *The Concept of Law* (Oxford, The Clarendon Press, 1961), pp. 59f, 106f, 109–114, for a discussion of when rules and legal systems may be said to exist.

long as these rules are designed to achieve ends generally accepted and others are not adversely affected. The publicity of the rules of an institution insures that those engaged in it know what limitations on conduct to expect of one another and what kinds of actions are permissible. There is a common basis for determining mutual expectations. Moreover, in a well-ordered society, one effectively regulated by a shared conception of justice, there is also a public understanding as to what is just and unjust. Later I assume that the principles of justice are chosen subject to the knowledge that they are to be public (§23). This condition is a natural one in a contractarian theory.

It is necessary to note the distinction between the constitutive rules of an institution, which establish its various rights and duties, and so on, and strategies and maxims for how best to take advantage of the institution for particular purposes.[2] Rational strategies and maxims are based upon an analysis of which permissible actions individuals and groups will decide upon in view of their interests, beliefs, and conjectures about one another's plans. These strategies and maxims are not themselves part of the institution. Rather they belong to the theory of it, for example, to the theory of parliamentary politics. Normally the theory of an institution, just as that of a game, takes the constitutive rules as given and analyzes the way in which power is distributed and explains how those engaged in it are likely to avail themselves of its opportunities. In designing and reforming social arrangements one must, of course, examine the schemes and tactics it allows and the forms of behavior which it tends to encourage. Ideally the rules should be set up so that men are led by their predominant interests to act in ways which further socially desirable ends. The conduct of individuals guided by their rational plans should be coordinated as far as possible to achieve results which although not intended or perhaps even foreseen by them are nevertheless the best ones from the standpoint of social justice. Bentham thinks of this coordination as the artificial identification of interests, Adam Smith as the work of the invisible hand.[3] It is the aim of the ideal legislator in enacting laws and of the moralist in urging their reform. Still, the strategies and tactics fol-

2. On constitutive rules and institutions, see J. R. Searle, *Speech Acts* (Cambridge, The University Press, 1969), pp. 33–42. See also G. E. M. Anscombe, "On Brute Facts," *Analysis,* vol. 18 (1958); and B. J. Diggs, "Rules and Utilitarianism," *American Philosophical Quarterly,* vol. 1 (1964), where various interpretations of rules are discussed.

3. The phrase "the artificial identification of interests" is from Elie Halévy's account of Bentham in *La Formation du radicalisme philosophique,* vol. 1 (Paris, Felix Alcan, 1901), pp. 20–24. On the invisible hand, see *The Wealth of Nations,* ed. Edwin Cannan (New York, The Modern Library, 1937), p. 423.

lowed by individuals, while essential to the assessment of institutions, are not part of the public systems of rules which define them.

We may also distinguish between a single rule (or group of rules), an institution (or a major part thereof), and the basic structure of the social system as a whole. The reason for doing this is that one or several rules of an arrangement may be unjust without the institution itself being so. Similarly, an institution may be unjust although the social system as a whole is not. There is the possibility not only that single rules and institutions are not by themselves sufficiently important but that within the structure of an institution or social system one apparent injustice compensates for another. The whole is less unjust than it would be if it contained but one of the unjust parts. Further, it is conceivable that a social system may be unjust even though none of its institutions are unjust taken separately: the injustice is a consequence of how they are combined together into a single system. One institution may encourage and appear to justify expectations which are denied or ignored by another. These distinctions are obvious enough. They simply reflect the fact that in appraising institutions we may view them in a wider or a narrower context.

There are, it should be remarked, institutions in regard to which the concept of justice does not ordinarily apply. A ritual, say, is not usually regarded as either just or unjust, although cases can no doubt be imagined in which this would not be true, for example, the ritual sacrifice of the first-born or of prisoners of war. A general theory of justice would consider when rituals and other practices not commonly thought of as just or unjust are indeed subject to this form of criticism. Presumably they must involve in some way the allocation among persons of certain rights and values. I shall not, however, pursue this larger inquiry. Our concern is solely with the basic structure of society and its major institutions and therefore with the standard cases of social justice.

Now let us suppose a certain basic structure to exist. Its rules satisfy a certain conception of justice. We may not ourselves accept its principles; we may even find them odious and unjust. But they are principles of justice in the sense that for this system they assume the role of justice: they provide an assignment of fundamental rights and duties and they determine the division of advantages from social cooperation. Let us also imagine that this conception of justice is by and large accepted in the society and that institutions are impartially and consistently administered by judges and other officials. That is, similar cases are treated similarly, the relevant similarities and differences being those identified by the existing norms. The correct rule as defined by institutions is regularly ad-

hered to and properly interpreted by the authorities. This impartial and consistent administration of laws and institutions, whatever their substantive principles, we may call formal justice. If we think of justice as always expressing a kind of equality, then formal justice requires that in their administration laws and institutions should apply equally (that is, in the same way) to those belonging to the classes defined by them. As Sidgwick emphasized, this sort of equality is implied in the very notion of a law or institution, once it is thought of as a scheme of general rules.[4] Formal justice is adherence to principle, or as some have said, obedience to system.[5]

It is obvious, Sidgwick adds, that law and institutions may be equally executed and yet be unjust. Treating similar cases similarly is not a sufficient guarantee of substantive justice. This depends upon the principles in accordance with which the basic structure is framed. There is no contradiction in supposing that a slave or caste society, or one sanctioning the most arbitrary forms of discrimination, is evenly and consistently administered, although this may be unlikely. Nevertheless, formal justice, or justice as regularity, excludes significant kinds of injustices. For if it is supposed that institutions are reasonably just, then it is of great importance that the authorities should be impartial and not influenced by personal, monetary, or other irrelevant considerations in their handling of particular cases. Formal justice in the case of legal institutions is simply an aspect of the rule of law which supports and secures legitimate expectations. One kind of injustice is the failure of judges and others in authority to adhere to the appropriate rules or interpretations thereof in deciding claims. A person is unjust to the extent that from character and inclination he is disposed to such actions. Moreover, even where laws and institutions are unjust, it is often better that they should be consistently applied. In this way those subject to them at least know what is demanded and they can try to protect themselves accordingly; whereas there is even greater injustice if those already disadvantaged are also arbitrarily treated in particular cases when the rules would give them some security. On the other hand, it might be still better in particular cases to alleviate the plight of those unfairly treated by departures from the existing norms. How far we are justified in doing this, especially at the expense of expectations founded in good faith on current institutions, is one of the tangled ques-

4. *The Methods of Ethics,* 7th ed. (London, Macmillan, 1907), p. 267.

5. See Ch. Perelman, *The Idea of Justice and the Problem of Argument,* trans. J. Petrie (London, Routledge and Kegan Paul, 1963), p. 41. All of the first two chapters, a translation of *De la Justice* (Brussels, 1943), is relevant here, but especially pp. 36–45.

tions of political justice. In general, all that can be said is that the strength of the claims of formal justice, of obedience to system, clearly depend upon the substantive justice of institutions and the possibilities of their reform.

Some have held that in fact substantive and formal justice tend to go together and therefore that at least grossly unjust institutions are never, or at any rate rarely, impartially and consistently administered.[6] Those who uphold and gain from unjust arrangements, and who deny with contempt the rights and liberties of others, are not likely, it is said, to let scruples concerning the rule of law interfere with their interests in particular cases. The inevitable vagueness of laws in general and the wide scope allowed for their interpretation encourages an arbitrariness in reaching decisions which only an allegiance to justice can allay. Thus it is maintained that where we find formal justice, the rule of law and the honoring of legitimate expectations, we are likely to find substantive justice as well. The desire to follow rules impartially and consistently, to treat similar cases similarly, and to accept the consequences of the application of public norms is intimately connected with the desire, or at least the willingness, to recognize the rights and liberties of others and to share fairly in the benefits and burdens of social cooperation. The one desire tends to be associated with the other. This contention is certainly plausible but I shall not examine it here. For it cannot be properly assessed until we know what are the most reasonable principles of substantive justice and under what conditions men come to affirm and to live by them. Once we understand the content of these principles and their basis in reason and human attitudes, we may be in a position to decide whether substantive and formal justice are tied together.

11. TWO PRINCIPLES OF JUSTICE

I shall now state in a provisional form the two principles of justice that I believe would be agreed to in the original position. The first formulation of these principles is tentative. As we go on I shall consider several formulations and approximate step by step the final statement to be given much later. I believe that doing this allows the exposition to proceed in a natural way.

6. See Lon Fuller, *The Morality of Law* (New Haven, Yale University Press, 1964), ch. IV.

The first statement of the two principles reads as follows.

First: each person is to have an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberties for others.

Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all.

There are two ambiguous phrases in the second principle, namely "everyone's advantage" and "open to all." Determining their sense more exactly will lead to a second formulation of the principle in §13. The final version of the two principles is given in §46; §39 considers the rendering of the first principle.

These principles primarily apply, as I have said, to the basic structure of society and govern the assignment of rights and duties and regulate the distribution of social and economic advantages. Their formulation presupposes that, for the purposes of a theory of justice, the social structure may be viewed as having two more or less distinct parts, the first principle applying to the one, the second principle to the other. Thus we distinguish between the aspects of the social system that define and secure the equal basic liberties and the aspects that specify and establish social and economic inequalities. Now it is essential to observe that the basic liberties are given by a list of such liberties. Important among these are political liberty (the right to vote and to hold public office) and freedom of speech and assembly; liberty of conscience and freedom of thought; freedom of the person, which includes freedom from psychological oppression and physical assault and dismemberment (integrity of the person); the right to hold personal property and freedom from arbitrary arrest and seizure as defined by the concept of the rule of law. These liberties are to be equal by the first principle.

The second principle applies, in the first approximation, to the distribution of income and wealth and to the design of organizations that make use of differences in authority and responsibility. While the distribution of wealth and income need not be equal, it must be to everyone's advantage, and at the same time, positions of authority and responsibility must be accessible to all. One applies the second principle by holding positions open, and then, subject to this constraint, arranges social and economic inequalities so that everyone benefits.

These principles are to be arranged in a serial order with the first principle prior to the second. This ordering means that infringements of

the basic equal liberties protected by the first principle cannot be justified, or compensated for, by greater social and economic advantages. These liberties have a central range of application within which they can be limited and compromised only when they conflict with other basic liberties. Since they may be limited when they clash with one another, none of these liberties is absolute; but however they are adjusted to form one system, this system is to be the same for all. It is difficult, and perhaps impossible, to give a complete specification of these liberties independently from the particular circumstances—social, economic, and technological—of a given society. The hypothesis is that the general form of such a list could be devised with sufficient exactness to sustain this conception of justice. Of course, liberties not on the list, for example, the right to own certain kinds of property (e.g., means of production) and freedom of contract as understood by the doctrine of laissez-faire are not basic; and so they are not protected by the priority of the first principle. Finally, in regard to the second principle, the distribution of wealth and income, and positions of authority and responsibility, are to be consistent with both the basic liberties and equality of opportunity.

The two principles are rather specific in their content, and their acceptance rests on certain assumptions that I must eventually try to explain and justify. For the present, it should be observed that these principles are a special case of a more general conception of justice that can be expressed as follows.

> All social values—liberty and opportunity, income and wealth, and the social bases of self-respect—are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage.

Injustice, then, is simply inequalities that are not to the benefit of all. Of course, this conception is extremely vague and requires interpretation.

As a first step, suppose that the basic structure of society distributes certain primary goods, that is, things that every rational man is presumed to want. These goods normally have a use whatever a person's rational plan of life. For simplicity, assume that the chief primary goods at the disposition of society are rights, liberties, and opportunities, and income and wealth. (Later on in Part Three the primary good of self-respect has a central place.) These are the social primary goods. Other primary goods such as health and vigor, intelligence and imagination, are natural goods; although their possession is influenced by the basic structure, they are not so directly under its control. Imagine, then, a hypothetical initial arrange-

ment in which all the social primary goods are equally distributed: everyone has similar rights and duties, and income and wealth are evenly shared. This state of affairs provides a benchmark for judging improvements. If certain inequalities of wealth and differences in authority would make everyone better off than in this hypothetical starting situation, then they accord with the general conception.

Now it is possible, at least theoretically, that by giving up some of their fundamental liberties men are sufficiently compensated by the resulting social and economic gains. The general conception of justice imposes no restrictions on what sort of inequalities are permissible; it only requires that everyone's position be improved. We need not suppose anything so drastic as consenting to a condition of slavery. Imagine instead that people seem willing to forego certain political rights when the economic returns are significant. It is this kind of exchange which the two principles rule out; being arranged in serial order they do not permit exchanges between basic liberties and economic and social gains except under extenuating circumstances (§§26, 39).

For the most part, I shall leave aside the general conception of justice and examine instead the two principles in serial order. The advantage of this procedure is that from the first the matter of priorities is recognized and an effort made to find principles to deal with it. One is led to attend throughout to the conditions under which the absolute weight of liberty with respect to social and economic advantages, as defined by the lexical order of the two principles, would be reasonable. Offhand, this ranking appears extreme and too special a case to be of much interest; but there is more justification for it than would appear at first sight. Or at any rate, so I shall maintain (§82). Furthermore, the distinction between fundamental rights and liberties and economic and social benefits marks a difference among primary social goods that suggests an important division in the social system. Of course, the distinctions drawn and the ordering proposed are at best only approximations. There are surely circumstances in which they fail. But it is essential to depict clearly the main lines of a reasonable conception of justice; and under many conditions anyway, the two principles in serial order may serve well enough.

The fact that the two principles apply to institutions has certain consequences. First of all, the rights and basic liberties referred to by these principles are those which are defined by the public rules of the basic structure. Whether men are free is determined by the rights and duties established by the major institutions of society. Liberty is a certain pattern

of social forms. The first principle simply requires that certain sorts of rules, those defining basic liberties, apply to everyone equally and that they allow the most extensive liberty compatible with a like liberty for all. The only reason for circumscribing basic liberties and making them less extensive is that otherwise they would interfere with one another.

Further, when principles mention persons, or require that everyone gain from an inequality, the reference is to representative persons holding the various social positions, or offices established by the basic structure. Thus in applying the second principle I assume that it is possible to assign an expectation of well-being to representative individuals holding these positions. This expectation indicates their life prospects as viewed from their social station. In general, the expectations of representative persons depend upon the distribution of rights and duties throughout the basic structure. Expectations are connected: by raising the prospects of the representative man in one position we presumably increase or decrease the prospects of representative men in other positions. Since it applies to institutional forms, the second principle (or rather the first part of it) refers to the expectations of representative individuals. As I shall discuss below (§14), neither principle applies to distributions of particular goods to particular individuals who may be identified by their proper names. The situation where someone is considering how to allocate certain commodities to needy persons who are known to him is not within the scope of the principles. They are meant to regulate basic institutional arrangements. We must not assume that there is much similarity from the standpoint of justice between an administrative allotment of goods to specific persons and the appropriate design of society. Our common sense intuitions for the former may be a poor guide to the latter.

Now the second principle insists that each person benefit from permissible inequalities in the basic structure. This means that it must be reasonable for each relevant representative man defined by this structure, when he views it as a going concern, to prefer his prospects with the inequality to his prospects without it. One is not allowed to justify differences in income or in positions of authority and responsibility on the ground that the disadvantages of those in one position are outweighed by the greater advantages of those in another. Much less can infringements of liberty be counterbalanced in this way. It is obvious, however, that there are indefinitely many ways in which all may be advantaged when the initial arrangement of equality is taken as a benchmark. How then are we to choose among these possibilities? The principles must be specified so that they yield a determinate conclusion. I now turn to this problem.

12. INTERPRETATIONS OF THE SECOND PRINCIPLE

I have already mentioned that since the phrases "everyone's advantage" and "equally open to all" are ambiguous, both parts of the second principle have two natural senses. Because these senses are independent of one another, the principle has four possible meanings. Assuming that the first principle of equal liberty has the same sense throughout, we then have four interpretations of the two principles. These are indicated in the table below.

	"Everyone's advantage"	
"Equally open"	Principle of efficiency	Difference principle
Equality as careers open to talents	System of Natural Liberty	Natural Aristocracy
Equality as equality of fair opportunity	Liberal Equality	Democratic Equality

I shall sketch in turn these three interpretations: the system of natural liberty, liberal equality, and democratic equality. In some respects this sequence is the more intuitive one, but the sequence via the interpretation of natural aristocracy is not without interest and I shall comment on it briefly. In working out justice as fairness, we must decide which interpretation is to be preferred. I shall adopt that of democratic equality, explaining in the next section what this notion means. The argument for its acceptance in the original position does not begin until the next chapter.

The first interpretation (in either sequence) I shall refer to as the system of natural liberty. In this rendering the first part of the second principle is understood as the principle of efficiency adjusted so as to apply to institutions or, in this case, to the basic structure of society; and the second part is understood as an open social system in which, to use the traditional phrase, careers are open to talents. I assume in all interpretations that the first principle of equal liberty is satisfied and that the economy is roughly a free market system, although the means of production may or may not be privately owned. The system of natural liberty asserts, then, that a basic structure satisfying the principle of efficiency and in which positions are open to those able and willing to strive for them will lead to a just distribution. Assigning rights and duties in this way is thought to give a scheme which allocates wealth and income, authority and responsibility, in a fair way whatever this allocation turns

out to be. The doctrine includes an important element of pure procedural justice which is carried over to the other interpretations.

At this point it is necessary to make a brief digression to explain the principle of efficiency. This principle is simply that of Pareto optimality (as economists refer to it) formulated so as to apply to the basic structure.[7] I shall always use the term "efficiency" instead because this is literally correct and the term "optimality" suggests that the concept is much broader than it is in fact.[8] To be sure, this principle was not originally intended to apply to institutions but to particular configurations of the economic system, for example, to distributions of goods among consumers or to modes of production. The principle holds that a configuration is efficient whenever it is impossible to change it so as to make some persons (at least one) better off without at the same time making other persons (at least one) worse off. Thus a distribution of a stock of commodities among certain individuals is efficient if there exists no redistribution of these goods that improves the circumstances of at least one of these individuals without another being disadvantaged. The organization of production is efficient if there is no way to alter inputs so as to produce more of some commodity without producing less of another. For if we could produce more of one good without having to give up some of another, the larger stock of goods could be used to better the circumstances of some persons without making that of others any worse. These applications of the principle show that it is, indeed, a principle of efficiency. A distribution of goods or a scheme of production is inefficient when there are ways of doing still better for some individuals without doing any worse for others. I shall assume that the parties in the original position accept this principle to judge the efficiency of economic and social arrangements. (See the accompanying discussion of the principle of efficiency.)

7. There are expositions of this principle in most any work on price theory or social choice. A perspicuous account is found in T. C. Koopmans, *Three Essays on the State of Economic Science* (New York, McGraw-Hill, 1957), pp. 41–66. See also A. K. Sen, *Collective Choice and Social Welfare* (San Francisco, Holden-Day Inc., 1970), pp. 21f. These works contain everything (and more) that is required for our purposes in this book; and the latter takes up the relevant philosophical questions. The principle of efficiency was introduced by Vilfredo Pareto in his *Manuel d'économie politique* (Paris, 1909), ch. VI, §53, and the appendix, §89. A translation of the relevant passages can be found in A. N. Page, *Utility Theory: A Book of Readings* (New York, John Wiley, 1968), pp. 38f. The related concept of indifference curves goes back to F. Y. Edgeworth, *Mathematical Psychics* (London, 1888), pp. 20–29; also in Page, pp. 160–167.

8. On this point see Koopmans, *Three Essays on the State of Economic Science,* p. 49. Koopmans remarks that a term like "allocative efficiency" would have been a more accurate name.

THE PRINCIPLE OF EFFICIENCY

Assume that there is a fixed stock of commodities to be distributed between two persons, x_1 and x_2. Let the line AB represent the points such that given x_1's gain at the corresponding level, there is no way to distribute the commodities so as to make x_2 better off than the point indicated by the curve. Consider the point D = (a,b). Then holding x_1, at the level a, the best that can be done for x_2 is the level b. In figure 3 the point O, the origin, represents the position before any commodities are distributed. The points on the line AB are the efficient points. Each point on AB can be seen to satisfy Pareto's criterion: there is no redistribution that makes either person better off without making the other worse off. This is conveyed by the fact that the line AB slopes downward to the right. Since there is but a fixed stock of items, it is supposed that as one person gains the other loses. (Of course, this assumption is dropped in the case of the basic structure which is a system of cooperation producing a sum of positive advantages.) Normally the region OAB is taken to be a convex set. This means that given any pair of points in the set, the points on the straight line joining these two points are also in the set. Circles, ellipses, squares, triangles, and so on are convex sets.

It is clear that there are many efficient points, in fact, all the points on the line AB. The principle of efficiency does not by itself select one particular distribution of commodities as the efficient one. To select among the efficient distributions some other principle, a principle of justice, say, is necessary.

Of two points, if one is northeast of the other, this point is superior by

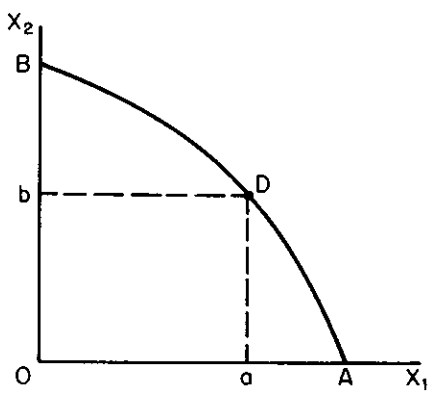


FIGURE 3

the principle of efficiency. Points to the northwest or southeast cannot be compared. The ordering defined by the principle of efficiency is but a partial one. Thus in figure 4 while C is superior to E, and D is superior to F, none of the points on the line AB are either superior or inferior to one another. The class of efficient points cannot be ranked. Even the extreme points A and B at which one of the parties has everything are efficient, just as other points on AB.

Observe that we cannot say that any point on the line AB is superior to all points in the interior of OAB. Each point on AB is superior only to those points in the interior southwest of it. Thus the point D is superior to all points inside the rectangle indicated by the dotted lines joining D to the points a and b. The point D is not superior to the point E. These points cannot be ordered. The point C, however, is superior to E and so are all the points on the line AB belonging to the small shaded triangular region that has the point E as a corner.

On the other hand, if one takes the 45° line as indicating the locus of equal distribution (this assumes an interpersonal cardinal interpretation of the axes, something not supposed in the preceding remarks), and if one counts this as an additional basis of decision, then all things considered, the point D may be preferable to both C and E. It is much closer to this line. One may even decide that an interior point such as F is to be preferred to C which is an efficient point. Actually, in justice as fairness the principles of justice are prior to considerations of efficiency and therefore, roughly speaking, the interior points that represent just distributions will generally be preferred to efficient points which represent unjust distributions. Of course, figure 4 depicts a very simple situation and cannot be applied to the basic structure.

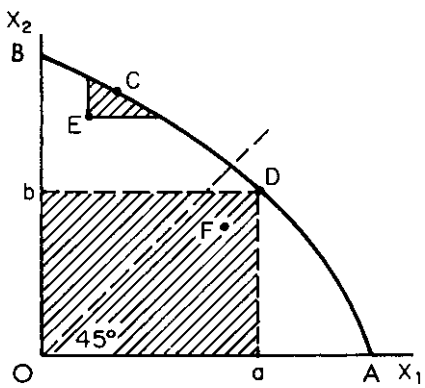


FIGURE 4

Now the principle of efficiency can be applied to the basic structure by reference to the expectations of representative men.[9] Thus we can say that an arrangement of rights and duties in the basic structure is efficient if and only if it is impossible to change the rules, to redefine the scheme of rights and duties, so as to raise the expectations of any representative man (at least one) without at the same time lowering the expectations of some (at least one) other representative man. Of course, these alterations must be consistent with the other principles. That is, in changing the basic structure we are not permitted to violate the principle of equal liberty or the requirement of open positions. What can be altered is the distribution of income and wealth and the way in which those in positions of authority and responsibility can regulate cooperative activities. Consistent with the constraints of liberty and accessibility, the allocation of these primary goods may be adjusted to modify the expectations of representative individuals. An arrangement of the basic structure is efficient when there is no way to change this distribution so as to raise the prospects of some without lowering the prospects of others.

There are, I shall assume, many efficient arrangements of the basic structure. Each of these specifies a division of advantages from social cooperation. The problem is to choose between them, to find a conception of justice that singles out one of these efficient distributions as also just. If we succeed in this, we shall have gone beyond mere efficiency yet in a way compatible with it. Now it is natural to try out the idea that as long as the social system is efficient there is no reason to be concerned with distribution. All efficient arrangements are in this case declared equally just. Of course, this suggestion would be outlandish for the allocation of particular goods to known individuals. No one would suppose that it is a matter of indifference from the standpoint of justice whether any one of a number of men happens to have everything. But the suggestion seems equally unreasonable for the basic structure. Thus it may be that under certain conditions serfdom cannot be significantly reformed without lowering the expectations of some other representative man, say that of landowners, in which case serfdom is efficient. Yet it may also happen under the same conditions that a system of free labor cannot be changed without

9. For the application of the Pareto criterion to systems of public rules, see J. M. Buchanan, "The Relevance of Pareto Optimality," *Journal of Conflict Resolution,* vol. 6 (1962), as well as his book with Gordon Tullock, *The Calculus of Consent* (Ann Arbor, The University of Michigan Press, 1962). In applying this and other principles to institutions I follow one of the points of "Two Concepts of Rules," *Philosophical Review,* vol. 64 (1955). Doing this has the advantage, among other things, of constraining the employment of principles by publicity effects. See §23, note 8.

lowering the expectations of some other representative man, say that of free laborers, so this arrangement is likewise efficient. More generally, whenever a society is relevantly divided into a number of classes, it is possible, let us suppose, to maximize with respect to any one of its representative men. These maxima give at least this many efficient positions, for none of them can be departed from to raise the expectations of others without lowering those of the representative man with respect to whom the maximum is defined. Thus each of these extremes is efficient but they surely cannot be all just.

Now these reflections show only what we knew all along, that is, that the principle of efficiency cannot serve alone as a conception of justice.[10] Therefore it must be supplemented in some way. Now in the system of natural liberty the principle of efficiency is constrained by certain background institutions; when these constraints are satisfied, any resulting efficient distribution is accepted as just. The system of natural liberty selects an efficient distribution roughly as follows. Let us suppose that we know from economic theory that under the standard assumptions defining a competitive market economy, income and wealth will be distributed in an efficient way, and that the particular efficient distribution which results in any period of time is determined by the initial distribution of assets, that is, by the initial distribution of income and wealth, and of natural talents and abilities. With each initial distribution, a definite efficient outcome is arrived at. Thus it turns out that if we are to accept the outcome as just, and not merely as efficient, we must accept the basis upon which over time the initial distribution of assets is determined.

In the system of natural liberty the initial distribution is regulated by the arrangements implicit in the conception of careers open to talents (as earlier defined). These arrangements presuppose a background of equal liberty (as specified by the first principle) and a free market economy. They require a formal equality of opportunity in that all have at least the same legal rights of access to all advantaged social positions. But since there is no effort to preserve an equality, or similarity, of social conditions, except insofar as this is necessary to preserve the requisite background institutions, the initial distribution of assets for any period of time is strongly influenced by natural and social contingencies. The existing

10. This fact is generally recognized in welfare economics, as when it is said that efficiency is to be balanced against equity. See for example Tibor Scitovsky, *Welfare and Competition* (London, George Allen and Unwin, 1952), pp. 60–69 and I. M. D. Little, *A Critique of Welfare Economics,* 2nd ed. (Oxford, The Clarendon Press, 1957), ch. VI, esp. pp. 112–116. See Sen's remarks on the limitations of the principle of efficiency, *Collective Choice and Social Welfare,* pp. 22, 24–26, 83–86.

distribution of income and wealth, say, is the cumulative effect of prior distributions of natural assets—that is, natural talents and abilities—as these have been developed or left unrealized, and their use favored or disfavored over time by social circumstances and such chance contingencies as accident and good fortune. Intuitively, the most obvious injustice of the system of natural liberty is that it permits distributive shares to be improperly influenced by these factors so arbitrary from a moral point of view.

The liberal interpretation, as I shall refer to it, tries to correct for this by adding to the requirement of careers open to talents the further condition of the principle of fair equality of opportunity. The thought here is that positions are to be not only open in a formal sense, but that all should have a fair chance to attain them. Offhand it is not clear what is meant, but we might say that those with similar abilities and skills should have similar life chances. More specifically, assuming that there is a distribution of natural assets, those who are at the same level of talent and ability, and have the same willingness to use them, should have the same prospects of success regardless of their initial place in the social system. In all sectors of society there should be roughly equal prospects of culture and achievement for everyone similarly motivated and endowed. The expectations of those with the same abilities and aspirations should not be affected by their social class.[11]

The liberal interpretation of the two principles seeks, then, to mitigate the influence of social contingencies and natural fortune on distributive shares. To accomplish this end it is necessary to impose further basic structural conditions on the social system. Free market arrangements must be set within a framework of political and legal institutions which regulates the overall trends of economic events and preserves the social conditions necessary for fair equality of opportunity. The elements of this framework are familiar enough, though it may be worthwhile to recall the importance of preventing excessive accumulations of property and wealth and of maintaining equal opportunities of education for all. Chances to acquire cultural knowledge and skills should not depend upon one's class position, and so the school system, whether public or private, should be designed to even out class barriers.

While the liberal conception seems clearly preferable to the system of

11. This definition follows Sidgwick's suggestion in *The Methods of Ethics,* p. 285n. See also R. H. Tawney, *Equality* (London, George Allen and Unwin, 1931), ch. II, sec. ii; and B. A. O. Williams, "The Idea of Equality," in *Philosophy, Politics, and Society,* ed. Peter Laslett and W. G. Runciman (Oxford, Basil Blackwell, 1962), pp. 125f.

natural liberty, intuitively it still appears defective. For one thing, even if it works to perfection in eliminating the influence of social contingencies, it still permits the distribution of wealth and income to be determined by the natural distribution of abilities and talents. Within the limits allowed by the background arrangements, distributive shares are decided by the outcome of the natural lottery; and this outcome is arbitrary from a moral perspective. There is no more reason to permit the distribution of income and wealth to be settled by the distribution of natural assets than by historical and social fortune. Furthermore, the principle of fair opportunity can be only imperfectly carried out, at least as long as some form of the family exists. The extent to which natural capacities develop and reach fruition is affected by all kinds of social conditions and class attitudes. Even the willingness to make an effort, to try, and so to be deserving in the ordinary sense is itself dependent upon happy family and social circumstances. It is impossible in practice to secure equal chances of achievement and culture for those similarly endowed, and therefore we may want to adopt a principle which recognizes this fact and also mitigates the arbitrary effects of the natural lottery itself. That the liberal conception fails to do this encourages one to look for another interpretation of the two principles of justice.

Before turning to the conception of democratic equality, we should note that of natural aristocracy. On this view no attempt is made to regulate social contingencies beyond what is required by formal equality of opportunity, but the advantages of persons with greater natural endowments are to be limited to those that further the good of the poorer sectors of society. The aristocratic ideal is applied to a system that is open, at least from a legal point of view, and the better situation of those favored by it is regarded as just only when less would be had by those below, if less were given to those above.[12] In this way the idea of *noblesse oblige* is carried over to the conception of natural aristocracy.

Now both the liberal conception and that of natural aristocracy are unstable. For once we are troubled by the influence of either social contingencies or natural chance on the determination of distributive shares, we are bound, on reflection, to be bothered by the influence of the other.

12. This formulation of the aristocratic ideal is derived from Santayana's account of aristocracy in ch. IV of *Reason and Society* (New York, Charles Scribner, 1905), pp. 109f. He says, for example, "an aristocratic regimen can only be justified by radiating benefit and by proving that were less given to those above, less would be attained by those beneath them." I am indebted to Robert Rodes for pointing out to me that natural aristocracy is a possible interpretation of the two principles of justice and that an ideal feudal system might also try to fulfill the difference principle.

From a moral standpoint the two seem equally arbitrary. So however we move away from the system of natural liberty, we cannot be satisfied short of the democratic conception. This conception I have yet to explain. And, moreover, none of the preceding remarks are an argument for this conception, since in a contract theory all arguments, strictly speaking, are to be made in terms of what it would be rational to agree to in the original position. But I am concerned here to prepare the way for the favored interpretation of the two principles so that these criteria, especially the second one, will not strike the reader as extreme. Once we try to find a rendering of them which treats everyone equally as a moral person, and which does not weight men's share in the benefits and burdens of social cooperation according to their social fortune or their luck in the natural lottery, the democratic interpretation is the best choice among the four alternatives. With these comments as a preface, I now turn to this conception.

13. DEMOCRATIC EQUALITY AND THE DIFFERENCE PRINCIPLE

The democratic interpretation, as the table suggests, is arrived at by combining the principle of fair equality of opportunity with the difference principle. This principle removes the indeterminateness of the principle of efficiency by singling out a particular position from which the social and economic inequalities of the basic structure are to be judged. Assuming the framework of institutions required by equal liberty and fair equality of opportunity, the higher expectations of those better situated are just if and only if they work as part of a scheme which improves the expectations of the least advantaged members of society. The intuitive idea is that the social order is not to establish and secure the more attractive prospects of those better off unless doing so is to the advantage of those less fortunate. (See the discussion of the difference principle that follows.)

THE DIFFERENCE PRINCIPLE

Assume that indifference curves now represent distributions that are judged equally just. Then the difference principle is a strongly egalitarian conception in the sense that unless there is a distribution that makes both

persons better off (limiting ourselves to the two-person case for simplicity), an equal distribution is to be preferred. The indifference curves take the form depicted in figure 5. These curves are actually made up of vertical and horizontal lines that intersect at right angles at the 45° line (again supposing an interpersonal and cardinal interpretation of the axes). No matter how much either person's situation is improved, there is no gain from the standpoint of the difference principle unless the other gains also.

Suppose that x_1 is the most favored representative man in the basic structure. As his expectations are increased so are the prospects of x_2, the least advantaged man. In figure 6 let the curve OP represent the contribution to x_2's expectations made by the greater expectations of x_1. The point O, the origin, represents the hypothetical state in which all social primary goods are distributed equally. Now the OP curve is always below the 45° line, since x_1 is always better off. Thus the only relevant parts of the indifference curves are those below this line, and for this reason the upper left-hand part of figure 6 is not drawn in. Clearly the difference principle is perfectly satisfied only when the OP curve is just tangent to the highest indifference curve that it touches. In figure 6 this is at the point a.

Note that the contribution curve, the curve OP, rises upward to the right because it is assumed that the social cooperation defined by the basic structure is mutually advantageous. It is no longer a matter of shuffling about a fixed stock of goods. Also, nothing is lost if an accurate interpersonal comparison of benefits is impossible. It suffices that the least favored person can be identified and his rational preference determined.

A view less egalitarian than the difference principle, and perhaps more

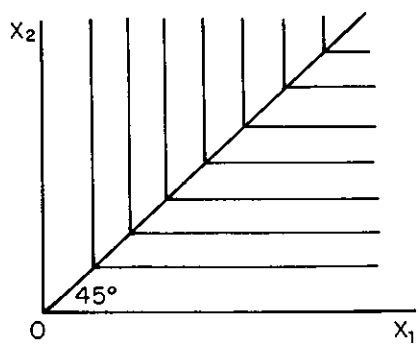


FIGURE 5

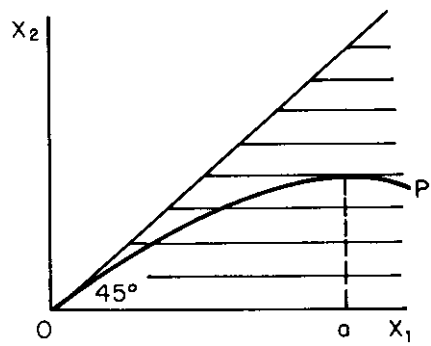


FIGURE 6

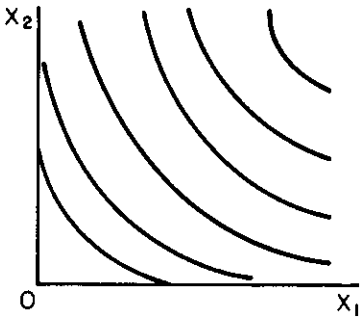


FIGURE 7

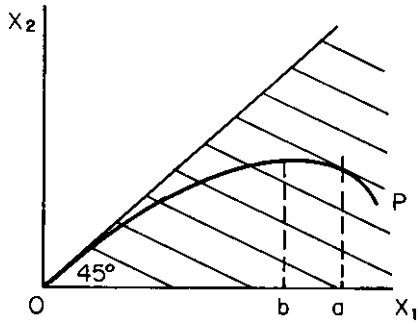


FIGURE 8

plausible at first sight, is one in which the indifference lines for just distributions (or for all things considered) are smooth curves convex to the origin, as in figure 7. The indifference curves for social welfare functions are often depicted in this fashion. This shape of the curves expresses the fact that as either person gains relative to the other, further benefits to him become less valuable from a social point of view.

A classical utilitarian, on the other hand, is indifferent as to how a constant sum of benefits is distributed. He appeals to equality only to break ties. If there are but two persons, then assuming an interpersonal cardinal interpretation of the axes, the utilitarian's indifference lines for distributions are straight lines perpendicular to the 45° line. Since, however, x_1 and x_2 are representative men, the gains to them have to be weighted by the number of persons they each represent. Since presumably x_2 represents rather more persons than x_1, the indifference lines become more horizontal, as seen in figure 8. The ratio of the number of advantaged to the number of disadvantaged defines the slope of these straight lines. Drawing the same contribution curve OP as before, we see that the best distribution from a utilitarian point of view is reached at the point which is beyond the point b where the OP curve reaches its maximum. Since the difference principle selects the point b and b is always to the left of a, utilitarianism allows, other things equal, larger inequalities.

To illustrate the difference principle, consider the distribution of income among social classes. Let us suppose that the various income groups correlate with representative individuals by reference to whose expectations we can judge the distribution. Now those starting out as members of the entrepreneurial class in property-owning democracy, say, have a better prospect than those who begin in the class of unskilled laborers. It

seems likely that this will be true even when the social injustices which now exist are removed. What, then, can possibly justify this kind of initial inequality in life prospects? According to the difference principle, it is justifiable only if the difference in expectation is to the advantage of the representative man who is worse off, in this case the representative unskilled worker. The inequality in expectation is permissible only if lowering it would make the working class even more worse off. Supposedly, given the rider in the second principle concerning open positions, and the principle of liberty generally, the greater expectations allowed to entrepreneurs encourages them to do things which raise the prospects of laboring class. Their better prospects act as incentives so that the economic process is more efficient, innovation proceeds at a faster pace, and so on. I shall not consider how far these things are true. The point is that something of this kind must be argued if these inequalities are to satisfy by the difference principle.

I shall now make a few remarks about this principle. First of all, in applying it, one should distinguish between two cases. The first case is that in which the expectations of the least advantaged are indeed maximized (subject, of course, to the mentioned constraints). No changes in the expectations of those better off can improve the situation of those worst off. The best arrangement obtains, what I shall call a perfectly just scheme. The second case is that in which the expectations of all those better off at least contribute to the welfare of the more unfortunate. That is, if their expectations were decreased, the prospects of the least advantaged would likewise fall. Yet the maximum is not yet achieved. Even higher expectations for the more advantaged would raise the expectations of those in the lowest position. Such a scheme is, I shall say, just throughout, but not the best just arrangement. A scheme is unjust when the higher expectations, one or more of them, are excessive. If these expectations were decreased, the situation of the least favored would be improved. How unjust an arrangement is depends on how excessive the higher expectations are and to what extent they depend upon the violation of the other principles of justice, for example, fair equality of opportunity; but I shall not attempt to measure the degrees of injustice. The point to note here is that while the difference principle is, strictly speaking, a maximizing principle, there is a significant distinction between the cases that fall short of the best arrangement. A society should try to avoid situations where the marginal contributions of those better off are negative, since, other things equal, this seems a greater fault than falling short of the best scheme when these contributions are positive. The even larger difference

between classes violates the principle of mutual advantage as well as democratic equality (§17).

A further point is this. We saw that the system of natural liberty and the liberal conception go beyond the principle of efficiency by setting up certain background institutions and leaving the rest to pure procedural justice. The democratic conception holds that while pure procedural justice may be invoked to some extent at least, the way previous interpretations do this still leaves too much to social and natural contingency. But it should be noted that the difference principle is compatible with the principle of efficiency. For when the former is fully satisfied, it is indeed impossible to make any one representative man better off without making another worse off, namely, the least advantaged representative man whose expectations we are to maximize. Thus justice is defined so that it is consistent with efficiency, at least when the two principles are perfectly fulfilled. Of course, if the basic structure is unjust, these principles will authorize changes that may lower the expectations of some of those better off; and therefore the democratic conception is not consistent with the principle of efficiency if this principle is taken to mean that only changes which improve everyone's prospects are allowed. Justice is prior to efficiency and requires some changes that are not efficient in this sense. Consistency obtains only in the sense that a perfectly just scheme is also efficient.

Next, we may consider a certain complication regarding the meaning of the difference principle. It has been taken for granted that if the principle is satisfied, everyone is benefited. One obvious sense in which this is so is that each man's position is improved with respect to the initial arrangement of equality. But it is clear that nothing depends upon being able to identify this initial arrangement; indeed, how well off men are in this situation plays no essential role in applying the difference principle. We simply maximize the expectations of the least favored position subject to the required constraints. As long as doing this is an improvement for everyone, as so far I have assumed it is, the estimated gains from the situation of hypothetical equality are irrelevant, if not largely impossible to ascertain anyway. There may be, however, a further sense in which everyone is advantaged when the difference principle is satisfied, at least if we make certain assumptions. Let us suppose that inequalities in expectations are chain-connected: that is, if an advantage has the effect of raising the expectations of the lowest position, it raises the expectations of all positions in between. For example, if the greater expectations for entrepreneurs benefit the unskilled worker, they also benefit the semi-

skilled. Notice that chain connection says nothing about the case where the least advantaged do not gain, so that it does not mean that all effects move together. Assume further that expectations are close-knit: that is, it is impossible to raise or lower the expectation of any representative man without raising or lowering the expectation of every other representative man, especially that of the least advantaged. There is no loose-jointedness, so to speak, in the way expectations hang together. Now with these assumptions there is a sense in which everyone benefits when the difference principle is satisfied. For the representative man who is better off in any two-way comparison gains by the advantages offered him, and the man who is worse off gains from the contributions which these inequalities make. Of course, these conditions may not hold. But in this case those who are better off should not have a veto over the benefits available for the least favored. We are still to maximize the expectations of those most disadvantaged. (See the accompanying discussion of chain connection.)

CHAIN CONNECTION

For simplicity assume that there are three representative men. Let x_1 be the most favored and x_3 the least favored with x_2 in between. Let the expectations of x_1 be marked off along the horizontal axis, the expectations of x_2 and x_3 along the vertical axis. The curves showing the contribution of the most favored to the other groups begin at the origin as the hypothetical position of equality. Moreover, there is a maximum gain permitted to the most favored on the assumption that, even if the difference principle would allow it, there would be unjust effects on the political system and the like excluded by the priority of liberty.

The difference principle selects the point where the curve for x_3 reaches its maximum, for example, the point a in figure 9.

Chain connection means that at any point where the x_3 curve is rising to the right, the x_2 curve is also rising, as in the intervals left of the points a and b in figures 9 and 10. Chain connection says nothing about the case where the x_3 curve is falling to the right, as in the interval to the right of the point a in figure 9. The x_2 curve may be either rising or falling (as indicated by the dashed line x_2'). Chain connection does not hold to the right of b in figure 10.

Intervals in which both the x_2 and the x_3 curves are rising define the intervals of positive contributions. Any more to the right increases the

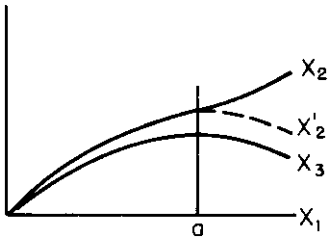


FIGURE 9

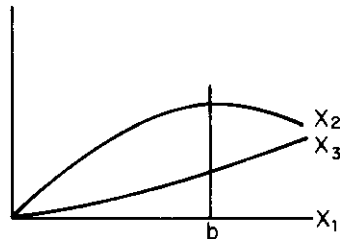


FIGURE 10

average expectation (average utility if utility is measured by expectations) and also satisfies the principle of efficiency as a criterion of change, that is, points to the right improve everyone's situation.

In figure 9 the average expectations may be rising beyond the point a, although the expectations of the least favored are falling. (This depends on the weights of the several groups.) The difference principle excludes this and selects the point a.

Close-knitness means that there are no flat stretches on the curves for x_2 and x_3. At each point both curves are either rising or falling. All the curves illustrated are close-knit.

I shall not examine how likely it is that chain connection and close-knitness hold. The difference principle is not contingent on these relations being satisfied. However, when the contributions of the more favored positions spread generally throughout society and are not confined to particular sectors, it seems plausible that if the least advantaged benefit so do others in between. Moreover, a wide diffusion of benefits is favored by two features of institutions both exemplified by the basic structure: first, they are set up to advance certain fundamental interests which everyone has in common, and second, offices and positions are open. Thus it seems probable that if the authority and powers of legislators and judges, say, improve the situation of the less favored, they improve that of citizens generally. Chain connection may often be true, provided the other principles of justice are fulfilled. If this is so, then we may observe that within the region of positive contributions (the region where the advantages of all those in favored positions raise the prospects of the least fortunate), any movement toward the perfectly just arrangement improves everyone's expectation. Under these circumstances the difference principle has somewhat similar practical consequences for the principles of efficiency and average utility (if utility is measured by primary goods). Of course, if

chain connection rarely holds, this similarity is unimportant. But it seems likely that within a just social scheme a general diffusion of benefits often takes place.

There is a further complication. Close-knitness is assumed in order to simplify the statement of the difference principle. It is clearly conceivable, however likely or important in practice, that the least advantaged are not affected one way or the other by some changes in expectations of the best off although these changes benefit others. In this sort of case close-knitness fails, and to cover the situation we can express a more general principle as follows: in a basic structure with n relevant representatives, first maximize the welfare of the worst off representative man; second, for equal welfare of the worst-off representative, maximize the welfare of the second worst-off representative man, and so on until the last case which is, for equal welfare of all the preceding n–1 representatives, maximize the welfare of the best-off representative man. We may think of this as the lexical difference principle.[13] I think, however, that in actual cases this principle is unlikely to be relevant, for when the greater potential benefits to the more advantaged are significant, there will surely be some way to improve the situation of the less advantaged as well. The general laws governing the institutions of the basic structure insure that cases requiring the lexical principle will not arise. Thus I shall always use the difference principle in the simpler form, and so the outcome of the last several sections is that the second principle reads as follows:

> Social and economic inequalities are to be arranged so that they are both (a) to the greatest expected benefit of the least advantaged and (b) attached to offices and positions open to all under conditions of fair equality of opportunity.

Finally, a comment about terminology. Economics may wish to refer to the difference principle as the maximin criterion, but I have carefully avoided this name for several reasons. The maximin criterion is generally understood as a rule for choice under great uncertainty (§26), whereas the difference principle is a principle of justice. It is undesirable to use the same name for two things that are so distinct. The difference principle is a very special criterion: it applies primarily to the basic structure of society via representative individuals whose expectations are to be estimated by an index of primary goods (§15). In addition, calling the difference principle the maximin criterion might wrongly suggest that the main argument for this principle from the original position derives from an

13. On this point, see Sen, *Collective Choice and Social Welfare,* p. 138n.

assumption of very high risk aversion. There is indeed a relation between the difference principle and such an assumption, but extreme attitudes to risk are not postulated (§28); and in any case, there are many considerations in favor of the difference principle in which the aversion to risk plays no role at all. Thus it is best to use the term "maximin criterion" solely for the rule of choice under uncertainty.

14. FAIR EQUALITY OF OPPORTUNITY AND PURE PROCEDURAL JUSTICE

I should now like to comment upon the second part of the second principle, henceforth to be understood as the liberal principle of fair equality of opportunity. It must not then be confused with the notion of careers open to talents; nor must one forget that since it is tied in with the difference principle its consequences are quite distinct from the liberal interpretation of the two principles taken together. In particular, I shall try to show further on (§17) that this principle is not subject to the objection that it leads to a meritocratic society. Here I wish to consider a few other points, especially its relation to the idea of pure procedural justice.

First, though, I should note that the reasons for requiring open positions are not solely, or even primarily, those of efficiency. I have not maintained that offices must be open if in fact everyone is to benefit from an arrangement. For it may be possible to improve everyone's situation by assigning certain powers and benefits to positions despite the fact that certain groups are excluded from them. Although access is restricted, perhaps these offices can still attract superior talent and encourage better performance. But the principle of open positions forbids this. It expresses the conviction that if some places were not open on a basis fair to all, those kept out would be right in feeling unjustly treated even though they benefited from the greater efforts of those who were allowed to hold them. They would be justified in their complaint not only because they were excluded from certain external rewards of office but because they were debarred from experiencing the realization of self which comes from a skillful and devoted exercise of social duties. They would be deprived of one of the main forms of human good.

Now I have said that the basic structure is the primary subject of justice. Of course, any ethical theory recognizes the importance of the basic structure as a subject of justice, but not all theories regard its importance in the same way. In justice as fairness society is interpreted as

a cooperative venture for mutual advantage. The basic structure is a public system of rules defining a scheme of activities that leads men to act together so as to produce a greater sum of benefits and assigns to each certain recognized claims to a share in the proceeds. What a person does depends upon what the public rules say he will be entitled to, and what a person is entitled to depends on what he does. The distribution which results is arrived at by honoring the claims determined by what persons undertake to do in the light of these legitimate expectations.

These considerations suggest the idea of treating the question of distributive shares as a matter of pure procedural justice.[14] The intuitive idea is to design the social system so that the outcome is just whatever it happens to be, at least so long as it is within a certain range. The notion of pure procedural justice is best understood by a comparison with perfect and imperfect procedural justice. To illustrate the former, consider the simplest case of fair division. A number of men are to divide a cake: assuming that the fair division is an equal one, which procedure, if any, will give this outcome? Technicalities aside, the obvious solution is to have one man divide the cake and get the last piece, the others being allowed their pick before him. He will divide the cake equally, since in this way he assures for himself the largest share possible. This example illustrates the two characteristic features of perfect procedural justice. First, there is an independent criterion for what is a fair division, a criterion defined separately from and prior to the procedure which is to be followed. And second, it is possible to devise a procedure that is sure to give the desired outcome. Of course, certain assumptions are made here, such as that the man selected can divide the cake equally, wants as large a piece as he can get, and so on. But we can ignore these details. The essential thing is that there is an independent standard for deciding which outcome is just and a procedure guaranteed to lead to it. Pretty clearly, perfect procedural justice is rare, if not impossible, in cases of much practical interest.

Imperfect procedural justice is exemplified by a criminal trial. The desired outcome is that the defendant should be declared guilty if and only if he has committed the offense with which he is charged. The trial procedure is framed to search for and to establish the truth in this regard.

14. For a general discussion of procedural justice, see Brian Barry, *Political Argument* (London, Routledge and Kegan Paul, 1965), ch. VI. On the problem of fair division, see R. D. Luce and Howard Raiffa, *Games and Decisions* (New York, John Wiley and Sons, Inc., 1957), pp. 363–368; and Hugo Steinhaus, "The Problem of Fair Division," *Econometrica,* vol. 16 (1948).

But it seems impossible to design the legal rules so that they always lead to the correct result. The theory of trials examines which procedures and rules of evidence, and the like, are best calculated to advance this purpose consistent with the other ends of the law. Different arrangements for hearing cases may reasonably be expected in different circumstances to yield the right results, not always but at least most of the time. A trial, then, is an instance of imperfect procedural justice. Even though the law is carefully followed, and the proceedings fairly and properly conducted, it may reach the wrong outcome. An innocent man may be found guilty, a guilty man may be set free. In such cases we speak of a miscarriage of justice: the injustice springs from no human fault but from a fortuitous combination of circumstances which defeats the purpose of the legal rules. The characteristic mark of imperfect procedural justice is that while there is an independent criterion for the correct outcome, there is no feasible procedure which is sure to lead to it.

By contrast, pure procedural justice obtains when there is no independent criterion for the right result: instead there is a correct or fair procedure such that the outcome is likewise correct or fair, whatever it is, provided that the procedure has been properly followed. This situation is illustrated by gambling. If a number of persons engage in a series of fair bets, the distribution of cash after the last bet is fair, or at least not unfair, whatever this distribution is. I assume here that fair bets are those having a zero expectation of gain, that the bets are made voluntarily, that no one cheats, and so on. The betting procedure is fair and freely entered into under conditions that are fair. Thus the background circumstances define a fair procedure. Now any distribution of cash summing to the initial stock held by all individuals could result from a series of fair bets. In this sense all of these particular distributions are equally fair. A distinctive feature of pure procedural justice is that the procedure for determining the just result must actually be carried out; for in these cases there is no independent criterion by reference to which a definite outcome can be known to be just. Clearly we cannot say that a particular state of affairs is just because it could have been reached by following a fair procedure. This would permit far too much. It would allow one to say that almost any distribution of goods is just, or fair, since it could have come about as a result of fair gambles. What makes the final outcome of betting fair, or not unfair, is that it is the one which has arisen after a series of fair gambles. A fair procedure translates its fairness to the outcome only when it is actually carried out.

In order, therefore, to apply the notion of pure procedural justice to distributive shares it is necessary to set up and to administer impartially a just system of institutions. Only against the background of a just basic structure, including a just political constitution and a just arrangement of economic and social institutions, can one say that the requisite just procedure exists. In Part Two I shall describe a basic structure that has the necessary features (§43). Its various institutions are explained and connected with the two principles of justice.

The role of the principle of fair opportunity is to insure that the system of cooperation is one of pure procedural justice. Unless it is satisfied, distributive justice could not be left to take care of itself, even within a restricted range. Now the practical advantage of pure procedural justice is that it is no longer necessary to keep track of the endless variety of circumstances and the changing relative positions of particular persons. One avoids the problem of defining principles to cope with the enormous complexities which would arise if such details were relevant. It is a mistake to focus attention on the varying relative positions of individuals and to require that every change, considered as a single transaction viewed in isolation, be in itself just. It is the arrangement of the basic structure which is to be judged, and judged from a general point of view. Unless we are prepared to criticize it from the standpoint of a relevant representative man in some particular position, we have no complaint against it. Thus the acceptance of the two principles constitutes an understanding to discard as irrelevant as a matter of social justice much of the information and many of the complications of everyday life.

In pure procedural justice, then, distributions of advantages are not appraised in the first instance by confronting a stock of benefits available with given desires and needs of known individuals. The allotment of the items produced takes place in accordance with the public system of rules, and this system determines what is produced, how much is produced, and by what means. It also determines legitimate claims the honoring of which yields the resulting distribution. Thus in this kind of procedural justice the correctness of the distribution is founded on the justice of the scheme of cooperation from which it arises and on answering the claims of individuals engaged in it. A distribution cannot be judged in isolation from the system of which it is the outcome or from what individuals have done in good faith in the light of established expectations. If it is asked in the abstract whether one distribution of a given stock of things to definite individuals with known desires and preferences is better than another, then there is simply no answer to this question. The conception of the two

principles does not interpret the primary problem of distributive justice as one of allocative justice.

By contrast allocative justice applies when a given collection of goods is to be divided among definite individuals with known desires and needs. The collection to be allotted is not the product of these individuals, nor do they stand in any existing cooperative relations. Since there are no prior claims on the things to be distributed, it is natural to share them out according to desires and needs, or even to maximize the net balance of satisfaction. Justice becomes a kind of efficiency, unless equality is preferred. Suitably generalized, the allocative conception leads to the classical utilitarian view. For as we have seen, this doctrine assimilates justice to the benevolence of the impartial spectator and the latter in turn to the most efficient design of institutions to promote the greatest balance of satisfaction. The point to note here is that utilitarianism does not interpret the basic structure as a scheme of pure procedural justice. For the utilitarian has, in principle anyway, an independent standard for judging all distributions, namely, whether they produce the greatest net balance of satisfaction. In his theory, institutions are more or less imperfect arrangements for bringing about this end. Thus given existing desires and preferences, and the developments into the future which they allow, the statesman's aim is to set up those social schemes that will best approximate an already specified goal. Since these arrangements are subject to the unavoidable constraints and hindrances of everyday life, the basic structure is a case of imperfect procedural justice.

For the time being I shall suppose that the two parts of the second principle are lexically ordered. Thus we have one lexical ordering within another. The advantage of the special conception is that it has a definite shape and suggests certain questions for investigation, for example, under what assumptions if any would the lexical ordering be chosen? Our inquiry is given a particular direction and is no longer confined to generalities. Of course, this conception of distributive shares is obviously a great simplification. It is designed to characterize in a clear way a basic structure that makes use of the idea of pure procedural justice. But all the same we should attempt to find simple concepts that can be assembled to give a reasonable conception of justice. The notions of the basic structure, of the veil of ignorance, of a lexical order, of the least favored position, as well as of pure procedural justice are all examples of this. By themselves none of these could be expected to work, but properly put together they may serve well enough. It is too much to suppose that there exists for all or even most moral problems a reasonable solution. Perhaps only a few can

be satisfactorily answered. In any case social wisdom consists in framing institutions so that intractable difficulties do not often arise and in accepting the need for clear and simple principles.

15. PRIMARY SOCIAL GOODS AS THE BASIS OF EXPECTATIONS

So much, then, for a brief statement and explanation of the two principles of justice and of the procedural conception which they express. In later chapters I shall present further details by describing an arrangement of institutions that realizes this conception. At the moment, however, there are several preliminary matters that must be faced. I begin with a discussion of expectations and how they are to be estimated.

The significance of this question can be brought out by a comparison with utilitarianism. When applied to the basic structure this view requires us to maximize the algebraic sum of expected utilities taken over all relevant positions. (The classical principle weights these expectations by the number of persons in these positions, the average principle by the fraction of persons.) Leaving aside for the next section the question as to what defines a relevant position, it is clear that utilitarianism assumes some fairly accurate measure of utility. Not only is it necessary to have a cardinal measure for each representative individual but some method of correlating the scales of different persons is presupposed if we are to say that the gains of some are to outweigh the losses of others. It is unreasonable to demand great precision, yet these estimates cannot be left to our unguided intuition. Moreover, they may be based on ethical and other notions, not to mention bias and self-interest, which puts their validity in question. Simply because we do in fact make what we call interpersonal comparisons of well-being does not mean that we understand the basis of these comparisons or that we should accept them as sound. To settle these matters we need to give an account of these judgments, to set out the criteria that underlie them (§49). For questions of social justice we should try to find some objective grounds for these comparisons, ones that men can recognize and agree to. I believe that the real objection to utilitarianism lies elsewhere. Even if interpersonal comparisons can be made, these comparisons must reflect values which it makes sense to pursue. The controversy about interpersonal comparisons tends to obscure the real question, namely, whether the total (or average) happiness is to be maximized in the first place.

The difference principle tries to establish objective grounds for interpersonal comparisons in two ways. First of all, as long as we can identify the least advantaged representative man, only ordinal judgments of well-being are required from then on. We know from what position the social system is to be judged. It does not matter how much worse off this representative individual is than the others. The further difficulties of cardinal measurement do not arise since no other interpersonal comparisons are necessary. The difference principle, then, asks less of our judgments of welfare. We never have to calculate a sum of advantages involving a cardinal measure. While qualitative interpersonal comparisons are made in finding the bottom position, for the rest the ordinal judgments of one representative man suffice.

Second, the difference principle introduces a simplification for the basis of interpersonal comparisons. These comparisons are made in terms of expectations of primary social goods. In fact, I define these expectations simply as the index of these goods which a representative individual can look forward to. One man's expectations are greater than another's if this index for some one in his position is greater. Now primary goods, as I have already remarked, are things which it is supposed a rational man wants whatever else he wants. Regardless of what an individual's rational plans are in detail, it is assumed that there are various things which he would prefer more of rather than less. With more of these goods men can generally be assured of greater success in carrying out their intentions and in advancing their ends, whatever these ends may be. The primary social goods, to give them in broad categories, are rights, liberties, and opportunities, and income and wealth. (A very important primary good is a sense of one's own worth; but for simplicity I leave this aside until much later, §67.) It seems evident that in general these things fit the description of primary goods. They are social goods in view of their connection with the basic structure; liberties and opportunities are defined by the rules of major institutions and the distribution of income and wealth is regulated by them.

The theory of the good adopted to account for primary goods will be presented more fully in Chapter VII. It is a familiar one going back to Aristotle, and something like it is accepted by philosophers so different in other respects as Kant and Sidgwick. It is not in dispute between the contract doctrine and utilitarianism. The main idea is that a person's good is determined by what is for him the most rational long-term plan of life given reasonably favorable circumstances. A man is happy when he is more or less successfully in the way of carrying out this plan. To put it

briefly, the good is the satisfaction of rational desire. We are to suppose, then, that each individual has a rational plan of life drawn up subject to the conditions that confront him. This plan is designed to permit the harmonious satisfaction of his interests. It schedules activities so that various desires can be fulfilled without interference. It is arrived at by rejecting other plans that are either less likely to succeed or do not provide for such an inclusive attainment of aims. Given the alternatives available, a rational plan is one which cannot be improved upon; there is no other plan which, taking everything into account, would be preferable.

Let us consider several difficulties. One problem clearly is the construction of the index of primary social goods. Assuming that the two principles of justice are serially ordered, this problem is greatly simplified. The basic liberties are always equal, and there is fair equality of opportunity; one does not need to balance these liberties and rights against other values. The primary social goods that vary in their distribution are the rights and prerogatives of authority, and income and wealth. But the difficulties are not so great as they might seem at first because of the nature of the difference principle. The only index problem that concerns us is that for the least advantaged group. The primary goods enjoyed by other representative individuals are adjusted to raise this index, subject of course to the usual constraints. It is unnecessary to define weights for the more favored positions in any detail, as long as we are sure that they are more favored. But often this is easy since they frequently have more of each primary good that is distributed unequally. If we know how the distribution of goods to the more favored affects the expectations of the most disfavored, this is sufficient. The index problem largely reduces, then, to that of weighting primary goods for the least advantaged. We try to do this by taking up the standpoint of the representative individual from this group and asking which combination of primary social goods it would be rational for him to prefer. In doing this we admittedly rely upon intuitive estimates. But this cannot be avoided entirely.

Another difficulty is this. It may be objected that expectations should not be defined as an index of primary goods anyway but rather as the satisfactions to be expected when plans are executed using these goods. After all, it is in the fulfillment of these plans that men gain happiness, and therefore the estimate of expectations should not be founded on the available means. Justice as fairness, however, takes a different view. For it does not look behind the use which persons make of the rights and opportunities available to them in order to measure, much less to maximize, the satisfactions they achieve. Nor does it try to evaluate the relative

merits of different conceptions of the good. Instead, it is assumed that the members of society are rational persons able to adjust their conceptions of the good to their situation. There is no necessity to compare the worth of the conceptions of different persons once it is supposed they are compatible with the principles of justice. Everyone is assured an equal liberty to pursue whatever plan of life he pleases as long as it does not violate what justice demands. Men share in primary goods on the principle that some can have more if they are acquired in ways which improve the situation of those who have less. Once the whole arrangement is set up and going no questions are asked about the totals of satisfaction or perfection.

It is worth noting that this interpretation of expectations represents, in effect, an agreement to compare men's situations solely by reference to things which it is assumed they all normally need to carry out their plans. This seems the most feasible way to establish a publicly recognized objective and common measure that reasonable persons can accept. Whereas there cannot be a similar agreement on how to estimate happiness as defined, say, by men's success in executing their rational plans, much less on the intrinsic value of these plans. Now founding expectations on primary goods is another simplifying device. I should like to comment in passing that this and other simplifications are accompanied by some sort of philosophical explanation, though this is not strictly necessary. Theoretical assumptions must, of course, do more than simplify; they must identify essential elements that explain the facts we want to understand. Similarly, the parts of a theory of justice must represent basic moral features of the social structure, and if it appears that some of these are being left aside, it is desirable to assure ourselves that such is not the case. I shall try to follow this rule. But even so, the soundness of the theory of justice is shown as much in its consequences as in the prima facie acceptability of its premises. Indeed, these cannot be usefully separated and therefore the discussion of institutional questions, particularly in Part Two, which may seem at first unphilosophical, is in fact unavoidable.

16. RELEVANT SOCIAL POSITIONS

In applying the two principles of justice to the basic structure of society one takes the position of certain representative individuals and considers how the social system looks to them. The perspective of those in these

situations defines a suitably general point of view. But certainly not all social positions are relevant. For not only are there farmers, say, but dairy farmers, wheat farmers, farmers working on large tracts of land, and so on for other occupations and groups indefinitely. We cannot have a coherent and manageable theory if we must take such a multiplicity of positions into account. The assessment of so many competing claims is impossible. Therefore we need to identify certain positions as more basic than the others and as providing an appropriate standpoint for judging the social system. Thus the choice of these positions becomes part of the theory of justice. On what principle, though, are they to be identified?

To answer this question we must keep in mind the fundamental problem of justice and the manner in which the two principles cope with it. The primary subject of justice, as I have emphasized, is the basic structure of society. The reason for this is that its effects are so profound and pervasive, and present from birth. This structure favors some starting places over others in the division of the benefits of social cooperation. It is these inequalities which the two principles are to regulate. Once these principles are satisfied, other inequalities are allowed to arise from men's voluntary actions in accordance with the principle of free association. Thus the relevant social positions are, so to speak, the starting places properly generalized and aggregated. By choosing these positions to specify the general point of view one follows the idea that the two principles attempt to mitigate the arbitrariness of natural contingency and social fortune.

I suppose, then, that for the most part each person holds two relevant positions: that of equal citizenship and that defined by his place in the distribution of income and wealth. The relevant representative men, therefore, are the representative citizen and the representatives of those with different expectations for the unequally distributed primary goods. Since I assume that in general other positions are entered into voluntarily, we need not consider the point of view of men in these positions in judging the basic structure. Instead, we are to adjust the whole scheme to suit the preferences of those in the so-called starting places.

Now as far as possible the basic structure should be appraised from the position of equal citizenship. This position is defined by the rights and liberties required by the principle of equal liberty and the principle of fair equality of opportunity. When the two principles are satisfied, all are equal citizens, and so everyone holds this position. In this sense, equal citizenship defines a general point of view. The problems of adjudicating among the basic liberties are settled by reference to it. These matters I

shall discuss in Chapter IV. But it should be noted here that many questions of social policy can also be considered from this position. For there are matters which concern the interests of everyone and in regard to which distributive effects are immaterial or irrelevant. In these cases the principle of the common interest can be applied. According to this principle institutions are ranked by how effectively they guarantee the conditions necessary for all equally to further their aims, or by how efficiently they advance shared ends that will similarly benefit everyone. Thus reasonable regulations to maintain public order and security, or efficient measures for public health and safety, promote the common interest in this sense. So do collective efforts for national defense in a just war. It may be suggested that maintaining public health and safety or achieving victory in a just war have distributive effects: those with higher expectations benefit more since they have more to lose. But if social and economic inequalities are just, these effects may be left aside and the principle of the common interest applied. The standpoint of equal citizenship is the appropriate one.

The definition of representative men for judging social and economic inequalities is less satisfactory. For one thing, taking these individuals as specified by the levels of income and wealth, I assume that these primary social goods are sufficiently correlated with differences in authority and responsibility. That is, I suppose that those with greater political authority, say, or those with more responsibility in various associations, are in general better off in other respects. On the whole, this assumption seems safe enough for our purposes. There is also a question about how many such representative men to single out, but this is not crucial because the difference principle selects one representative for a special role. The serious difficulty is how to define the least fortunate group.

To fix ideas, let us single out the least advantaged as those who are least favored by each of the three main kinds of contingencies. Thus this group includes persons whose family and class origins are more disadvantaged than others, whose natural endowments (as realized) permit them to fare less well, and whose fortune and luck in the course of life turn out to be less happy, all within the normal range (as noted below) and with the relevant measures based on social primary goods. Various refinements will certainly be necessary in practice, but this rough definition of the least advantaged suitably expresses the connection with the problem of contingency and should suffice for our purposes here. I shall assume that everyone has physical needs and psychological capacities within the normal range, so that the questions of health care and

mental capacity do not arise. Besides prematurely introducing matters that may take us beyond the theory of justice, the consideration of these hard cases can distract our moral perception by leading us to think of persons distant from us whose fate arouses pity and anxiety. The first problem of justice concerns the relations among those who in the everyday course of things are full and active participants in society and directly or indirectly associated together over the whole span of their life. Thus the difference principle is to apply to citizens engaged in social cooperation; if the principle fails for this case, it would seem to fail in general.

Now it seems impossible to avoid a certain arbitrariness in actually identifying the least favored group. One possibility is to choose a particular social position, say that of the unskilled worker, and then to count as the least favored all those with approximately the income and wealth of those in this position, or less. Another criterion is one in terms of relative income and wealth with no reference to social positions. For example, all persons with less than half of the median may be regarded as the least advantaged segment. This criterion depends only on the lower half of the distribution and has the merit of focusing attention on the social distance between those who have the least and the average citizen.[15] Either of these criteria would appear to cover those most disfavored by the various contingencies and provide a basis for determining at what level a reasonable social minimum might be set and from which, in conjunction with other measures, society could proceed to fulfill the difference principle. Any procedure is bound to be somewhat ad hoc. Yet we are entitled at some point to plead practical considerations, for sooner or later the capacity of philosophical or other arguments to make finer discriminations must run out. I assume that the persons in the original position understand these matters, and that they assess the difference principle in comparison with the other alternatives accordingly.[16]

As far as possible, then, justice as fairness appraises the social system from the position of equal citizenship and the various levels of income and wealth. Sometimes, however, other positions may need to be taken into account. If, for example, there are unequal basic rights founded on fixed natural characteristics, these inequalities will single out relevant positions. Since these characteristics cannot be changed, the positions

15. For this definition, see M. J. Bowman's discussion of the so-called Fuchs criterion in "Poverty in an Affluent Society," an essay in *Contemporary Economic Issues,* ed. N. W. Chamberlain (Homewood, Ill., R. D. Irwin, 1969), pp. 53–56.

16. I am indebted to Scott Boorman for clarification on this point.

they define count as starting places in the basic structure. Distinctions based on sex are of this type, and so are those depending upon race and culture. Thus if, say, men are favored in the assignment of basic rights, this inequality is justified by the difference principle (in the general interpretation) only if it is to the advantage of women and acceptable from their standpoint. And the analogous condition applies to the justification of caste systems, or racial and ethnic inequalities (§39). Such inequalities multiply relevant positions and complicate the application of the two principles. On the other hand, these inequalities are seldom, if ever, to the advantage of the less favored, and therefore in a just society the smaller number of relevant positions should ordinarily suffice.

Now it is essential that the judgments made from the perspective of the relevant positions override the claims that we are prone to make in more particular situations. Not everyone always benefits by what the two principles require if we think of ourselves in terms of our more specific positions. And unless the viewpoint of the relevant positions has priority, one still has a chaos of competing claims. Thus the two principles express, in effect, an understanding to order our interests by giving certain of them a special weight. For example, persons engaged in a particular industry often find that free trade is contrary to their interests. Perhaps the industry cannot remain prosperous without tariffs or other restrictions. But if free trade is desirable from the point of view of equal citizens or of the least advantaged, it is justified even though more specific interests temporarily suffer. For we are to agree in advance to the principles of justice and their consistent application from the standpoint of certain positions. There is no way to guarantee the protection of every interest over each period of time once the situation of representative men is defined more narrowly. Having acknowledged certain principles and a certain way of applying them, we are bound to accept the consequences. This does not mean, of course, that the rigors of free trade should be allowed to go unchecked. But the arrangements for softening them are to be considered from an appropriately general perspective.

The relevant social positions specify, then, the general point of view from which the two principles of justice are to be applied to the basic structure. In this way everyone's interests are taken into account, for each person is an equal citizen and all have a place in the distribution of income and wealth or in the range of fixed natural characteristics upon which distinctions are based. Some selection of relevant positions is necessary for a coherent theory of social justice and the ones chosen should accord with its first principles. By selecting the so-called starting

places one follows out the idea of mitigating the effects of natural accident and social circumstance. No one is to benefit from these contingencies except in ways that redound to the well-being of others.

17. THE TENDENCY TO EQUALITY

I wish to conclude this discussion of the two principles by explaining the sense in which they express an egalitarian conception of justice. Also I should like to forestall the objection to the principle of fair opportunity that it leads to a meritocratic society. In order to prepare the way for doing this, I note several aspects of the conception of justice that I have set out.

First we may observe that the difference principle gives some weight to the considerations singled out by the principle of redress. This is the principle that undeserved inequalities call for redress; and since inequalities of birth and natural endowment are undeserved, these inequalities are to be somehow compensated for.[17] Thus the principle holds that in order to treat all persons equally, to provide genuine equality of opportunity, society must give more attention to those with fewer native assets and to those born into the less favorable social positions. The idea is to redress the bias of contingencies in the direction of equality. In pursuit of this principle greater resources might be spent on the education of the less rather than the more intelligent, at least over a certain time of life, say the earlier years of school.

Now the principle of redress has not to my knowledge been proposed as the sole criterion of justice, as the single aim of the social order. It is plausible as most such principles are only as a prima facie principle, one that is to be weighed in the balance with others. For example, we are to weigh it against the principle to improve the average standard of life, or to advance the common good.[18] But whatever other principles we hold, the claims of redress are to be taken into account. It is thought to represent one of the elements in our conception of justice. Now the difference principle is not of course the principle of redress. It does not require society to try to even out handicaps as if all were expected to compete on a fair basis in the same race. But the difference principle would allocate

17. See Herbert Spiegelberg, "A Defense of Human Equality," *Philosophical Review,* vol. 53 (1944), pp. 101, 113–123; and D. D. Raphael, "Justice and Liberty," *Proceedings of the Aristotelian Society,* vol. 51 (1950–1951), pp. 187f.

18. See, for example, Spiegelberg, pp. 120f.

resources in education, say, so as to improve the long-term expectation of the least favored. If this end is attained by giving more attention to the better endowed, it is permissible; otherwise not. And in making this decision, the value of education should not be assessed solely in terms of economic efficiency and social welfare. Equally if not more important is the role of education in enabling a person to enjoy the culture of his society and to take part in its affairs, and in this way to provide for each individual a secure sense of his own worth.

Thus although the difference principle is not the same as that of redress, it does achieve some of the intent of the latter principle. It transforms the aims of the basic structure so that the total scheme of institutions no longer emphasizes social efficiency and technocratic values. The difference principle represents, in effect, an agreement to regard the distribution of natural talents as in some respects a common asset and to share in the greater social and economic benefits made possible by the complementarities of this distribution. Those who have been favored by nature, whoever they are, may gain from their good fortune only on terms that improve the situation of those who have lost out. The naturally advantaged are not to gain merely because they are more gifted, but only to cover the costs of training and education and for using their endowments in ways that help the less fortunate as well. No one deserves his greater natural capacity nor merits a more favorable starting place in society. But, of course, this is no reason to ignore, much less to eliminate these distinctions. Instead, the basic structure can be arranged so that these contingencies work for the good of the least fortunate. Thus we are led to the difference principle if we wish to set up the social system so that no one gains or loses from his arbitrary place in the distribution of natural assets or his initial position in society without giving or receiving compensating advantages in return.

In view of these remarks we may reject the contention that the ordering of institutions is always defective because the distribution of natural talents and the contingencies of social circumstance are unjust, and this injustice must inevitably carry over to human arrangements. Occasionally this reflection is offered as an excuse for ignoring injustice, as if the refusal to acquiesce in injustice is on a par with being unable to accept death. The natural distribution is neither just nor unjust; nor is it unjust that persons are born into society at some particular position. These are simply natural facts. What is just and unjust is the way that institutions deal with these facts. Aristocratic and caste societies are unjust because they make these contingencies the ascriptive basis for belonging to more

or less enclosed and privileged social classes. The basic structure of these societies incorporates the arbitrariness found in nature. But there is no necessity for men to resign themselves to these contingencies. The social system is not an unchangeable order beyond human control but a pattern of human action. In justice as fairness men agree to avail themselves of the accidents of nature and social circumstance only when doing so is for the common benefit. The two principles are a fair way of meeting the arbitrariness of fortune; and while no doubt imperfect in other ways, the institutions which satisfy these principles are just.

A further point is that the difference principle expresses a conception of reciprocity. It is a principle of mutual benefit. At first sight, however, it may appear unfairly biased towards the least favored. To consider this question in an intuitive way, suppose for simplicity that there are only two groups in society, one noticeably more fortunate than the other. Subject to the usual constraints (defined by the priority of the first principle and fair equality of opportunity), society could maximize the expectations of either group but not both, since we can maximize with respect to only one aim at a time. It seems clear that society should not do the best it can for those initially more advantaged; so if we reject the difference principle, we must prefer maximizing some weighted mean of the two expectations. But if we give any weight to the more fortunate, we are valuing for their own sake the gains to those already more favored by natural and social contingencies. No one had an antecedent claim to be benefited in this way, and so to maximize a weighted mean is, so to speak, to favor the more fortunate twice over. Thus the more advantaged, when they view the matter from a general perspective, recognize that the well-being of each depends on a scheme of social cooperation without which no one could have a satisfactory life; they recognize also that they can expect the willing cooperation of all only if the terms of the scheme are reasonable. So they regard themselves as already compensated, as it were, by the advantages to which no one (including themselves) had a prior claim. They forego the idea of maximizing a weighted mean and regard the difference principle as a fair basis for regulating the basic structure.

One may object that those better situated deserve the greater advantages they could acquire for themselves under other schemes of cooperation whether or not these advantages are gained in ways that benefit others. Now it is true that given a just system of cooperation as a framework of public rules, and the expectations set up by it, those who, with the prospect of improving their condition, have done what the system announces it will reward are entitled to have their expectations met. In

this sense the more fortunate have title to their better situation; their claims are legitimate expectations established by social institutions and the community is obligated to fulfill them. But this sense of desert is that of entitlement. It presupposes the existence of an ongoing cooperative scheme and is irrelevant to the question whether this scheme itself is to be designed in accordance with the difference principle or some other criterion (§48).

Thus it is incorrect that individuals with greater natural endowments and the superior character that has made their development possible have a right to a cooperative scheme that enables them to obtain even further benefits in ways that do not contribute to the advantages of others. We do not deserve our place in the distribution of native endowments, any more than we deserve our initial starting place in society. That we deserve the superior character that enables us to make the effort to cultivate our abilities is also problematic; for such character depends in good part upon fortunate family and social circumstances in early life for which we can claim no credit. The notion of desert does not apply here. To be sure, the more advantaged have a right to their natural assets, as does everyone else; this right is covered by the first principle under the basic liberty protecting the integrity of the person. And so the more advantaged are entitled to whatever they can acquire in accordance with the rules of a fair system of social cooperation. Our problem is how this scheme, the basic structure of society, is to be designed. From a suitably general standpoint, the difference principle appears acceptable to both the more advantaged and the less advantaged individual. Of course, none of this is strictly speaking an argument for the principle, since in a contract theory arguments are made from the point of view of the original position. But these intuitive considerations help to clarify the principle and the sense in which it is egalitarian.

I noted earlier (§13) that a society should try to avoid the region where the marginal contributions of those better off to the well-being of the less favored are negative. It should operate only on the upward rising part of the contribution curve (including of course the maximum). On this segment of the curve the criterion of mutual benefit is always fulfilled. Moreover, there is a natural sense in which the harmony of social interests is achieved; representative men do not gain at one another's expense since only reciprocal advantages are allowed. To be sure, the shape and slope of the contribution curve is determined in part at least by the natural lottery in native assets, and as such it is neither just nor unjust. But suppose we think of the forty-five degree line as representing the ideal of

a perfect harmony of interests; it is the contribution curve (a straight line in this case) along which everyone gains equally. Then it seems that the consistent realization of the two principles of justice tends to raise the curve closer to the ideal of a perfect harmony of interests. Once a society goes beyond the maximum it operates along the downward sloping part of the curve and a harmony of interests no longer exists. As the more favored gain the less advantaged lose, and vice versa. Thus it is to realize the ideal of the harmony of interests on terms that nature has given us, and to meet the criterion of mutual benefit, that we should stay in the region of positive contributions.

A further merit of the difference principle is that it provides an interpretation of the principle of fraternity. In comparison with liberty and equality, the idea of fraternity has had a lesser place in democratic theory. It is thought to be less specifically a political concept, not in itself defining any of the democratic rights but conveying instead certain attitudes of mind and forms of conduct without which we would lose sight of the values expressed by these rights.[19] Or closely related to this, fraternity is held to represent a certain equality of social esteem manifest in various public conventions and in the absence of manners of deference and servility.[20] No doubt fraternity does imply these things, as well as a sense of civic friendship and social solidarity, but so understood it expresses no definite requirement. We have yet to find a principle of justice that matches the underlying idea. The difference principle, however, does seem to correspond to a natural meaning of fraternity: namely, to the idea of not wanting to have greater advantages unless this is to the benefit of others who are less well off. The family, in its ideal conception and often in practice, is one place where the principle of maximizing the sum of advantages is rejected. Members of a family commonly do not wish to gain unless they can do so in ways that further the interests of the rest. Now wanting to act on the difference principle has precisely this consequence. Those better circumstanced are willing to have their greater advantages only under a scheme in which this works out for the benefit of the less fortunate.

The ideal of fraternity is sometimes thought to involve ties of sentiment and feeling which it is unrealistic to expect between members of the wider society. And this is surely a further reason for its relative neglect in

19. See J. R. Pennock, *Liberal Democracy: Its Merits and Prospects* (New York, Rinehart, 1950), pp. 94f.

20. See R. B. Perry, *Puritanism and Democracy* (New York, The Vanguard Press, 1944), ch. XIX, sec. 8.

democratic theory. Many have felt that it has no proper place in political affairs. But if it is interpreted as incorporating the requirements of the difference principle, it is not an impracticable conception. It does seem that the institutions and policies which we most confidently think to be just satisfy its demands, at least in the sense that the inequalities permitted by them contribute to the well-being of the less favored. Or at any rate, so I shall try to make plausible in Chapter V. On this interpretation, then, the principle of fraternity is a perfectly feasible standard. Once we accept it we can associate the traditional ideas of liberty, equality, and fraternity with the democratic interpretation of the two principles of justice as follows: liberty corresponds to the first principle, equality to the idea of equality in the first principle together with equality of fair opportunity, and fraternity to the difference principle. In this way we have found a place for the conception of fraternity in the democratic interpretation of the two principles, and we see that it imposes a definite requirement on the basic structure of society. The other aspects of fraternity should not be forgotten, but the difference principle expresses its fundamental meaning from the standpoint of social justice.

Now it seems evident in the light of these observations that the democratic interpretation of the two principles will not lead to a meritocratic society.[21] This form of social order follows the principle of careers open to talents and uses equality of opportunity as a way of releasing men's energies in the pursuit of economic prosperity and political dominion. There exists a marked disparity between the upper and lower classes in both means of life and the rights and privileges of organizational authority. The culture of the poorer strata is impoverished while that of the governing and technocratic elite is securely based on the service of the national ends of power and wealth. Equality of opportunity means an equal chance to leave the less fortunate behind in the personal quest for influence and social position.[22] Thus a meritocratic society is a danger for the other interpretations of the principles of justice but not for the democratic conception. For, as we have just seen, the difference principle transforms the aims of society in fundamental respects. This consequence is even more obvious once we note that we must when necessary take into account the essential primary good of self-respect and the fact that a

21. The problem of a meritocratic society is the subject of Michael Young's fantasy, *The Rise of Meritocracy* (London, Thames and Hudson, 1958).

22. For elaborations of this point to which I am indebted, see John Schaar, "Equality of Opportunity and Beyond," *Nomos IX: Equality,* ed. by J. R. Pennock and J. W. Chapman (New York, Atherton Press, 1967); and B. A. O. Williams, "The Idea of Equality," pp. 125–129.

well-ordered society is a social union of social unions (§79). It follows that the confident sense of their own worth should be sought for the least favored and this limits the forms of hierarchy and the degrees of inequality that justice permits. Thus, for example, resources for education are not to be allotted solely or necessarily mainly according to their return as estimated in productive trained abilities, but also according to their worth in enriching the personal and social life of citizens, including here the less favored. As a society progresses the latter consideration becomes increasingly more important.

These remarks must suffice to sketch the conception of social justice expressed by the two principles for institutions. Before taking up the principles for individuals I should mention one further question. I have assumed so far that the distribution of natural assets is a fact of nature and that no attempt is made to change it, or even to take it into account. But to some extent this distribution is bound to be affected by the social system. A caste system, for example, tends to divide society into separate biological populations, while an open society encourages the widest genetic diversity.[23] In addition, it is possible to adopt eugenic policies, more or less explicit. I shall not consider questions of eugenics, confining myself throughout to the traditional concerns of social justice. We should note, though, that it is not in general to the advantage of the less fortunate to propose policies which reduce the talents of others. Instead, by accepting the difference principle, they view the greater abilities as a social asset to be used for the common advantage. But it is also in the interest of each to have greater natural assets. This enables him to pursue a preferred plan of life. In the original position, then, the parties want to insure for their descendants the best genetic endowment (assuming their own to be fixed). The pursuit of reasonable policies in this regard is something that earlier generations owe to later ones, this being a question that arises between generations. Thus over time a society is to take steps at least to preserve the general level of natural abilities and to prevent the diffusion of serious defects. These measures are to be guided by principles that the parties would be willing to consent to for the sake of their successors. I mention this speculative and difficult matter to indicate once again the manner in which the difference principle is likely to transform problems of social justice. We might conjecture that in the long run, if there is an upper bound on ability, we would eventually reach a society with the

23. See Theodosius Dobzhansky, *Mankind Evolving* (New Haven, Yale University Press, 1962), pp. 242–252, for a discussion of this question.

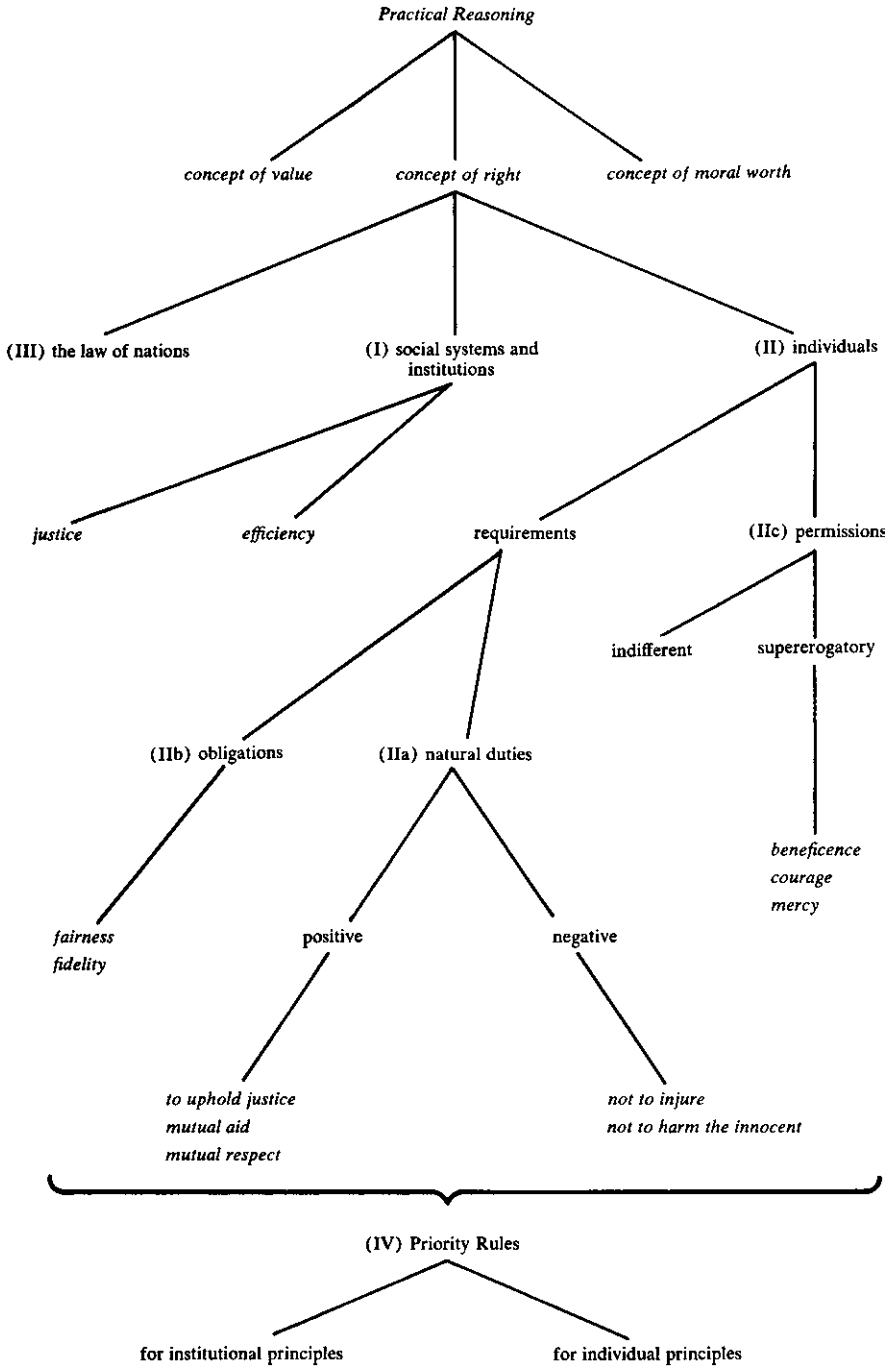
greatest equal liberty the members of which enjoy the greatest equal talent. But I shall not pursue this thought further.

18. PRINCIPLES FOR INDIVIDUALS: THE PRINCIPLE OF FAIRNESS

In the discussion so far I have considered the principles which apply to institutions or, more exactly, to the basic structure of society. It is clear, however, that principles of another kind must also be chosen, since a complete theory of right includes principles for individuals as well. In fact, as the accompanying diagram indicates, one needs in addition principles for the law of nations and of course priority rules for assigning weights when principles conflict. I shall not take up the principles for the law of nations, except in passing (§58); nor shall I attempt any systematic discussion of the principles for individuals. But certain principles of this type are an essential part of any theory of justice. In this and the next section the meaning of several of these principles is explained, although the examination of the reasons for choosing them is postponed until later (§§51–52).

The accompanying diagram is purely schematic. It is not suggested that the principles associated with the concepts lower down in the tree are deduced from the higher ones. The diagram simply indicates the kinds of principles that must be chosen before a full conception of right is on hand. The Roman numerals express the order in which the various sorts of principles are to be acknowledged in the original position. Thus the principles for the basic structure of society are to be agreed to first, principles for individuals next, followed by those for the law of nations. Last of all the priority rules are adopted, although we may tentatively choose these earlier contingent on subsequent revision.

Now the order in which principles are chosen raises a number of questions which I shall skip over. The important thing is that the various principles are to be adopted in a definite sequence and the reasons for this ordering are connected with the more difficult parts of the theory of justice. To illustrate: while it would be possible to choose many of the natural duties before those for the basic structure without changing the principles in any substantial way, the sequence in either case reflects the fact that obligations presuppose principles for social forms. And some natural duties also presuppose such principles, for example, the duty to support just institutions. For this reason it seems simpler to adopt all principles


Practical Reasoning
concept of value
concept of right
concept of moral worth
(III) the law of nations
(I) social systems and institutions
(II) individuals
justice
efficiency
requirements
(IIc) permissions
indifferent
supererogatory
(IIb) obligations
(IIa) natural duties
beneficence
courage
mercy
fairness
fidelity
positive
negative
to uphold justice
mutual aid
mutual respect
not to injure
not to harm the innocent
(IV) Priority Rules
for institutional principles
for individual principles

for individuals after those for the basic structure. That principles for institutions are chosen first shows the social nature of the virtue of justice, its intimate connection with social practices so often noted by idealists. When Bradley says that the individual is a bare abstraction, he can be interpreted to say, without too much distortion, that a person's obligations and duties presuppose a moral conception of institutions and therefore that the content of just institutions must be defined before the requirements for individuals can be set out.[24] And this is to say that, in most cases, the principles for obligations and duties should be settled upon after those for the basic structure.

Therefore, to establish a complete conception of right, the parties in the original position are to choose in a definite order not only a conception of justice but also principles to go with each major concept falling under the concept of right. These concepts are I assume relatively few in number and have a determinate relation to each other. Thus, in addition to principles for institutions there must be an agreement on principles for such notions as fairness and fidelity, mutual respect and beneficence as these apply to individuals, as well as on principles for the conduct of states. The intuitive idea is this: the concept of something's being right is the same as, or better, may be replaced by, the concept of its being in accordance with the principles that in the original position would be acknowledged to apply to things of its kind. I do not interpret this concept of right as providing an analysis of the meaning of the term "right" as normally used in moral contexts. It is not meant as an analysis of the concept of right in the traditional sense. Rather, the broader notion of rightness as fairness is to be understood as a replacement for existing conceptions. There is no necessity to say that sameness of meaning holds between the word "right" (and its relatives) in its ordinary use and the more elaborate locutions needed to express this ideal contractarian concept of right. For our purposes here I accept the view that a sound analysis is best understood as providing a satisfactory substitute, one that meets certain desiderata while avoiding certain obscurities and confusions. In other words, explication is elimination: we start with a concept the expression for which is somehow troublesome; but it serves certain ends that cannot be given up. An explication achieves these ends in other ways that are relatively free of difficulty.[25] Thus if the theory of justice as fairness, or more generally of rightness as fairness, fits our considered

24. See F. H. Bradley, *Ethical Studies,* 2nd ed. (Oxford, The Clarendon Press, 1927), pp. 163–189.

25. See W. V. Quine, *Word and Object* (Cambridge, Mass., M.I.T. Press, 1960), pp. 257–262, whom I follow here.

judgments in reflective equilibrium, and if it enables us to say all that on due examination we want to say, then it provides a way of eliminating customary phrases in favor of other expressions. So understood one may think of justice as fairness and rightness as fairness as providing a definition or explication of the concepts of justice and right.

I now turn to one of the principles that applies to individuals, the principle of fairness. I shall try to use this principle to account for all requirements that are obligations as distinct from natural duties. This principle holds that a person is required to do his part as defined by the rules of an institution when two conditions are met: first, the institution is just (or fair), that is, it satisfies the two principles of justice; and second, one has voluntarily accepted the benefits of the arrangement or taken advantage of the opportunities it offers to further one's interests. The main idea is that when a number of persons engage in a mutually advantageous cooperative venture according to rules, and thus restrict their liberty in ways necessary to yield advantages for all, those who have submitted to these restrictions have a right to a similar acquiescence on the part of those who have benefited from their submission.[26] We are not to gain from the cooperative labors of others without doing our fair share. The two principles of justice define what is a fair share in the case of institutions belonging to the basic structure. So if these arrangements are just, each person receives a fair share when all (himself included) do their part.

Now by definition the requirements specified by the principle of fairness are the obligations. All obligations arise in this way. It is important, however, to note that the principle of fairness has two parts, the first which states that the institutions or practices in question must be just, the second which characterizes the requisite voluntary acts. The first part formulates the conditions necessary if these voluntary acts are to give rise to obligations. By the principle of fairness it is not possible to be bound to unjust institutions, or at least to institutions which exceed the limits of tolerable injustice (so far undefined). In particular, it is not possible to have an obligation to autocratic and arbitrary forms of government. The necessary background does not exist for obligations to arise from consensual or other acts, however expressed. Obligatory ties presuppose just institutions, or ones reasonably just in view of the circumstances. It is, therefore, a mistake to argue against justice as fairness and contract

26. I am indebted here to H. L. A. Hart, "Are There Any Natural Rights?" *Philosophical Review,* vol. 64 (1955), pp. 185f.

theories generally that they have the consequence that citizens are under an obligation to unjust regimes which coerce their consent or win their tacit acquiescence in more refined ways. Locke especially has been the object of this mistaken criticism which overlooks the necessity for certain background conditions.[27]

There are several characteristic features of obligations which distinguish them from other moral requirements. For one thing, they arise as a result of our voluntary acts; these acts may be the giving of express or tacit undertakings, such as promises and agreements, but they need not be, as in the case of accepting benefits. Further, the content of obligations is always defined by an institution or practice the rules of which specify what it is that one is required to do. And finally, obligations are normally owed to definite individuals, namely, those who are cooperating together to maintain the arrangement in question.[28] As an example illustrating these features, consider the political act of running for and (if successful) holding public office in a constitutional regime. This act gives rise to the obligation to fulfill the duties of office, and these duties determine the content of the obligation. Here I think of duties not as moral duties but as tasks and responsibilities assigned to certain institutional positions. It is nevertheless the case that one may have a moral reason (one based on a moral principle) for discharging these duties, as when one is bound to do so by the principle of fairness. Also, one who assumes public office is obligated to his fellow citizens whose trust and confidence he has sought and with whom he is cooperating in running a democratic society. Similarly, we assume obligations when we marry as well as when we accept positions of judicial, administrative, or other authority. We acquire obligations by promising and by tacit understandings, and even when we join a game, namely, the obligation to play by the rules and to be a good sport.

All of these obligations are, I believe, covered by the principle of fairness. There are two important cases though that are somewhat problematical, namely, political obligation as it applies to the average citizen, rather than, say, to those who hold office, and the obligation to keep

27. Locke holds that conquest gives no right, nor does violence and injury however much "colored with the name, pretences, or forms of law." *Second Treatise of Government,* pars. 176, 20. See Hanna Pitkin's discussion of Locke in "Obligation and Consent I," *American Political Science Review,* vol. 59 (1965), esp. pp. 994–997, the essentials of which I accept.

28. In distinguishing between obligations and natural duties I have drawn upon H. L. A. Hart, "Legal and Moral Obligation," in *Essays in Moral Philosophy,* ed. by A. I. Melden (Seattle, University of Washington Press, 1958), pp. 100–105; C. H. Whiteley, "On Duties," *Proceedings of the Aristotelian Society,* vol. 53 (1952–53); and R. B. Brandt, "The Concepts of Obligation and Duty," *Mind,* vol. 73 (1964).

promises. In the first case it is not clear what is the requisite binding action or who has performed it. There is, I believe, no political obligation, strictly speaking, for citizens generally. In the second case an explanation is needed as to how fiduciary obligations arise from taking advantage of a just practice. We need to look into the nature of the relevant practice in this instance. These matters I shall discuss at another place (§§51–52).

19. PRINCIPLES FOR INDIVIDUALS: THE NATURAL DUTIES

Whereas all obligations are accounted for by the principle of fairness, there are many natural duties, positive and negative. I shall make no attempt to bring them under one principle. Admittedly this lack of unity runs the risk of putting too much strain on priority rules, but I shall have to leave this difficulty aside. The following are examples of natural duties: the duty of helping another when he is in need or jeopardy, provided that one can do so without excessive risk or loss to oneself; the duty not to harm or injure another; and the duty not to cause unnecessary suffering. The first of these duties, the duty of mutual aid, is a positive duty in that it is a duty to do something good for another; whereas the last two duties are negative in that they require us not to do something that is bad. The distinction between positive and negative duties is intuitively clear in many cases, but often gives way. I shall not put any stress upon it. The distinction is important only in connection with the priority problem, since it seems plausible to hold that, when the distinction is clear, negative duties have more weight than positive ones. But I shall not pursue this question here.

Now in contrast with obligations, it is characteristic of natural duties that they apply to us without regard to our voluntary acts. Moreover, they have no necessary connection with institutions or social practices; their content is not, in general, defined by the rules of these arrangements. Thus we have a natural duty not to be cruel, and a duty to help another, whether or not we have committed ourselves to these actions. It is no defense or excuse to say that we have made no promise not to be cruel or vindictive, or to come to another's aid. Indeed, a promise not to kill, for example, is normally ludicrously redundant, and the suggestion that it establishes a moral requirement where none already existed is mistaken. Such a promise is in order, if it ever is so, only when for special reasons one has the right to kill, perhaps in a situation arising in a just war. A

further feature of natural duties is that they hold between persons irrespective of their institutional relationships; they obtain between all as equal moral persons. In this sense the natural duties are owed not only to definite individuals, say to those cooperating together in a particular social arrangement, but to persons generally. This feature in particular suggests the propriety of the adjective "natural." One aim of the law of nations is to assure the recognition of these duties in the conduct of states. This is especially important in constraining the means used in war, assuming that, in certain circumstances anyway, wars of self-defense are justified (§58).

From the standpoint of justice as fairness, a fundamental natural duty is the duty of justice. This duty requires us to support and to comply with just institutions that exist and apply to us. It also constrains us to further just arrangements not yet established, at least when this can be done without too much cost to ourselves. Thus if the basic structure of society is just, or as just as it is reasonable to expect in the circumstances, everyone has a natural duty to do his part in the existing scheme. Each is bound to these institutions independent of his voluntary acts, performative or otherwise. Thus even though the principles of natural duty are derived from a contractarian point of view, they do not presuppose an act of consent, express or tacit, or indeed any voluntary act, in order to apply. The principles that hold for individuals, just as the principles for institutions, are those that would be acknowledged in the original position. These principles are understood as the outcome of a hypothetical agreement. If their formulation shows that no binding action, consensual or otherwise, is a presupposition of their application, then they apply unconditionally. The reason why obligations depend upon voluntary acts is given by the second part of the principle of fairness which states this condition. It has nothing to do with the contractual nature of justice as fairness.[29] In fact, once the full set of principles, a complete conception of right, is on hand, we can simply forget about the conception of original position and apply these principles as we would any others.

29. For clarification on these points I am indebted to Robert Amdur. Views seeking to derive political ties solely from consensual acts are found in Michael Walzer, *Obligations: Essays on Disobedience, War, and Citizenship* (Cambridge, Mass., Harvard University Press, 1970), esp. pp. ix-xvi, 7–10, 18–21, and ch. 5; and Joseph Tussman, *Obligation and the Body Politic* (New York, Oxford University Press, 1960). On the latter, see Hanna Pitkin, "Obligation and Consent I," pp. 997f. For further discussions of the problems of consent theory in addition to Pitkin, see Alan Gewirth, "Political Justice." in *Social Justice,* ed. R. B. Brandt (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1962), pp. 128–141; and J. P. Plamenatz, *Consent, Freedom, and Political Obligation,* 2nd ed. (London, Oxford University Press, 1968).

There is nothing inconsistent, or even surprising, in the fact that justice as fairness allows unconditional principles. It suffices to show that the parties in the original position would agree to principles defining the natural duties which as formulated hold unconditionally. We should note that, since the principle of fairness may establish a bond to existing just arrangements, the obligations covered by it can support a tie already present that derives from the natural duty of justice. Thus a person may have both a natural duty and an obligation to comply with an institution and to do his part. The thing to observe here is that there are several ways in which one may be bound to political institutions. For the most part the natural duty of justice is the more fundamental, since it binds citizens generally and requires no voluntary acts in order to apply. The principle of fairness, on the other hand, binds only those who assume public office, say, or those who, being better situated, have advanced their aims within the system. There is, then, another sense of *noblesse oblige:* namely, that those who are more privileged are likely to acquire obligations tying them even more strongly to a just scheme.

I shall say very little about the other kind of principles for individuals. For while permissions are not an unimportant class of actions, I must limit the discussion to the theory of social justice. It may be observed, though, that once all the principles defining requirements are chosen, no further acknowledgments are necessary to define permissions. This is so because permissions are those acts which we are at liberty both to do and not to do. They are acts which violate no obligation or natural duty. In studying permissions one wishes to single out those that are significant from a moral point of view and to explain their relation to duties and obligations. Many such actions are morally indifferent or trivial. But among permissions is the interesting class of supererogatory actions. These are acts of benevolence and mercy, of heroism and self-sacrifice. It is good to do these actions but it is not one's duty or obligation. Supererogatory acts are not required, though normally they would be were it not for the loss or risk involved for the agent himself. A person who does a supererogatory act does not invoke the exemption which the natural duties allow. For while we have a natural duty to bring about a great good, say, if we can do so relatively easily, we are released from this duty when the cost to ourselves is considerable. Supererogatory acts raise questions of first importance for ethical theory. For example, it seems offhand that the classical utilitarian view cannot account for them. It would appear that we are bound to perform actions which bring about a greater good for others whatever the cost to ourselves provided that the sum of advantages

altogether exceeds that of other acts open to us. There is nothing corresponding to the exemptions included in the formulation of the natural duties. Thus some of the actions which justice as fairness counts as supererogatory may be required by the utility principle. I shall not, however, pursue this matter further. Supererogatory acts are mentioned here for the sake of completeness. We must now turn to the interpretation of the initial situation.

CHAPTER III. THE ORIGINAL POSITION

In this chapter I discuss the favored philosophical interpretation of the initial situation. I refer to this interpretation as the original position. I begin by sketching the nature of the argument for conceptions of justice and explaining how the alternatives are presented so that the parties are to choose from a definite list of traditional conceptions. Then I describe the conditions which characterize the initial situation under several headings: the circumstances of justice, the formal constraints of the concept of right, the veil of ignorance, and the rationality of the contracting parties. In each case I try to indicate why the features adopted for the favored interpretation are reasonable from a philosophical point of view. Next the natural lines of reasoning leading to the two principles of justice and to the principle of average utility are examined prior to a consideration of the relative advantages of these conceptions of justice. I argue that the two principles would be acknowledged and set out some of the main grounds to support this contention. In order to clarify the differences between the various conceptions of justice, the chapter concludes with another look at the classical principle of utility.

20. THE NATURE OF THE ARGUMENT FOR CONCEPTIONS OF JUSTICE

The intuitive idea of justice as fairness is to think of the first principles of justice as themselves the object of an original agreement in a suitably defined initial situation. These principles are those which rational persons concerned to advance their interests would accept in this position of equality to settle the basic terms of their association. It must be shown, then, that the two principles of justice are the solution for the problem of choice presented by the original position. In order to do this, one must establish that, given the circumstances of the parties, and their knowl-

edge, beliefs, and interests, an agreement on these principles is the best way for each person to secure his ends in view of the alternatives available.

Now obviously no one can obtain everything he wants; the mere existence of other persons prevents this. The absolutely best for any man is that everyone else should join with him in furthering his conception of the good whatever it turns out to be. Or failing this, that all others are required to act justly but that he is authorized to exempt himself as he pleases. Since other persons will never agree to such terms of association these forms of egoism would be rejected. The two principles of justice, however, seem to be a reasonable proposal. In fact, I should like to show that these principles are everyone's best reply, so to speak, to the corresponding demands of the others. In this sense, the choice of this conception of justice is the unique solution to the problem set by the original position.

By arguing in this way one follows a procedure familiar in social theory. That is, a simplified situation is described in which rational individuals with certain ends and related to each other in certain ways are to choose among various courses of action in view of their knowledge of the circumstances. What these individuals will do is then derived by strictly deductive reasoning from these assumptions about their beliefs and interests, their situation and the options open to them. Their conduct is, in the phrase of Pareto, the resultant of tastes and obstacles.[1] In the theory of price, for example, the equilibrium of competitive markets is thought of as arising when many individuals each advancing his own interests give way to each other what they can best part with in return for what they most desire. Equilibrium is the result of agreements freely struck between willing traders. For each person it is the best situation that he can reach by free exchange consistent with the right and freedom of others to further their interests in the same way. It is for this reason that this state of affairs is an equilibrium, one that will persist in the absence of further changes in the circumstances. No one has any incentive to alter it. If a departure from this situation sets in motion tendencies which restore it, the equilibrium is stable.

Of course, the fact that a situation is one of equilibrium, even a stable one, does not entail that it is right or just. It only means that given men's estimate of their position, they act effectively to preserve it. Clearly a

1. *Manuel d'économie politique* (Paris, 1909), ch. III, §23. Pareto says: "L'equilibre résulte précisément de cette opposition des goûts et des obstacles."

balance of hatred and hostility may be a stable equilibrium; each may think that any feasible change will be worse. The best that each can do for himself may be a condition of lesser injustice rather than of greater good. The moral assessment of equilibrium situations depends upon the background circumstances which determine them. It is at this point that the conception of the original position embodies features peculiar to moral theory. For while the theory of price, say, tries to account for the movements of the market by assumptions about the actual tendencies at work, the philosophically favored interpretation of the initial situation incorporates conditions which it is thought reasonable to impose on the choice of principles. By contrast with social theory, the aim is to characterize this situation so that the principles that would be chosen, whatever they turn out to be, are acceptable from a moral point of view. The original position is defined in such a way that it is a status quo in which any agreements reached are fair. It is a state of affairs in which the parties are equally represented as moral persons and the outcome is not conditioned by arbitrary contingencies or the relative balance of social forces. Thus justice as fairness is able to use the idea of pure procedural justice from the beginning.

It is clear, then, that the original position is a purely hypothetical situation. Nothing resembling it need ever take place, although we can by deliberately following the constraints it expresses simulate the reflections of the parties. The conception of the original position is not intended to explain human conduct except insofar as it tries to account for our moral judgments and helps to explain our having a sense of justice. Justice as fairness is a theory of our moral sentiments as manifested by our considered judgments in reflective equilibrium. These sentiments presumably affect our thought and action to some degree. So while the conception of the original position is part of the theory of conduct, it does not follow at all that there are actual situations that resemble it. What is necessary is that the principles that would be accepted play the requisite part in our moral reasoning and conduct.

One should note also that the acceptance of these principles is not conjectured as a psychological law or probability. Ideally anyway, I should like to show that their acknowledgment is the only choice consistent with the full description of the original position. The argument aims eventually to be strictly deductive. To be sure, the persons in the original position have a certain psychology, since various assumptions are made about their beliefs and interests. These assumptions appear along with other premises in the description of this initial situation. But clearly arguments

from such premises can be fully deductive, as theories in politics and economics attest. We should strive for a kind of moral geometry with all the rigor which this name connotes. Unhappily the reasoning I shall give will fall far short of this, since it is highly intuitive throughout. Yet it is essential to have in mind the ideal one would like to achieve.

A final remark. There are, as I have said, many possible interpretations of the initial situation. This conception varies depending upon how the contracting parties are conceived, upon what their beliefs and interests are said to be, upon which alternatives are available to them, and so on. In this sense, there are many different contract theories. Justice as fairness is but one of these. But the question of justification is settled, as far as it can be, by showing that there is one interpretation of the initial situation which best expresses the conditions that are widely thought reasonable to impose on the choice of principles yet which, at the same time, leads to a conception that characterizes our considered judgments in reflective equilibrium. This most favored, or standard, interpretation I shall refer to as the original position. We may conjecture that for each traditional conception of justice there exists an interpretation of the initial situation in which its principles are the preferred solution. Thus, for example, there are interpretations that lead to the classical as well as the average principle of utility. These variations of the initial situation will be mentioned as we go along. The procedure of contract theories provides, then, a general analytic method for the comparative study of conceptions of justice. One tries to set out the different conditions embodied in the contractual situation in which their principles would be chosen. In this way one formulates the various underlying assumptions on which these conceptions seem to depend. But if one interpretation is philosophically most favored, and if its principles characterize our considered judgments, we have a procedure for justification as well. We cannot know at first whether such an interpretation exists, but at least we know what to look for.

21. THE PRESENTATION OF ALTERNATIVES

Let us now turn from these remarks on method to the description of the original position. I shall begin with the question of the alternatives open to the persons in this situation. Ideally of course one would like to say that they are to choose among all possible conceptions of justice. One obvious difficulty is how these conceptions are to be characterized so that those in the original position can be presented with them. Yet granting

that these conceptions could be defined, there is no assurance that the parties could make out the best option; the principles that would be most preferred might be overlooked. Indeed, there may exist no best alternative: conceivably for each conception of justice there is another that is better. Even if there is a best alternative, it seems difficult to describe the parties' intellectual powers so that this optimum, or even the more plausible conceptions, are sure to occur to them. Some solutions to the choice problem may be clear enough on careful reflection; it is another matter to describe the parties so that their deliberations generate these alternatives. Thus although the two principles of justice may be superior to those conceptions known to us, perhaps some hitherto unformulated set of principles is still better.

In order to handle this problem I shall resort to the following device. I shall simply take as given a short list of traditional conceptions of justice, for example those discussed in the first chapter, together with a few other possibilities suggested by the two principles of justice. I then assume that the parties are presented with this list and required to agree unanimously that one conception is the best among those enumerated. We may suppose that this decision is arrived at by making a series of comparisons in pairs. Thus the two principles would be shown to be preferable once all agree that they are to be chosen over each of the other alternatives. In this chapter I shall consider for the most part the choice between the two principles of justice and two forms of the principle of utility (the classical and the average principle). Later on, the comparisons with perfectionism and mixed theories are discussed. In this manner I try to show that the two principles would be chosen from the list.

Now admittedly this is an unsatisfactory way to proceed. It would be better if we could define necessary and sufficient conditions for a uniquely best conception of justice and then exhibit a conception that fulfilled these conditions. Eventually one may be able to do this. For the time being, however, I do not see how to avoid rough and ready methods. Moreover, using such procedures may point to a general solution of our problem. Thus it may turn out that, as we run through these comparisons, the reasoning of the parties singles out certain features of the basic structure as desirable, and that these features have natural maximum and minimum properties. Suppose, for example, that it is rational for the persons in the original position to prefer a society with the greatest equal liberty. And suppose further that while they prefer social and economic advantages to work for the common good they insist that they mitigate the ways in which men are advantaged or disadvantaged by natural and

social contingencies. If these two features are the only relevant ones, and if the principle of equal liberty is the natural maximum of the first feature, and the difference principle (constrained by fair equality of opportunity) the natural maximum of the second, then, leaving aside the problem of priority, the two principles are the optimum solution. The fact that one cannot constructively characterize or enumerate all possible conceptions of justice, or describe the parties so that they are bound to think of them, is no obstacle to this conclusion.

It would not be profitable to pursue these speculations any further. For the present, no attempt is made to deal with the general problem of the best solution. I limit the argument throughout to the weaker contention that the two principles would be chosen from the conceptions of justice on the following list.

A. The Two Principles of Justice (in serial order)
    1. The principle of greatest equal liberty
    2. (a) The principle of (fair) equality of opportunity
       (b) The difference principle
B. Mixed Conceptions. Substitute one for A2 above
    1. The principle of average utility; or
    2. The principle of average utility, subject to a constraint, either:
       (a) That a certain social minimum be maintained, or
       (b) That the overall distribution not be too wide; or
    3. The principle of average utility subject to either constraint in B2 plus that of equality of fair opportunity
C. Classical Teleological Conceptions
    1. The classical principle of utility
    2. The average principle of utility
    3. The principle of perfection
D. Intuitionistic Conceptions
    1. To balance total utility against the principle of equal distribution
    2. To balance average utility against the principle of redress
    3. To balance a list of prima facie principles (as appropriate)
E. Egoistic Conceptions (See §23 where it is explained why strictly speaking the egoistic conceptions are not alternatives.)
    1. First-person dictatorship: Everyone is to serve my interests
    2. Free-rider: Everyone is to act justly except for myself, if I choose not to
    3. General: Everyone is permitted to advance his interests as he pleases

The merits of these traditional theories surely suffice to justify the effort to rank them. And in any case, the study of this ranking is a useful way of feeling one's way into the larger question. Now each of these conceptions presumably has its assets and liabilities; there are reasons for and against any alternative one selects. The fact that a conception is open to criticism is not necessarily decisive against it, nor are certain desirable features always conclusive in its favor. The decision of the persons in the original position hinges, as we shall see, on a balance of various considerations. In this sense, there is an appeal to intuition at the basis of the theory of justice. Yet when everything is tallied up, it may be perfectly clear where the balance of reason lies. The relevant reasons may have been so factored and analyzed by the description of the original position that one conception of justice is distinctly preferable to the others. The argument for it is not strictly speaking a proof, not yet anyway; but, in Mill's phrase, it may present considerations capable of determining the intellect.[2]

The list of conceptions is largely self-explanatory. A few brief comments, however, may be useful. Each conception is expressed in a reasonably simple way, and each holds unconditionally, that is, whatever the circumstances or state of society. None of the principles is contingent upon certain social or other conditions. Now one reason for this is to keep things simple. It would be easy to formulate a family of conceptions each designed to apply only if special circumstances obtain, these various conditions being exhaustive and mutually exclusive. For example one conception might hold at one stage of culture, a different conception at another. Such a family could be counted as itself a conception of justice; it would consist of a set of ordered pairs, each pair being a conception of justice matched with the circumstances in which it applies. But if conceptions of this kind were added to the list, our problem would become very complicated if not unmanageable. Moreover, there is a reason for excluding alternatives of this kind, for it is natural to ask what underlying principle determines the ordered pairs. Here I assume that some recognizably ethical conception specifies the appropriate principles given each of the conditions. It is really this unconditional principle that defines the conception expressed by the set of ordered pairs. Thus to allow such families on the list is to include alternatives that conceal their proper basis. So for this reason as well I shall exclude them. It also turns out to be desirable to characterize the original position so that the parties are to choose princi-

2. *Utilitarianism,* ch. I, par. 5.

ples that hold unconditionally whatever the circumstances. This fact is connected with the Kantian interpretation of justice as fairness. But I leave this matter aside until later (§40).

Finally, an obvious point. An argument for the two principles, or indeed for any conception, is always relative to some list of alternatives. If we change the list, the argument will, in general, have to be different. A similar sort of remark applies to all features of the original position. There are indefinitely many variations of the initial situation and therefore no doubt indefinitely many theorems of moral geometry. Only a few of these are of any philosophical interest, since most variations are irrelevant from a moral point of view. We must try to steer clear of side issues while at the same time not losing sight of the special assumptions of the argument.

22. THE CIRCUMSTANCES OF JUSTICE

The circumstances of justice may be described as the normal conditions under which human cooperation is both possible and necessary.[3] Thus, as I noted at the outset, although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to try to live solely by his own efforts. There is a conflict of interests since men are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share. Thus principles are needed for choosing among the various social arrangements which determine this division of advantages and for underwriting an agreement on the proper distributive shares. These requirements define the role of justice. The background conditions that give rise to these necessities are the circumstances of justice.

These conditions may be divided into two kinds. First, there are the objective circumstances which make human cooperation both possible and necessary. Thus, many individuals coexist together at the same time on a definite geographical territory. These individuals are roughly similar

3. My account largely follows that of Hume in *A Treatise of Human Nature,* bk. III, pt. II, sec. ii, and *An Enquiry Concerning the Principles of Morals,* sec. III, pt. I. But see also H. L. A. Hart, *The Concept of Law* (Oxford, The Clarendon Press, 1961), pp. 189–195, and J. R. Lucas, *The Principles of Politics* (Oxford, The Clarendon Press, 1966), pp. 1–10.

in physical and mental powers; or at any rate, their capacities are comparable in that no one among them can dominate the rest. They are vulnerable to attack, and all are subject to having their plans blocked by the united force of others. Finally, there is the condition of moderate scarcity understood to cover a wide range of situations. Natural and other resources are not so abundant that schemes of cooperation become superfluous, nor are conditions so harsh that fruitful ventures must inevitably break down. While mutually advantageous arrangements are feasible, the benefits they yield fall short of the demands men put forward.

The subjective circumstances are the relevant aspects of the subjects of cooperation, that is, of the persons working together. Thus while the parties have roughly similar needs and interests, or needs and interests in various ways complementary, so that mutually advantageous cooperation among them is possible, they nevertheless have their own plans of life. These plans, or conceptions of the good, lead them to have different ends and purposes, and to make conflicting claims on the natural and social resources available. Moreover, although the interests advanced by these plans are not assumed to be interests in the self, they are the interests of a self that regards its conception of the good as worthy of recognition and that advances claims in its behalf as deserving satisfaction. I also suppose that men suffer from various shortcomings of knowledge, thought, and judgment. Their knowledge is necessarily incomplete, their powers of reasoning, memory, and attention are always limited, and their judgment is likely to be distorted by anxiety, bias, and a preoccupation with their own affairs. Some of these defects spring from moral faults, from selfishness and negligence; but to a large degree, they are simply part of men's natural situation. As a consequence individuals not only have different plans of life but there exists a diversity of philosophical and religious belief, and of political and social doctrines.

Now this constellation of conditions I shall refer to as the circumstances of justice. Hume's account of them is especially perspicuous and the preceding summary adds nothing essential to his much fuller discussion. For simplicity I often stress the condition of moderate scarcity (among the objective circumstances), and that of conflict of interests (among the subjective circumstances). Thus, one can say, in brief, that the circumstances of justice obtain whenever persons put forward conflicting claims to the division of social advantages under conditions of moderate scarcity. Unless these circumstances existed there would be no occasion for the virtue of justice, just as in the absence of threats of injury to life and limb there would be no occasion for physical courage.

Several clarifications should be noted. First of all, I shall, of course, assume that the persons in the original position know that these circumstances of justice obtain. This much they take for granted about the conditions of their society. A further assumption is that the parties try to advance their conception of the good as best they can, and that in attempting to do this they are not bound by prior moral ties to each other.

The question arises, however, whether the persons in the original position have obligations and duties to third parties, for example, to their immediate descendants. To say that they do would be one way of handling questions of justice between generations. However, the aim of justice as fairness is to try to derive all duties and obligations of justice from other reasonable conditions. So, if possible, this way out should be avoided. There are several other courses open to us. We can adopt a motivation assumption and think of the parties as representing a continuing line of claims. For example, we can assume that they are heads of families and therefore have a desire to further the well-being of at least their more immediate descendants. Or we can require the parties to agree to principles subject to the constraint that they wish all preceding generations to have followed the very same principles. By an appropriate combination of such stipulations, I believe that the whole chain of generations can be tied together and principles agreed to that suitably take into account the interests of each (§§24, 44). If this is right, we will have succeeded in deriving duties to other generations from reasonable conditions.

It should be noted that I make no restrictive assumptions about the parties' conceptions of the good except that they are rational long-term plans. While these plans determine the aims and interests of a self, the aims and interests are not presumed to be egoistic or selfish. Whether this is the case depends upon the kinds of ends which a person pursues. If wealth, position, and influence, and the accolades of social prestige, are a person's final purposes, then surely his conception of the good is egoistic. His dominant interests are in himself, not merely, as they must always be, interests of a self.[4] There is no inconsistency, then, in supposing that once the veil of ignorance is removed, the parties find that they have ties of sentiment and affection, and want to advance the interests of others and to see their ends attained. But the postulate of mutual disinterest in the original position is made to insure that the principles of justice do not depend upon strong assumptions. Recall that the original position is meant to incorporate widely shared and yet weak conditions. A conception of

4. On this point see W. T. Stace, *The Concept of Morals* (London, Macmillan, 1937), pp. 221–223.

justice should not presuppose, then, extensive ties of natural sentiment. At the basis of the theory, one tries to assume as little as possible.

Finally, I shall assume that the parties in the original position are mutually disinterested: they are not willing to have their interests sacrificed to the others. The intention is to model men's conduct and motives in cases where questions of justice arise. The spiritual ideals of saints and heroes can be as irreconcilably opposed as any other interests. Conflicts in pursuit of these ideals are the most tragic of all. Thus justice is the virtue of practices where there are competing interests and where persons feel entitled to press their rights on each other. In an association of saints agreeing on a common ideal, if such a community could exist, disputes about justice would not occur. Each would work selflessly for one end as determined by their common religion, and reference to this end (assuming it to be clearly defined) would settle every question of right. But a human society is characterized by the circumstances of justice. The account of these conditions involves no particular theory of human motivation. Rather, its aim is to reflect in the description of the original position the relations of individuals to one another which set the stage for questions of justice.

23. THE FORMAL CONSTRAINTS OF THE CONCEPT OF RIGHT

The situation of the persons in the original position reflects certain constraints. The alternatives open to them and their knowledge of their circumstances are limited in various ways. These restrictions I refer to as the constraints of the concept of right since they hold for the choice of all ethical principles and not only for those of justice. If the parties were to acknowledge principles for the other virtues as well, these constraints would also apply.

I shall consider first the constraints on the alternatives. There are certain formal conditions that it seems reasonable to impose on the conceptions of justice that are to be allowed on the list presented to the parties. I do not claim that these conditions follow from the concept of right, much less from the meaning of morality. I avoid an appeal to the analysis of concepts at crucial points of this kind. There are many constraints that can reasonably be associated with the concept of right, and different selections can be made from these and counted as definitive within a particular theory. The merit of any definition depends upon the

soundness of the theory that results; by itself, a definition cannot settle any fundamental question.[5]

The propriety of these formal conditions is derived from the task of principles of right in adjusting the claims that persons make on their institutions and one another. If the principles of justice are to play their role, that of assigning basic rights and duties and determining the division of advantages, these requirements are natural enough. Each of them is suitably weak and I assume that they are satisfied by the traditional conceptions of justice. These conditions do, however, exclude the various forms of egoism, as I note below, which shows that they are not without moral force. This makes it all the more necessary that the conditions not be justified by definition or the analysis of concepts, but only by the reasonableness of the theory of which they are a part. I arrange them under five familiar headings.

First of all, principles should be general. That is, it must be possible to formulate them without the use of what would be intuitively recognized as proper names, or rigged definite descriptions. Thus the predicates used in their statement should express general properties and relations. Unfortunately deep philosophical difficulties seem to bar the way to a satisfactory account of these matters.[6] I shall not try to deal with them here. In presenting a theory of justice one is entitled to avoid the problem of defining general properties and relations and to be guided by what seems reasonable. Further, since the parties have no specific information about themselves or their situation, they cannot identify themselves anyway. Even if a person could get others to agree, he does not know how to tailor principles to his advantage. The parties are effectively forced to stick to general principles, understanding the notion here in an intuitive fashion.

The naturalness of this condition lies in part in the fact that first

5. Various interpretations of the concept of morality are discussed by W. K. Frankena, "Recent Conceptions of Morality," in *Morality and the Language of Conduct,* ed. H. N. Castañeda and George Nakhnikian (Detroit, Wayne State University Press, 1965), and "The Concept of Morality," *Journal of Philosophy,* vol. 63 (1966). The first of these essays contains numerous references. The account in the text is perhaps closest to that of Kurt Baier in *The Moral Point of View* (Ithaca, N.Y., Cornell University Press, 1958), ch. VIII. I follow Baier in emphasizing the conditions of publicity (he does not use this term, but it is implied by his stipulation of universal teachability, pp. 195f), ordering, finality, and material content (although on the contract view the last condition follows as a consequence, see §25 and note 16 below). For other discussions, see R. M. Hare, *The Language of Morals* (Oxford, The Clarendon Press, 1952), W. D. Falk, "Morality, Self, and Others," also in *Morality and the Language of Conduct,* and P. F. Strawson, "Social Morality and Individual Ideal," *Philosophy,* vol. 36 (1961).

6. See, for example, W. V. Quine, *Ontological Relativity and Other Essays* (New York, Columbia University Press, 1969), ch. 5 entitled "Natural Kinds."

principles must be capable of serving as a public charter of a well-ordered society in perpetuity. Being unconditional, they always hold (under the circumstances of justice), and the knowledge of them must be open to individuals in any generation. Thus, to understand these principles should not require a knowledge of contingent particulars, and surely not a reference to individuals or associations. Traditionally the most obvious test of this condition is the idea that what is right is that which accords with God's will. But in fact this doctrine is normally supported by an argument from general principles. For example, Locke held that the fundamental principle of morals is the following: if one person is created by another (in the theological sense), then that person has a duty to comply with the precepts set to him by his creator.[7] This principle is perfectly general and given the nature of the world on Locke's view, it singles out God as the legitimate moral authority. The generality condition is not violated, although it may appear so at first sight.

Next, principles are to be universal in application. They must hold for everyone in virtue of their being moral persons. Thus I assume that each can understand these principles and use them in his deliberations. This imposes an upper bound of sorts on how complex they can be, and on the kinds and number of distinctions they draw. Moreover, a principle is ruled out if it would be self-contradictory, or self-defeating, for everyone to act upon it. Similarly, should a principle be reasonable to follow only when others conform to a different one, it is also inadmissible. Principles are to be chosen in view of the consequences of everyone's complying with them.

As defined, generality and universality are distinct conditions. For example, egoism in the form of first-person dictatorship (Everyone is to serve my—or Pericles'—interests) satisfies universality but not generality. While all could act in accordance with this principle, and the results might in some cases not be at all bad, depending on the interests of the dictator, the personal pronoun (or the name) violates the first condition. Again, general principles may not be universal. They may be framed to hold for a restricted class of individuals, for instance those singled out by special biological or social characteristics, such as hair color or class situation, or whatever. To be sure, in the course of their lives individuals acquire obligations and assume duties that are peculiar to them. Never-

7. See *Essays on the Laws of Nature,* ed. W. von Leyden (Oxford, The Clarendon Press, 1954), the fourth essay, especially pp. 151–157.

theless these various duties and obligations are the consequence of first principles that hold for all as moral persons; the derivation of these requirements has a common basis.

A third condition is that of publicity, which arises naturally from a contractarian standpoint. The parties assume that they are choosing principles for a public conception of justice.[8] They suppose that everyone will know about these principles all that he would know if their acceptance were the result of an agreement. Thus the general awareness of their universal acceptance should have desirable effects and support the stability of social cooperation. The difference between this condition and that of universality is that the latter leads one to assess principles on the basis of their being intelligently and regularly followed by everyone. But it is possible that all should understand and follow a principle and yet this fact not be widely known or explicitly recognized. The point of the publicity condition is to have the parties evaluate conceptions of justice as publicly acknowledged and fully effective moral constitutions of social life. The publicity condition is clearly implicit in Kant's doctrine of the categorical imperative insofar as it requires us to act in accordance with principles that one would be willing as a rational being to enact as law for a kingdom of ends. He thought of this kingdom as an ethical commonwealth, as it were, which has such moral principles for its public charter.

A further condition is that a conception of right must impose an ordering on conflicting claims. This requirement springs directly from the role of its principles in adjusting competing demands. There is a difficulty, however, in deciding what counts as an ordering. It is clearly desirable that a conception of justice be complete, that is, able to order all the claims that can arise (or that are likely to in practice). And the ordering

8. Publicity is clearly implied in Kant's notion of the moral law, but the only place I know of where he discusses it expressly is in *Perpetual Peace,* appendix II; see *Political Writings,* ed. Hans Reiss and trans. H. B. Nisbet (Cambridge, The University Press, 1970), pp. 125–130. There are of course brief statements elsewhere. For example, in *The Metaphysics of Morals,* pt. I *(Rechtslehre),* §43, he says: "Public Right is the sum total of those laws which require to be made universally public in order to produce a state of right." In "Theory and Practice" he remarks in a footnote: "No right in a state can be tacitly and treacherously included by a secret reservation, and least of all a right which the people claim to be a part of the constitution, for all laws within it must be thought of as arising out of a public will. Thus if a constitution allowed rebellion, it would have to declare this right publicly and make clear how it might be implemented." *Political Writings,* pp. 136, 84n, respectively. I believe Kant intends this condition to apply to a society's conception of justice. See also note 4, §51, below; and Baier, cited in note 5 above. There is a discussion of common knowledge and its relation to agreement in D. K. Lewis, *Convention* (Cambridge, Mass., Harvard University Press, 1969), esp. pp. 52–60, 83–88.

should in general be transitive: if, say, a first arrangement of the basic structure is ranked more just than a second, and the second more just than a third, then the first should be more just than the third. These formal conditions are natural enough, though not always easy to satisfy.[9] But is trial by combat a form of adjudication? After all, physical conflict and resort to arms result in an ordering; certain claims do win out over others. The main objection to this ordering is not that it may be intransitive. Rather, it is to avoid the appeal to force and cunning that the principles of right and justice are accepted. Thus I assume that to each according to his threat advantage is not a conception of justice. It fails to establish an ordering in the required sense, an ordering based on certain relevant aspects of persons and their situation which are independent from their social position, or their capacity to intimidate and coerce.[10]

The fifth and last condition is that of finality. The parties are to assess the system of principles as the final court of appeal in practical reasoning. There are no higher standards to which arguments in support of claims can be addressed; reasoning successfully from these principles is conclusive. If we think in terms of the fully general theory which has principles for all the virtues, then such a theory specifies the totality of relevant considerations and their appropriate weights, and its requirements are decisive. They override the demands of law and custom, and of social rules generally. We are to arrange and respect social institutions as the princi-

9. For a discussion of orderings and preference relations, see A. K. Sen, *Collective Choice and Social Welfare* (San Francisco, Holden-Day Inc., 1970), chs. 1 and 1*; and K. J. Arrow, *Social Choice and Individual Values,* 2nd ed. (New York, John Wiley, 1963), ch. II.

10. *Theory of Games as a Tool for the Moral Philosopher* (Cambridge, The University Press, 1955). On the analysis he presents, it turns out that the fair division of playing time between Matthew and Luke depends on their preferences, and these in turn are connected with the instruments they wish to play. Since Matthew has a threat advantage over Luke, arising from the fact that Matthew, the trumpeter, prefers both of them playing at once to neither of them playing, whereas Luke, the pianist, prefers silence to cacophony, Matthew is allotted twenty-six evenings of play to Luke's seventeen. If the situation were reversed, the threat advantage would be with Luke. See pp. 36f. But we have only to suppose that Matthew is a jazz enthusiast who plays the drums, and Luke a violinist who plays sonatas, in which case it will be fair on this analysis for Matthew to play whenever and as often as he likes, assuming as it is plausible to assume that he does not care whether Luke plays or not. Clearly something has gone wrong. What is lacking is a suitable definition of a status quo that is acceptable from a moral point of view. We cannot take various contingencies as known and individual preferences as given and expect to elucidate the concept of justice (or fairness) by theories of bargaining. The conception of the original position is designed to meet the problem of the appropriate status quo. A similar objection to Braithwaite's analysis is found in J. R. Lucas, "Moralists and Gamesmen," *Philosophy,* vol. 34 (1959), pp. 9f. For another discussion, consult Sen, *Collective Choice and Social Welfare,* pp. 118–123, who argues that the solution of J. F. Nash in "The Bargaining Problem," *Econometrica,* vol. 18 (1950), is similarly defective from an ethical point of view.

ples of right and justice direct. Conclusions from these principles also override considerations of prudence and self-interest. This does not mean that these principles insist upon self-sacrifice; for in drawing up the conception of right the parties take their interests into account as best they can. The claims of personal prudence are already given an appropriate weight within the full system of principles. The complete scheme is final in that when the course of practical reasoning it defines has reached its conclusion, the question is settled. The claims of existing social arrangements and of self-interest have been duly allowed for. We cannot at the end count them a second time because we do not like the result.

Taken together, then, these conditions on conceptions of right come to this: a conception of right is a set of principles, general in form and universal in application, that is to be publicly recognized as a final court of appeal for ordering the conflicting claims of moral persons. Principles of justice are identified by their special role and the subject to which they apply. Now by themselves the five conditions exclude none of the traditional conceptions of justice. It should be noted, however, that they do rule out the listed variants of egoism. The generality condition eliminates both first-person dictatorship and the free-rider forms, since in each case a proper name, or pronoun, or a rigged definite description is needed, either to single out the dictator or to characterize the free-rider. Generality does not, however, exclude general egoism, for each person is allowed to do whatever, in his judgment, is most likely to further his own aims. The principle here can clearly be expressed in a perfectly general way. It is the ordering condition which renders general egoism inadmissible, for if everyone is authorized to advance his aims as he pleases, or if everyone ought to advance his own interests, competing claims are not ranked at all and the outcome is determined by force and cunning.

The several kinds of egoism, then, do not appear on the list presented to the parties. They are eliminated by the formal constraints. Of course, this is not a surprising conclusion, since it is obvious that by choosing one of the other conceptions the persons in the original position can do much better for themselves. Once they ask which principles all should agree to, no form of egoism is a serious candidate for consideration in any case. This only confirms what we knew already, namely, that although egoism is logically consistent and in this sense not irrational, it is incompatible with what we intuitively regard as the moral point of view. The significance of egoism philosophically is not as an alternative conception of right but as a challenge to any such conception. In justice as fairness this

is reflected in the fact that we can interpret general egoism as the no-agreement point. It is what the parties would be stuck with if they were unable to reach an understanding.

24. THE VEIL OF IGNORANCE

The idea of the original position is to set up a fair procedure so that any principles agreed to will be just. The aim is to use the notion of pure procedural justice as a basis of theory. Somehow we must nullify the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage. Now in order to do this I assume that the parties are situated behind a veil of ignorance. They do not know how the various alternatives will affect their own particular case and they are obliged to evaluate principles solely on the basis of general considerations.[11]

It is assumed, then, that the parties do not know certain kinds of particular facts. First of all, no one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like. Nor, again, does anyone know his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism. More than this, I assume that the parties do not know the particular circumstances of their own society. That is, they do not know its economic or political situation, or the level of civilization and culture it has been able to achieve. The persons in the original position have no information as to which generation they belong. These broader restrictions on knowledge are appropriate in part because questions of social justice arise between generations as well as within them, for example, the question of the appropriate rate of capital saving and of the conservation of natural re-

11. The veil of ignorance is so natural a condition that something like it must have occurred to many. The formulation in the text is implicit, I believe, in Kant's doctrine of the categorical imperative, both in the way this procedural criterion is defined and the use Kant makes of it. Thus when Kant tells us to test our maxim by considering what would be the case were it a universal law of nature, he must suppose that we do not know our place within this imagined system of nature. See, for example, his discussion of the topic of practical judgment in *The Critique of Practical Reason,* Academy Edition, vol. 5, pp. 68–72. A similar restriction on information is found in J. C. Harsanyi, "Cardinal Utility in Welfare Economics and in the Theory of Risk-taking," *Journal of Political Economy,* vol. 61 (1953). However, other aspects of Harsanyi's view are quite different, and he uses the restriction to develop a utilitarian theory. See the last paragraph of §27.

sources and the environment of nature. There is also, theoretically anyway, the question of a reasonable genetic policy. In these cases too, in order to carry through the idea of the original position, the parties must not know the contingencies that set them in opposition. They must choose principles the consequences of which they are prepared to live with whatever generation they turn out to belong to.

As far as possible, then, the only particular facts which the parties know is that their society is subject to the circumstances of justice and whatever this implies. It is taken for granted, however, that they know the general facts about human society. They understand political affairs and the principles of economic theory; they know the basis of social organization and the laws of human psychology. Indeed, the parties are presumed to know whatever general facts affect the choice of the principles of justice. There are no limitations on general information, that is, on general laws and theories, since conceptions of justice must be adjusted to the characteristics of the systems of social cooperation which they are to regulate, and there is no reason to rule out these facts. It is, for example, a consideration against a conception of justice that, in view of the laws of moral psychology, men would not acquire a desire to act upon it even when the institutions of their society satisfied it. For in this case there would be difficulty in securing the stability of social cooperation. An important feature of a conception of justice is that it should generate its own support. Its principles should be such that when they are embodied in the basic structure of society men tend to acquire the corresponding sense of justice and develop a desire to act in accordance with its principles. In this case a conception of justice is stable. This kind of general information is admissible in the original position.

The notion of the veil of ignorance raises several difficulties. Some may object that the exclusion of nearly all particular information makes it difficult to grasp what is meant by the original position. Thus it may be helpful to observe that one or more persons can at any time enter this position, or perhaps better, simulate the deliberations of this hypothetical situation, simply by reasoning in accordance with the appropriate restrictions. In arguing for a conception of justice we must be sure that it is among the permitted alternatives and satisfies the stipulated formal constraints. No considerations can be advanced in its favor unless they would be rational ones for us to urge were we to lack the kind of knowledge that is excluded. The evaluation of principles must proceed in terms of the general consequences of their public recognition and universal application, it being assumed that they will be complied with by everyone. To say

that a certain conception of justice would be chosen in the original position is equivalent to saying that rational deliberation satisfying certain conditions and restrictions would reach a certain conclusion. If necessary, the argument to this result could be set out more formally. I shall, however, speak throughout in terms of the notion of the original position. It is more economical and suggestive, and brings out certain essential features that otherwise one might easily overlook.

These remarks show that the original position is not to be thought of as a general assembly which includes at one moment everyone who will live at some time; or, much less, as an assembly of everyone who could live at some time. It is not a gathering of all actual or possible persons. If we conceived of the original position in either of these ways, the conception would cease to be a natural guide to intuition and would lack a clear sense. In any case, the original position must be interpreted so that one can at any time adopt its perspective. It must make no difference when one takes up this viewpoint, or who does so: the restrictions must be such that the same principles are always chosen. The veil of ignorance is a key condition in meeting this requirement. It insures not only that the information available is relevant, but that it is at all times the same.

It may be protested that the condition of the veil of ignorance is irrational. Surely, some may object, principles should be chosen in the light of all the knowledge available. There are various replies to this contention. Here I shall sketch those which emphasize the simplifications that need to be made if one is to have any theory at all. (Those based on the Kantian interpretation of the original position are given later, §40.) To begin with, it is clear that since the differences among the parties are unknown to them, and everyone is equally rational and similarly situated, each is convinced by the same arguments. Therefore, we can view the agreement in the original position from the standpoint of one person selected at random. If anyone after due reflection prefers a conception of justice to another, then they all do, and a unanimous agreement can be reached. We can, to make the circumstances more vivid, imagine that the parties are required to communicate with each other through a referee as intermediary, and that he is to announce which alternatives have been suggested and the reasons offered in their support. He forbids the attempt to form coalitions, and he informs the parties when they have come to an understanding. But such a referee is actually superfluous, assuming that the deliberations of the parties must be similar.

Thus there follows the very important consequence that the parties have no basis for bargaining in the usual sense. No one knows his situ-

ation in society nor his natural assets, and therefore no one is in a position to tailor principles to his advantage. We might imagine that one of the contractees threatens to hold out unless the others agree to principles favorable to him. But how does he know which principles are especially in his interests? The same holds for the formation of coalitions: if a group were to decide to band together to the disadvantage of the others, they would not know how to favor themselves in the choice of principles. Even if they could get everyone to agree to their proposal, they would have no assurance that it was to their advantage, since they cannot identify themselves either by name or description. The one case where this conclusion fails is that of saving. Since the persons in the original position know that they are contemporaries (taking the present time of entry interpretation), they can favor their generation by refusing to make any sacrifices at all for their successors; they simply acknowledge the principle that no one has a duty to save for posterity. Previous generations have saved or they have not; there is nothing the parties can now do to affect that. So in this instance the veil of ignorance fails to secure the desired result. Therefore, to handle the question of justice between generations, I modify the motivation assumption and add a further constraint (§22). With these adjustments, no generation is able to formulate principles especially designed to advance its own cause and some significant limits on savings principles can be derived (§44). Whatever a person's temporal position, each is forced to choose for all.[12]

The restrictions on particular information in the original position are, then, of fundamental importance. Without them we would not be able to work out any definite theory of justice at all. We would have to be content with a vague formula stating that justice is what would be agreed to without being able to say much, if anything, about the substance of the agreement itself. The formal constraints of the concept of right, those applying to principles directly, are not sufficient for our purpose. The veil of ignorance makes possible a unanimous choice of a particular conception of justice. Without these limitations on knowledge the bargaining problem of the original position would be hopelessly complicated. Even if theoretically a solution were to exist, we would not, at present anyway, be able to determine it.

The notion of the veil of ignorance is implicit, I think, in Kant's ethics (§40). Nevertheless the problem of defining the knowledge of the parties and of characterizing the alternatives open to them has often been passed

12. Rousseau, *The Social Contract,* bk. II, ch. IV, par. 5.

over, even by contract theories. Sometimes the situation definitive of moral deliberation is presented in such an indeterminate way that one cannot ascertain how it will turn out. Thus Perry's doctrine is essentially contractarian: he holds that social and personal integration must proceed by entirely different principles, the latter by rational prudence, the former by the concurrence of persons of good will. He would appear to reject utilitarianism on much the same grounds suggested earlier: namely, that it improperly extends the principle of choice for one person to choices facing society. The right course of action is characterized as that which best advances social aims as these would be formulated by reflective agreement, given that the parties have full knowledge of the circumstances and are moved by a benevolent concern for one another's interests. No effort is made, however, to specify in any precise way the possible outcomes of this sort of agreement. Indeed, without a far more elaborate account, no conclusions can be drawn.[13] I do not wish here to criticize others; rather, I want to explain the necessity for what may seem at times like so many irrelevant details.

Now the reasons for the veil of ignorance go beyond mere simplicity. We want to define the original position so that we get the desired solution. If a knowledge of particulars is allowed, then the outcome is biased by arbitrary contingencies. As already observed, to each according to his threat advantage is not a principle of justice. If the original position is to yield agreements that are just, the parties must be fairly situated and treated equally as moral persons. The arbitrariness of the world must be corrected for by adjusting the circumstances of the initial contractual situation. Moreover, if in choosing principles we required unanimity even when there is full information, only a few rather obvious cases could be decided. A conception of justice based on unanimity in these circumstances would indeed be weak and trivial. But once knowledge is excluded, the requirement of unanimity is not out of place and the fact that it can be satisfied is of great importance. It enables us to say of the preferred conception of justice that it represents a genuine reconciliation of interests.

A final comment. For the most part I shall suppose that the parties possess all general information. No general facts are closed to them. I do this mainly to avoid complications. Nevertheless a conception of justice is to be the public basis of the terms of social cooperation. Since common

13. See R. B. Perry, *The General Theory of Value* (New York, Longmans, Green and Company, 1926), pp. 674–682.

understanding necessitates certain bounds on the complexity of principles, there may likewise be limits on the use of theoretical knowledge in the original position. Now clearly it would be very difficult to classify and to grade the complexity of the various sorts of general facts. I shall make no attempt to do this. We do however recognize an intricate theoretical construction when we meet one. Thus it seems reasonable to say that other things equal one conception of justice is to be preferred to another when it is founded upon markedly simpler general facts, and its choice does not depend upon elaborate calculations in the light of a vast array of theoretically defined possibilities. It is desirable that the grounds for a public conception of justice should be evident to everyone when circumstances permit. This consideration favors, I believe, the two principles of justice over the criterion of utility.

25. THE RATIONALITY OF THE PARTIES

I have assumed throughout that the persons in the original position are rational. But I have also assumed that they do not know their conception of the good. This means that while they know that they have some rational plan of life, they do not know the details of this plan, the particular ends and interests which it is calculated to promote. How, then, can they decide which conceptions of justice are most to their advantage? Or must we suppose that they are reduced to mere guessing? To meet this difficulty, I postulate that they accept the account of the good touched upon in the preceding chapter: they assume that they normally prefer more primary social goods rather than less. Of course, it may turn out, once the veil of ignorance is removed, that some of them for religious or other reasons may not, in fact, want more of these goods. But from the standpoint of the original position, it is rational for the parties to suppose that they do want a larger share, since in any case they are not compelled to accept more if they do not wish to. Thus even though the parties are deprived of information about their particular ends, they have enough knowledge to rank the alternatives. They know that in general they must try to protect their liberties, widen their opportunities, and enlarge their means for promoting their aims whatever these are. Guided by the theory of the good and the general facts of moral psychology, their deliberations are no longer guesswork. They can make a rational decision in the ordinary sense.

The concept of rationality invoked here, with the exception of one

essential feature, is the standard one familiar in social theory.[14] Thus in the usual way, a rational person is thought to have a coherent set of preferences between the options open to him. He ranks these options according to how well they further his purposes; he follows the plan which will satisfy more of his desires rather than less, and which has the greater chance of being successfully executed. The special assumption I make is that a rational individual does not suffer from envy. He is not ready to accept a loss for himself if only others have less as well. He is not downcast by the knowledge or perception that others have a larger index of primary social goods. Or at least this is true as long as the differences between himself and others do not exceed certain limits, and he does not believe that the existing inequalities are founded on injustice or are the result of letting chance work itself out for no compensating social purpose (§80).

The assumption that the parties are not moved by envy raises certain questions. Perhaps we should also assume that they are not liable to various other feelings such as shame and humiliation (§67). Now a satisfactory account of justice will eventually have to deal with these matters too, but for the present I shall leave these complications aside. Another objection to our procedure is that it is too unrealistic. Certainly men are afflicted with these feelings. How can a conception of justice ignore this fact? I shall meet this problem by dividing the argument for the principles of justice into two parts. In the first part, the principles are derived on the supposition that envy does not exist; while in the second, we consider whether the conception arrived at is feasible in view of the circumstances of human life.

One reason for this procedure is that envy tends to make everyone worse off. In this sense it is collectively disadvantageous. Presuming its absence amounts to supposing that in the choice of principles men should think of themselves as having their own plan of life which is sufficient for

14. For this notion of rationality, see the references to Sen and Arrow above, §23, note 9. The discussion in I. M. D. Little, *The Critique of Welfare Economics,* 2nd ed. (Oxford, Clarendon Press, 1957), ch. II, is also relevant here. For rational choice under uncertainty, see below §26, note 18. H. A. Simon discusses the limitations of the classical conceptions of rationality and the need for a more realistic theory in "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics,* vol. 69 (1955). See also his essay in *Surveys of Economic Theory,* vol. 3 (London, Macmillan, 1967). For philosophical discussions see Donald Davidson, "Actions, Reasons, and Causes," *Journal of Philosophy,* vol. 60 (1963); C. G. Hempel, *Aspects of Scientific Explanation* (New York, The Free Press, 1965), pp. 463–486; Jonathan Bennett, *Rationality* (London, Routledge and Kegan Paul, 1964), and J. D. Mabbott, "Reason and Desire," *Philosophy,* vol. 28 (1953).

itself. They have a secure sense of their own worth so that they have no desire to abandon any of their aims provided others have less means to further theirs. I shall work out a conception of justice on this stipulation to see what happens. Later I shall try to show that when the principles adopted are put into practice, they lead to social arrangements in which envy and other destructive feelings are not likely to be strong. The conception of justice eliminates the conditions that give rise to disruptive attitudes. It is, therefore, inherently stable (§§80–81).

The assumption of mutually disinterested rationality, then, comes to this: the persons in the original position try to acknowledge principles which advance their system of ends as far as possible. They do this by attempting to win for themselves the highest index of primary social goods, since this enables them to promote their conception of the good most effectively whatever it turns out to be. The parties do not seek to confer benefits or to impose injuries on one another; they are not moved by affection or rancor. Nor do they try to gain relative to each other; they are not envious or vain. Put in terms of a game, we might say: they strive for as high an absolute score as possible. They do not wish a high or a low score for their opponents, nor do they seek to maximize or minimize the difference between their successes and those of others. The idea of a game does not really apply, since the parties are not concerned to win but to get as many points as possible judged by their own system of ends.

There is one further assumption to guarantee strict compliance. The parties are presumed to be capable of a sense of justice and this fact is public knowledge among them. This condition is to insure the integrity of the agreement made in the original position. It does not mean that in their deliberations the parties apply some particular conception of justice, for this would defeat the point of the motivation assumption. Rather, it means that the parties can rely on each other to understand and to act in accordance with whatever principles are finally agreed to. Once principles are acknowledged the parties can depend on one another to conform to them. In reaching an agreement, then, they know that their undertaking is not in vain: their capacity for a sense of justice insures that the principles chosen will be respected. It is essential to observe, however, that this assumption still permits the consideration of men's capacity to act on the various conceptions of justice. The general facts of human psychology and the principles of moral learning are relevant matters for the parties to examine. If a conception of justice is unlikely to generate its own support, or lacks stability, this fact must not be overlooked. For then a different

conception of justice might be preferred. The assumption only says that the parties have a capacity for justice in a purely formal sense: taking everything relevant into account, including the general facts of moral psychology, the parties will adhere to the principles eventually chosen. They are rational in that they will not enter into agreements they know they cannot keep, or can do so only with great difficulty. Along with other considerations, they count the strains of commitment (§29). Thus in assessing conceptions of justice the persons in the original position are to assume that the one they adopt will be strictly complied with. The consequences of their agreement are to be worked out on this basis.

With the preceding remarks about rationality and motivation of the parties the description of the original position is for the most part complete. We can summarize this description with the following list of elements of the initial situation and their variations. (The asterisks mark the interpretations that constitute the original position.)

1. The Nature of the Parties (§22)
   - *a. continuing persons (family heads, or genetic lines)
   - b. single individuals
   - c. associations (states, churches, or other corporate bodies)
2. Subject of Justice (§2)
   - *a. basic structure of society
   - b. rules of corporate associations
   - c. law of nations
3. Presentation of Alternatives (§21)
   - *a. shorter (or longer) list
   - b. general characterization of the possibilities
4. Time of Entry (§24)
   - *a. any time (during age of reason) for living persons
   - b. all actual persons (those alive at some time) simultaneously
   - c. all possible persons simultaneously
5. Circumstances of Justice (§22)
   - *a. Hume's conditions of moderate scarcity
   - b. the above plus further extremes
6. Formal Conditions on Principles (§23)
   - *a. generality, universality, publicity, ordering, and finality
   - b. the above less publicity, say
7. Knowledge and Beliefs (§24)
   - *a. veil of ignorance
   - b. full information
   - c. partial knowledge

8. Motivation of the Parties (§25)
   *a. mutual disinterestedness (limited altruism)
   b. elements of social solidarity and good will
   c. perfect altruism
9. Rationality (§§25, 28)
   *a. taking effective means to ends with unified expectations and objective interpretation of probability
   b. as above but without unified expectations and using the principle of insufficient reason
10. Agreement Condition (§24)
   *a. unanimity in perpetuity
   b. majority acceptance, or whatever, for limited period
11. Compliance Condition (§25)
   *a. strict compliance
   b. partial compliance in various degrees
12. No Agreement Point (§23)
   *a. general egoism
   b. the state of nature

We can turn now to the choice of principles. But first I shall mention a few misunderstandings to be avoided. First of all, we must keep in mind that the parties in the original position are theoretically defined individuals. The grounds for their consent are set out by the description of the contractual situation and their preference for primary goods. Thus to say that the principles of justice would be adopted is to say how these persons would decide being moved in ways our account describes. Of course, when we try to simulate the original position in everyday life, that is, when we try to conduct ourselves in moral argument as its constraints require, we will presumably find that our deliberations and judgments are influenced by our special inclinations and attitudes. Surely it will prove difficult to correct for our various propensities and aversions in striving to adhere to the conditions of this idealized situation. But none of this affects the contention that in the original position rational persons so characterized would make a certain decision. This proposition belongs to the theory of justice. It is another question how well human beings can assume this role in regulating their practical reasoning.

Since the persons in the original position are assumed to take no interest in one another's interests (although they may have a concern for third parties), it may be thought that justice as fairness is itself an egoistic theory. It is not, of course, one of the three forms of egoism mentioned earlier, but some may think, as Schopenhauer thought of Kant's doctrine,

that it is egoistic nevertheless.[15] Now this is a misconception. For the fact that in the original position the parties are characterized as mutually disinterested does not entail that persons in ordinary life, or in a well-ordered society, who hold the principles that would be agreed to are similarly disinterested in one another. Clearly the two principles of justice and the principles of obligation and natural duty require us to consider the rights and claims of others. And the sense of justice is a normally effective desire to comply with these restrictions. The motivation of the persons in the original position must not be confused with the motivation of persons in everyday life who accept the principles of justice and who have the corresponding sense of justice. In practical affairs an individual does have a knowledge of his situation and he can, if he wishes, exploit contingencies to his advantage. Should his sense of justice move him to act on the principles of right that would be adopted in the original position, his desires and aims are surely not egoistic. He voluntarily takes on the limitations expressed by this interpretation of the moral point of view. Thus, more generally, the motivation of the parties in the original position does not determine directly the motivation of people in a just society. For in the latter case, we assume that its members grow up and live under a just basic structure, as the two principles require; and then we try to work out what kind of conception of the good and moral sentiments people would acquire (Ch. VIII). Therefore the mutual disinterestedness of the parties determines other motivations only indirectly, that is, via its effects on the agreement on principles. It is these principles, together with the laws of psychology (as these work under the conditions of just institutions), which shape the aims and moral sentiments of citizens of a well-ordered society.

Once we consider the idea of a contract theory it is tempting to think that it will not yield the principles we want unless the parties are to some degree at least moved by benevolence, or an interest in one another's interests. Perry, as I mentioned before, thinks of the right standards and decisions as those promoting the ends reached by reflective agreement under circumstances making for impartiality and good will. Now the combination of mutual disinterest and the veil of ignorance achieves much the same purpose as benevolence. For this combination of conditions forces each person in the original position to take the good of others

15. See *On the Basis of Ethics* (1840), trans. E. F. J. Payne (New York, The Liberal Arts Press, Inc., 1965), pp. 89–92.

into account. In justice as fairness, then, the effects of good will are brought about by several conditions working jointly. The feeling that this conception of justice is egoistic is an illusion fostered by looking at but one of the elements of the original position. Furthermore, this pair of assumptions has enormous advantages over that of benevolence plus knowledge. As I have noted, the latter is so complex that no definite theory at all can be worked out. Not only are the complications caused by so much information insurmountable, but the motivational assumption requires clarification. For example, what is the relative strength of benevolent desires? In brief, the combination of mutual disinterestedness plus the veil of ignorance has the merits of simplicity and clarity while at the same time insuring the effects of what are at first sight morally more attractive assumptions.

Finally, if the parties are conceived as themselves making proposals, they have no incentive to suggest pointless or arbitrary principles. For example, none would urge that special privileges be given to those exactly six feet tall or born on a sunny day. Nor would anyone put forward the principle that basic rights should depend on the color of one's skin or the texture of one's hair. No one can tell whether such principles would be to his advantage. Furthermore, each such principle is a limitation of one's liberty of action, and such restrictions are not to be accepted without a reason. Certainly we might imagine peculiar circumstances in which these characteristics are relevant. Those born on a sunny day might be blessed with a happy temperament, and for some positions of authority this might be a qualifying attribute. But such distinctions would never be proposed in first principles, for these must have some rational connection with the advancement of human interests broadly defined. The rationality of the parties and their situation in the original position guarantees that ethical principles and conceptions of justice have this general content.[16] Inevitably, then, racial and sexual discrimination presupposes that some hold a favored place in the social system which they are willing to exploit to their advantage. From the standpoint of persons similarly situated in an initial situation which is fair, the principles of explicit racist doctrines are not only unjust. They are irrational. For this reason we could say that they

16. For a different way of reaching this conclusion, see Philippa Foot, "Moral Arguments," *Mind,* vol. 67 (1958), and "Moral Beliefs," *Proceedings of the Aristotelian Society,* vol. 59 (1958–1959); and R. W. Beardsmore, *Moral Reasoning* (New York, Schocken Books, 1969), especially ch. IV. The problem of content is discussed briefly in G. F. Warnock, *Contemporary Moral Philosophy* (London, Macmillan, 1967), pp. 55–61.

are not moral conceptions at all, but simply means of suppression. They have no place on a reasonable list of traditional conceptions of justice.[17] Of course, this contention is not at all a matter of definition. It is rather a consequence of the conditions characterizing the original position, especially the conditions of the rationality of the parties and the veil of ignorance. That conceptions of right have a certain content and exclude arbitrary and pointless principles is, therefore, an inference from the theory.

26. THE REASONING LEADING TO THE TWO PRINCIPLES OF JUSTICE

In this and the next two sections I take up the choice between the two principles of justice and the principle of average utility. Determining the rational preference between these two options is perhaps the central problem in developing the conception of justice as fairness as a viable alternative to the utilitarian tradition. I shall begin in this section by presenting some intuitive remarks favoring the two principles. I shall also discuss briefly the qualitative structure of the argument that needs to be made if the case for these principles is to be conclusive.

Now consider the point of view of anyone in the original position. There is no way for him to win special advantages for himself. Nor, on the other hand, are there grounds for his acquiescing in special disadvantages. Since it is not reasonable for him to expect more than an equal share in the division of social primary goods, and since it is not rational for him to agree to less, the sensible thing is to acknowledge as the first step a principle of justice requiring an equal distribution. Indeed, this principle is so obvious given the symmetry of the parties that it would occur to everyone immediately. Thus the parties start with a principle requiring equal basic liberties for all, as well as fair equality of opportunity and equal division of income and wealth.

But even holding firm to the priority of the basic liberties and fair equality of opportunity, there is no reason why this initial acknowledgment should be final. Society should take into account economic efficiency and the requirements of organization and technology. If there are inequalities in income and wealth, and differences in authority and de-

17. For a similar view, see B. A. O. Williams, "The Idea of Equality," *Philosophy, Politics, and Society,* Second Series, ed. Peter Laslett and W. G. Runciman (Oxford, Basil Blackwell, 1962), p. 113.

grees of responsibility, that work to make everyone better off in comparison with the benchmark of equality, why not permit them? One might think that ideally individuals should want to serve one another. But since the parties are assumed to be mutually disinterested, their acceptance of these economic and institutional inequalities is only the recognition of the relations of opposition in which men stand in the circumstances of justice. They have no grounds for complaining of one another's motives. Thus the parties would object to these differences only if they would be dejected by the bare knowledge or perception that others are better situated; but I suppose that they decide as if they are not moved by envy. Thus the basic structure should allow these inequalities so long as these improve everyone's situation, including that of the least advantaged, provided that they are consistent with equal liberty and fair opportunity. Because the parties start from an equal division of all social primary goods, those who benefit least have, so to speak, a veto. Thus we arrive at the difference principle. Taking equality as the basis of comparison, those who have gained more must do so on terms that are justifiable to those who have gained the least.

By some such reasoning, then, the parties might arrive at the two principles of justice in serial order. I shall not try to justify this ordering here, but the following remarks may convey the intuitive idea. I assume that the parties view themselves as free persons who have fundamental aims and interests in the name of which they think it legitimate for them to make claims on one another concerning the design of the basic structure of society. The religious interest is a familiar historical example; the interest in the integrity of the person is another. In the original position the parties do not know what particular forms these interests take; but they do assume that they have such interests and that the basic liberties necessary for their protection are guaranteed by the first principle. Since they must secure these interests, they rank the first principle prior to the second. The case for the two principles can be strengthened by spelling out in more detail the notion of a free person. Very roughly the parties regard themselves as having a highest-order interest in how all their other interests, including even their fundamental ones, are shaped and regulated by social institutions. They do not think of themselves as inevitably bound to, or as identical with, the pursuit of any particular complex of fundamental interests that they may have at any given time, although they want the right to advance such interests (provided they are admissible). Rather, free persons conceive of themselves as beings who can revise and alter their final ends and who give first priority to preserving their liberty

in these matters. Hence, they not only have final ends that they are in principle free to pursue or to reject, but their original allegiance and continued devotion to these ends are to be formed and affirmed under conditions that are free. Since the two principles secure a social form that maintains these conditions, they would be agreed to rather than the principle of utility. Only by this agreement can the parties be sure that their highest-order interest as free persons is guaranteed.

The priority of liberty means that whenever the basic liberties can be effectively established, a lesser or an unequal liberty cannot be exchanged for an improvement in economic well-being. It is only when social circumstances do not allow the effective establishment of these basic rights that one can concede their limitation; and even then these restrictions can be granted only to the extent that they are necessary to prepare the way for the time when they are no longer justified. The denial of the equal liberties can be defended only when it is essential to change the conditions of civilization so that in due course these liberties can be enjoyed. Thus in adopting the serial order of the two principles, the parties are assuming that the conditions of their society, whatever they are, admit the effective realization of the equal liberties. Or that if they do not, circumstances are nevertheless sufficiently favorable so that the priority of the first principle points out the most urgent changes and identifies the preferred path to the social state in which all the basic liberties can be fully instituted. The complete realization of the two principles in serial order is the long-run tendency of this ordering, at least under reasonably fortunate conditions.

It seems from these remarks that the two principles are at least a plausible conception of justice. The question, though, is how one is to argue for them more systematically. Now there are several things to do. One can work out their consequences for institutions and note their implications for fundamental social policy. In this way they are tested by a comparison with our considered judgments of justice. Part II is devoted to this. But one can also try to find arguments in their favor that are decisive from the standpoint of the original position. In order to see how this might be done, it is useful as a heuristic device to think of the two principles as the maximin solution to the problem of social justice. There is a relation between the two principles and the maximin rule for choice under uncertainty.[18] This is evident from the fact that the two principles

18. An accessible discussion of this and other rules of choice under uncertainty can be found in W. J. Baumol, *Economic Theory and Operations Analysis,* 2nd ed. (Englewood Cliffs, N.J., Prentice-

are those a person would choose for the design of a society in which his enemy is to assign him his place. The maximin rule tells us to rank alternatives by their worst possible outcomes: we are to adopt the alternative the worst outcome of which is superior to the worst outcomes of the others.[19] The persons in the original position do not, of course, assume that their initial place in society is decided by a malevolent opponent. As I note below, they should not reason from false premises. The veil of ignorance does not violate this idea, since an absence of information is not misinformation. But that the two principles of justice would be chosen if the parties were forced to protect themselves against such a contingency explains the sense in which this conception is the maximin solution. And this analogy suggests that if the original position has been described so that it is rational for the parties to adopt the conservative attitude expressed by this rule, a conclusive argument can indeed be constructed for these principles. Clearly the maximin rule is not, in general, a suitable guide for choices under uncertainty. But it holds only in situations marked by certain special features. My aim, then, is to show that a good case can be made for the two principles based on the fact that the original position has these features to a very high degree.

Hall Inc., 1965), ch. 24. Baumol gives a geometric interpretation of these rules, including the diagram used in §13 to illustrate the difference principle. See pp. 558–562. See also R. D. Luce and Howard Raiffa, *Games and Decisions* (New York, John Wiley and Sons, Inc., 1957), ch. XIII, for a fuller account.

19. Consider the gain-and-loss table below. It represents the gains and losses for a situation which is not a game of strategy. There is no one playing against the person making the decision; instead he is faced with several possible circumstances which may or may not obtain. Which circumstances happen to exist does not depend upon what the person choosing decides or whether he announces his moves in advance. The numbers in the table are monetary values (in hundreds of dollars) in comparison with some initial situation. The gain (g) depends upon the individual's decision (d) and the circumstances (c). Thus $g = f(d, c)$. Assuming that there are three possible decisions and three possible circumstances, we might have this gain-and-loss table.

	Circumstances		
Decisions	c_1	c_2	c_3
d_1	–7	8	12
d_2	–8	7	14
d_3	5	6	8

The maximin rule requires that we make the third decision. For in this case the worst that can happen is that one gains five hundred dollars, which is better than the worst for the other actions. If we adopt one of these we may lose either eight or seven hundred dollars. Thus, the choice of d_3 maximizes f (d,c) for that value of c, which for a given d, minimizes f. The term "maximin" means the *maximum minimorum;* and the rule directs our attention to the worst that can happen under any proposed course of action, and to decide in the light of that.

Now there appear to be three chief features of situations that give plausibility to this unusual rule.[20] First, since the rule takes no account of the likelihoods of the possible circumstances, there must be some reason for sharply discounting estimates of these probabilities. Offhand, the most natural rule of choice would seem to be to compute the expectation of monetary gain for each decision and then to adopt the course of action with the highest prospect. (This expectation is defined as follows: let us suppose that g_{ij} represent the numbers in the gain-and-loss table, where i is the row index and j is the column index; and let p_j, $j = 1, 2, 3$, be the likelihoods of the circumstances, with $\Sigma p_j = 1$. Then the expectation for the ith decision is equal to $\Sigma\ p_j g_{ij}$.) Thus it must be, for example, that the situation is one in which a knowledge of likelihoods is impossible, or at best extremely insecure. In this case it is unreasonable not to be skeptical of probabilistic calculations unless there is no other way out, particularly if the decision is a fundamental one that needs to be justified to others.

The second feature that suggests the maximin rule is the following: the person choosing has a conception of the good such that he cares very little, if anything, for what he might gain above the minimum stipend that he can, in fact, be sure of by following the maximin rule. It is not worthwhile for him to take a chance for the sake of a further advantage, especially when it may turn out that he loses much that is important to him. This last provision brings in the third feature, namely, that the rejected alternatives have outcomes that one can hardly accept. The situation involves grave risks. Of course these features work most effectively in combination. The paradigm situation for following the maximin rule is when all three features are realized to the highest degree.

Let us review briefly the nature of the original position with these three special features in mind. To begin with, the veil of ignorance excludes all knowledge of likelihoods. The parties have no basis for determining the probable nature of their society, or their place in it. Thus they have no basis for probability calculations. They must also take into account the fact that their choice of principles should seem reasonable to others, in particular their descendants, whose rights will be deeply affected by it. These considerations are strengthened by the fact that the parties know very little about the possible states of society. Not only are they unable to conjecture the likelihoods of the various possible circumstances, they cannot say much about what the possible circumstances are, much less

20. Here I borrow from William Fellner, *Probability and Profit* (Homewood, Ill., R. D. Irwin, Inc., 1965), pp. 140–142, where these features are noted.

enumerate them and foresee the outcome of each alternative available. Those deciding are much more in the dark than illustrations by numerical tables suggest. It is for this reason that I have spoken only of a relation to the maximin rule.

Several kinds of arguments for the two principles of justice illustrate the second feature. Thus, if we can maintain that these principles provide a workable theory of social justice, and that they are compatible with reasonable demands of efficiency, then this conception guarantees a satisfactory minimum. There may be, on reflection, little reason for trying to do better. Thus much of the argument, especially in Part Two, is to show, by their application to some main questions of social justice, that the two principles are a satisfactory conception. These details have a philosophical purpose. Moreover, this line of thought is practically decisive if we can establish the priority of liberty. For this priority implies that the persons in the original position have no desire to try for greater gains at the expense of the basic equal liberties. The minimum assured by the two principles in lexical order is not one that the parties wish to jeopardize for the sake of greater economic and social advantages (§§33–35).

Finally, the third feature holds if we can assume that other conceptions of justice may lead to institutions that the parties would find intolerable. For example, it has sometimes been held that under some conditions the utility principle (in either form) justifies, if not slavery or serfdom, at any rate serious infractions of liberty for the sake of greater social benefits. We need not consider here the truth of this claim. For the moment, this contention is only to illustrate the way in which conceptions of justice may allow for outcomes which the parties may not be able to accept. And having the ready alternative of the two principles of justice which secure a satisfactory minimum, it seems unwise, if not irrational, for them to take a chance that these conditions are not realized.

So much, then, for a brief sketch of the features of situations in which the maximin rule is a useful maxim and of the way in which the arguments for the two principles of justice can be subsumed under them. Thus if the list of traditional views (§21) represents the possible decisions, these principles would be selected by the rule. The original position exhibits these special features to a sufficiently high degree in view of the fundamental character of the choice of a conception of justice. These remarks about the maximin rule are intended only to clarify the structure of the choice problem in the original position. I shall conclude this section by taking up an objection which is likely to be made against the difference principle and which leads into an important question. The

objection is that since we are to maximize (subject to the usual constraints) the prospects of the least advantaged, it seems that the justice of large increases or decreases in the expectations of the more advantaged may depend upon small changes in the prospects of those worst off. To illustrate: the most extreme disparities in wealth and income are allowed provided that they are necessary to raise the expectations of the least fortunate in the slightest degree. But at the same time similar inequalities favoring the more advantaged are forbidden when those in the worst position lose by the least amount. Yet it seems extraordinary that the justice of increasing the expectations of the better placed by a billion dollars, say, should turn on whether the prospects of the least favored increase or decrease by a penny. This objection is analogous to the following familiar difficulty with the maximin rule. Consider the sequence of gain-and-loss tables:

0	n
1/n	1

for all natural numbers n. Even if for some smallish number it is reasonable to select the second row, surely there is another point later in the sequence when it is irrational not to choose the first row contrary to the rule.

Part of the answer is that the difference principle is not intended to apply to such abstract possibilities. As I have said, the problem of social justice is not that of allocating *ad libitum* various amounts of something, whether it be money, or property, or whatever, among given individuals. Nor is there some substance of which expectations are made that can be shuffled from one representative man to another in all possible combinations. The possibilities which the objection envisages cannot arise in real cases; the feasible set is so restricted that they are excluded.[21] The reason for this is that the two principles are tied together as one conception of justice which applies to the basic structure of society as a whole. The operation of the principles of equal liberty and fair equality of opportunity prevents these contingencies from occurring. For we raise the expectations of the more advantaged only in ways required to improve the situation of the worst off. For the greater expectations of the more favored presumably cover the costs of training or answer to organizational requirements, thereby contributing to the general advantage. While nothing

21. I am indebted to S. A. Marglin for this point,

guarantees that inequalities will not be significant, there is a persistent tendency for them to be leveled down by the increasing availability of educated talent and ever widening opportunities. The conditions established by the other principles insure that the disparities likely to result will be much less than the differences that men have often tolerated in the past.

We should also observe that the difference principle not only assumes the operation of other principles, but it presupposes as well a certain theory of social institutions. In particular, as I shall discuss in Chapter V, it relies on the idea that in a competitive economy (with or without private ownership) with an open class system excessive inequalities will not be the rule. Given the distribution of natural assets and the laws of motivation, great disparities will not long persist. Now the point to stress here is that there is no objection to resting the choice of first principles upon the general facts of economics and psychology. As we have seen, the parties in the original position are assumed to know the general facts about human society. Since this knowledge enters into the premises of their deliberations, their choice of principles is relative to these facts. What is essential, of course, is that these premises be true and sufficiently general. It is often objected, for example, that utilitarianism may allow for slavery and serfdom, and for other infractions of liberty. Whether these institutions are justified is made to depend upon whether actuarial calculations show that they yield a higher balance of happiness. To this the utilitarian replies that the nature of society is such that these calculations are normally against such denials of liberty.

Contract theory agrees, then, with utilitarianism in holding that the fundamental principles of justice quite properly depend upon the natural facts about men in society. This dependence is made explicit by the description of the original position: the decision of the parties is taken in the light of general knowledge. Moreover, the various elements of the original position presuppose many things about the circumstances of human life. Some philosophers have thought that ethical first principles should be independent of all contingent assumptions, that they should take for granted no truths except those of logic and others that follow from these by an analysis of concepts. Moral conceptions should hold for all possible worlds. Now this view makes moral philosophy the study of the ethics of creation: an examination of the reflections an omnipotent deity might entertain in determining which is the best of all possible worlds. Even the general facts of nature are to be chosen. Certainly we have a natural religious interest in the ethics of creation. But it would

appear to outrun human comprehension. From the point of view of contract theory it amounts to supposing that the persons in the original position know nothing at all about themselves or their world. How, then, can they possibly make a decision? A problem of choice is well defined only if the alternatives are suitably restricted by natural laws and other constraints, and those deciding already have certain inclinations to choose among them. Without a definite structure of this kind the question posed is indeterminate. For this reason we need have no hesitation in making the choice of the principles of justice presuppose a certain theory of social institutions. Indeed, one cannot avoid assumptions about general facts any more than one can do without a conception of the good on the basis of which the parties rank alternatives. If these assumptions are true and suitably general, everything is in order, for without these elements the whole scheme would be pointless and empty.

It is evident from these remarks that both general facts as well as moral conditions are needed even in the argument for the first principles of justice. (Of course, it has always been obvious that secondary moral rules and particular ethical judgments depend upon factual premises as well as normative principles.) In a contract theory, these moral conditions take the form of a description of the initial contractual situation. It is also clear that there is a division of labor between general facts and moral conditions in arriving at conceptions of justice, and this division can be different from one theory to another. As I have noted before, principles differ in the extent to which they incorporate the desired moral ideal. It is characteristic of utilitarianism that it leaves so much to arguments from general facts. The utilitarian tends to meet objections by holding that the laws of society and of human nature rule out the cases offensive to our considered judgments. Justice as fairness, by contrast, embeds the ideals of justice, as ordinarily understood, more directly into its first principles. This conception relies less on general facts in reaching a match with our judgments of justice. It insures this fit over a wider range of possible cases.

There are two reasons that justify this embedding of ideals into first principles. First of all, and most obviously, the utilitarian's standard assumptions that lead to the wanted consequences may be only probably true, or even doubtfully so. Moreover, their full meaning and application may be highly conjectural. And the same may hold for all the requisite general suppositions that support the principle of utility. From the standpoint of the original position it may be unreasonable to rely upon these hypotheses and therefore far more sensible to embody the ideal more expressly in the principles chosen. Thus it seems that the parties would

prefer to secure their liberties straightway rather than have them depend upon what may be uncertain and speculative actuarial calculations. These remarks are further confirmed by the desirability of avoiding complicated theoretical arguments in arriving at a public conception of justice (§24). In comparison with the reasoning for the two principles, the grounds for the utility criterion trespass upon this constraint. But secondly, there is a real advantage in persons' announcing to one another once and for all that even though theoretical computations of utility always happen to favor the equal liberties (assuming that this is indeed the case here), they do not wish that things had been different. Since in justice as fairness moral conceptions are public, the choice of the two principles is, in effect, such an announcement. And the benefits of this collective profession favor these principles even though the utilitarian's assumptions should be true. These matters I shall consider in more detail in connection with publicity and stability (§29). The relevant point here is that while, in general, an ethical theory can certainly invoke natural facts, there may nevertheless be good reasons for embedding convictions of justice more directly into first principles than a theoretically complete grasp of the contingencies of the world may actually require.

27. THE REASONING LEADING TO THE PRINCIPLE OF AVERAGE UTILITY

I now wish to examine the reasoning that favors the principle of average utility. The classical principle is discussed later (§30). One of the merits of contract theory is that it reveals these principles to be markedly distinct conceptions, however much their practical consequences may coincide. Their underlying analytic assumptions are far apart in the sense that they are associated with contrasting interpretations of the initial situation. But first a word about the meaning of utility. It is understood in the traditional sense as the satisfaction of desire; and it admits of interpersonal comparisons that can at least be summed at the margin. I assume also that utility is measured by some procedure that is independent of choices involving risk, say by postulating an ability to rank differences between levels of satisfaction. These are the traditional assumptions; and while they are very strong, they will not be criticized here. As far as possible, I want to examine the historical doctrine on its own terms.

Applied to the basic structure, the classical principle requires that institutions be arranged to maximize the absolute weighted sum of the ex-

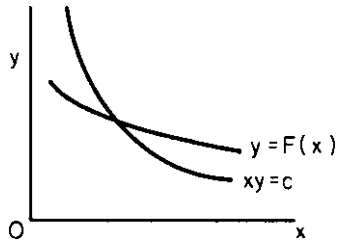
pectations of the relevant representative men. This sum is arrived at by weighting each expectation by the number of persons in the corresponding position and then adding. Thus, other things equal, when the number of persons in society doubles, total utility is twice as great. (Of course, on the utilitarian view expectations are to measure total satisfactions enjoyed and foreseen. They are not, as in justice as fairness, merely indexes of primary goods.) By contrast, the principle of average utility directs society to maximize not the total but the average utility (per capita). This seems to be a more modern view: it was held by Mill and Wicksell, and recently others have given it a new foundation.[22] To apply this conception to the basic structure, institutions are set up so as to maximize the percentage weighted sum of the expectations of representative individuals. To compute this sum we multiply expectations by the fraction of society at the corresponding position. Thus it is no longer true that, other things equal, when a community doubles its population the utility is twice as great. To the contrary, as long as the percentages in the various positions are unchanged, the utility remains the same.

Which of these principles of utility would be preferred in the original position? To answer this question, one should note that both variations come to the same thing if population size is constant. But when population is subject to change, there is a difference. The classical principle requires that so far as institutions affect the size of families, the age of marriage, and the like, they should be arranged so that the maximum of total utility is achieved. This entails that so long as the average utility per person falls slowly enough when the number of individuals increases, the population should be encouraged to grow indefinitely no matter how low the average has fallen. In this case the sum of utilities added by the greater number of persons is sufficiently great to make up for the decline in the share per capita. As a matter of justice and not of preference, a very low average of well-being may be required. (See the following figure.)

22. For Mill and Wicksell, see Gunnar Myrdal, *The Political Element in the Development of Economic Theory,* trans. Paul Streeten (London, Routledge and Kegan Paul, Ltd., 1953), pp. 38f. J. J. C. Smart in *An Outline of a System of Utilitarian Ethics* (Cambridge, The University Press, 1961), p. 18, leaves the matter unsettled, but affirms the classical principle in the case where it is necessary to break ties. As unambiguous proponents of the average doctrine, see J. C. Harsanyi, "Cardinal Utility in Welfare Economics and the Theory of Risk Taking," *Journal of Political Economy,* vol. 61 (1953), and "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy,* vol. 63 (1955); and R. B. Brandt, "Some Merits of One Form of Rule Utilitarianism," in *University of Colorado Studies* (Boulder, Colo., 1967), pp. 39–65. But note the qualification regarding Brandt's view in §29 below, note 31. For a discussion of Harsanyi, see P. K. Pattanaik, "Risk, Impersonality, and the Social Welfare Function," *Journal of Political Economy,* vol. 76 (1968), and Sen, *Collective Choice and Social Welfare,* pp. 141–146.

INDEFINITE INCREASE OF POPULATION

Formally the condition for increasing population size indefinitely is that the curve $y = F(x)$, where y is average per capita utility and x is population size, should be flatter than the rectangular hyperbola $xy = c$. For xy



equals the total utility, and the area of the rectangle representing this total increases as x increases whenever the curve $y = F(x)$ is flatter than $xy = c$.

Now this consequence of the classical principle seems to show that it would be rejected by the parties in favor of the average principle. The two principles would be equivalent only if it is supposed that average well-being always falls sufficiently fast (beyond a certain point anyway) so that there is no serious conflict between them. But this assumption seems questionable. From the standpoint of the persons in the original position, it would appear more rational to agree to some sort of floor to hold up average welfare. Since the parties aim to advance their own interests, they have no desire in any event to maximize the sum total of satisfaction. I assume, therefore, that the more plausible utilitarian alternative to the two principles of justice is the average and not the classical principle.

I now wish to consider how the parties might arrive at the average principle. The reasoning I shall sketch is perfectly general and if it were sound it would sidestep entirely the problem of how to present the alternatives. The average principle would be recognized as the only reasonable candidate. Imagine a situation in which a single rational individual can choose which of several societies to enter.[23] To fix ideas, assume first that the members of these societies all have the same preferences. And assume also that these preferences satisfy conditions that enable one to

23. Here I follow the first stages of W. S. Vickrey's presentation in "Utility, Strategy, and Social Decision Rules," *Quarterly Journal of Economics,* vol. 74 (1960), pp. 523f.

define a cardinal utility. Further, each society has the same resources and the same distribution of natural talents. Nevertheless, individuals with different talents have different incomes; and each society has a redistribution policy which if pushed beyond a certain point weakens incentives and thereby lowers production. Supposing that different policies are followed in these societies, how will a single individual decide which society to join? If he knows his own abilities and interests precisely, and if he has detailed information about these societies, he may be able to foresee the well-being that he will almost certainly enjoy in each one. He can then decide on this basis. There is no need for him to make any probabilistic calculations.

But this case is rather special. Let us alter it step by step so that it increasingly resembles that of someone in the original position. Thus, suppose first that the hypothetical joiner is unsure about the role his talents will enable him to fill in these various societies. If he assumes that his preferences are the same as everyone else, he may decide by trying to maximize his expected well-being. He computes his prospect for a given society by taking as the alternative utilities those of the representative members of that society and as the likelihoods for each position his estimates of his chances of attaining it. His expectation is defined, then, by a weighted sum of utilities of representative individuals, that is, by the expression $\Sigma p_i u_i$, where p_i is the likelihood of his achieving the ith position, and u_i the utility of the corresponding representative man. He then chooses the society offering the highest prospect.

Several further modifications bring the situation closer to that of the original position. Assume that the hypothetical joiner knows nothing about either his abilities or the place he is likely to hold in each society. It is still assumed, though, that his preferences are the same as the people in these societies. Now suppose that he continues to reason along probabilistic lines by holding that he has an equal chance of being any individual (that is, that his chance of falling under any representative man is the fraction of society that this man represents). In this case his prospects are still identical with the average utility for each society. These modifications have at last brought his expected gains for each society in line with its average welfare.

So far we have assumed that all individuals have similar preferences whether or not they belong to the same society. Their conceptions of the good are roughly the same. Once this highly restrictive assumption is dropped, we take the final step and arrive at a variation of the initial situation. Nothing is known, let us say, about the particular preferences of

the members of these societies or of the person deciding. These facts as well as a knowledge of the structure of these societies are ruled out. The veil of ignorance is now complete. But one can still imagine that the hypothetical newcomer reasons much as before. He assumes that there is an equal likelihood of his turning out to be anyone, fully endowed with that person's preferences, abilities, and social position. Once again his prospect is highest for that society with the greatest average utility. We can see this in the following way. Let n be the number of persons in a society. Let their levels of well-being be $u_1, u_2, \ldots, u_n$. Then the total utility is Σu_i and the average is $\Sigma u_i/n$. Assuming that one has an equal chance of being any person, one's prospect is: $1/n\ u_1 + 1/n\ u_2 + \ldots + 1/n\ u_n$ or $\Sigma u_i/n$. The value of the prospect is identical with the average utility.

Thus if we waive the problem of interpersonal comparisons of utility, and if the parties are viewed as rational individuals who have no aversion to risk and who follow the principle of insufficient reason in computing likelihoods (the principle that underlies the preceding probabilistic calculations), then the idea of the initial situation leads naturally to the average principle. By choosing it the parties maximize their expected well-being as seen from this point of view. Some form of contract theory provides, then, a way of arguing for the average over the classical view. In fact, how else is the average principle to be accounted for? After all, it is not a teleological doctrine, strictly speaking, as the classical view is, and therefore it lacks some of the intuitive appeal of the idea of maximizing the good. Presumably one who held the average principle would want to invoke the contract theory at least to this extent.

In the preceding discussion I have assumed that utility is understood in the traditional sense as the satisfaction of desire and cardinal interpersonal comparisons are regarded as possible. But this notion of utility has been largely abandoned by economic theory in recent decades; it is thought to be too vague and to play no essential role in explaining economic behavior. Utility is now understood as a way of representing the choices of economic agents and not as a measure of satisfaction. The main kind of cardinal utility presently recognized derives from the Neuman-Morgenstern construction, which is based on choices between prospects involving risks (§49). Unlike the traditional notion, this measure takes attitudes to uncertainty into account and it does not seek to provide a basis for interpersonal comparisons. Nevertheless, it is still possible to formulate the principle of average utility using this kind of measure: one supposes the parties in the original position, or some variant thereof, to

have Neuman-Morgenstern utility function and to assess their prospects accordingly.[24] Of course, certain precautions must be taken; for example, these utility functions cannot take into account all kinds of considerations but must reflect the parties' estimate of what furthers their good. If they were influenced by other reasons, we would not have a teleological theory.

When these restrictions are observed, however, an average utilitarian view can be stated that takes into account the high level of risk aversion that it seems any normal person would have in the original position; and the greater this risk aversion the more this form of utility principle would resemble the difference principle, at least when the evaluation of economic benefits is in question. Of course, these two principles are not the same, since there are many important differences between them. But there is this similarity: risk and uncertainty from a suitably general perspective leads both views to weight more heavily the advantages of those whose situation is less fortunate. In fact, reasonable risk aversion may be so great, once the enormous hazards of the decision in the original position are fully appreciated, that the utilitarian weighting may be, for practical purposes, so close to the difference principle as to make the simplicity of the latter (§49) decisive in its favor.

28. SOME DIFFICULTIES WITH THE AVERAGE PRINCIPLE

Before taking up the arguments for the two principles of justice I wish to mention several difficulties with the average principle of utility. First, though, we should note an objection which turns out to be only apparent. As we have seen, this principle may be viewed as the ethics of a single rational individual prepared to take whatever chances necessary to maximize his prospects from the standpoint of the initial situation. (If there is no objective basis for probabilities, they are computed by the principle of insufficient reason.) Now it is tempting to argue against this principle that

24. How this might be done was shown by J. C. Harsanyi. See his "Cardinal Utility in Welfare Economics and the Theory of Risk Taking," *Journal of Political Economy,* vol. 61 (1953), and "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy,* vol. 63 (1955). For a discussion of some of the difficulties with this formulation, see P. K. Pattanaik, *Voting and Collective Choice* (Cambridge, The University Press, 1971), ch. 9, and A. K. Sen, *Collective Choice and Social Welfare,* pp. 141–146. An accessible account of the contrast between the traditional and Neuman-Morgenstern notions of utility can be found in Daniel Ellsberg, "Classic and Current Notions of 'Measurable Utility,'" *Economic Journal,* vol. 64 (1963).

it presupposes a real and equal acceptance of risk by all members of society. At some time, one wants to say, everyone must actually have agreed to take the same chances. Since clearly there was no such occasion, the principle is unsound. Consider an extreme case: a slaveholder when confronted by his slaves attempts to justify his position to them by claiming that, first of all, given the circumstances of their society, the institution of slavery is in fact necessary to produce the greatest average happiness; and secondly, that in the initial contractual situation he would choose the average principle even at the risk of its subsequently happening that he is justifiably held a slave. Now offhand we are inclined to reject the slaveholder's argument as beside the point, if not outrageous. One may think that it makes no difference what he would choose. Unless individuals have actually agreed to a conception of justice subject to real risks, no one is bound by its requirements.

On the contract view, however, the general form of the slaveholder's argument is correct. It would be a mistake for the slaves to retort that his contentions are irrelevant since there has been no actual occasion of choice, no equal sharing of risk as to how things would turn out. The contract doctrine is purely hypothetical: if a conception of justice would be agreed to in the original position, its principles are the right ones to apply. It is no objection that such an understanding has never been nor ever will be entered into. We cannot have it both ways: we cannot interpret the theory of justice hypothetically when the appropriate occasions of consent cannot be found to explain individuals' duties and obligations, and then insist upon real situations of risk-bearing to throw out principles of justice that we do not want.[25] Thus in justice as fairness the way to refute the slaveholder's argument is to show that the principle he invokes would be rejected in the original position. We have no alternative but to exploit the various aspects of this initial situation (on the favored interpretation) to establish that the balance of reasons favors the two principles of justice. In the next section I shall start on this task.

The first difficulty with the average principle I have already mentioned in discussing the maximin rule as a heuristic device for arranging the arguments favoring the two principles. It concerns the way that a rational individual is to estimate probabilities. This question arises because there seem to be no objective grounds in the initial situation for assuming that one has an equal chance of turning out to be anybody. That is, this

25. I have myself been in error on this matter. See "Constitutional Liberty and the Concept of Justice," *Nomos VI: Justice,* ed. C. J. Friedrich and J. W. Chapman (New York, Atherton Press, 1963), pp. 109–114. I am grateful to O. H. Harman for clarification on this point.

assumption is not founded upon known features of one's society. In the early stages of the argument leading to the average principle, the hypothetical newcomer does have some knowledge of his abilities and of the design of the societies among which he is choosing. The estimates of his chances are based upon this information. But at the last stage there is complete ignorance of particular facts (with the exception of those implied by the circumstances of justice). The construction of the individual's prospect depends at this stage solely upon the principle of insufficient reason. This principle is used to assign probabilities to outcomes in the absence of any information. When we have no evidence at all, the possible cases are taken to be equally probable. Thus Laplace reasoned that when we are drawing from two urns each containing a different ratio of black to red balls, but we have no information as to which urn we are faced with, then we should assume initially that the chance of drawing from each of these urns is the same. The idea is that the state of ignorance on the basis of which these prior probabilities are assigned presents the same sort of problem as the situation where one has a lot of evidence showing that a particular coin is unbiased. What is distinctive about the use of the principle is that it enables one to incorporate different kinds of information within one strictly probabilistic framework and to draw inferences about probabilities even in the absence of knowledge. Prior probabilities however arrived at are part of one theory along with estimates of chances based on random sampling. The limiting case of no information does not pose a theoretical problem.[26] As evidence accumulates the prior probabilities are revised anyway and the principle of insufficient reason at least insures that no possibilities are excluded from the outset.

Now I shall assume that the parties discount likelihoods arrived at solely on the basis of this principle. This supposition is plausible in view of the fundamental importance of the original agreement and the desire to have one's decision appear responsible to one's descendants who will be affected by it. We are more reluctant to take great risks for them than for ourselves; and we are willing to do so only when there is no way to avoid these uncertainties, or when the probable gains, as estimated by objective information, are so large that it would appear to them irresponsible to

26. See William Fellner, *Probability and Profit,* pp. 27f. The principle of insufficient reason in its classical form is known to lead to difficulties. See J. M. Keynes, *A Treatise on Probability* (London, Macmillan, 1921), ch. IV. Part of Rudolf Carnap's aim in his *Logical Foundations of Probability,* 2nd ed. (Chicago: University of Chicago Press, 1962), is to construct a system of inductive logic by finding other theoretical means to do what the classical principle was intended to do. See pp. 344f.

have refused the chance offered even though accepting it should actually turn out badly. Since the parties have the alternative of the two principles of justice, they can in large part sidestep the uncertainties of the original position. They can guarantee the protection of their liberties and a reasonably satisfactory standard of life as the conditions of their society permit. In fact, as I argue in the next section, it is questionable whether the choice of the average principle really offers a better prospect anyway, waiving the fact that it is based on the principle of insufficient reason. It seems, then, that the effect of the veil of ignorance is to favor the two principles. This conception of justice is better suited to the situation of complete ignorance.

There are, to be sure, assumptions about society that, if they were sound, would allow the parties to arrive at objective estimates of equal probability. To see this one can convert an argument of Edgeworth for the classical principle into one for average utility.[27] In fact, his reasoning can be adjusted to support nearly any general standard of policy. Edgeworth's idea is to formulate certain reasonable assumptions under which it would be rational for self-interested parties to agree to the standard of utility as a political principle to assess social policies. The necessity for such a principle arises because the political process is not a competitive one and these decisions cannot be left to the market. Some other method must be found to reconcile divergent interests. Edgeworth believes that the principle of utility would be agreed to by self-interested parties as the desired criterion. His thought seems to be that over the long run of many occasions, the policy of maximizing utility on each occasion is most likely to give the greatest utility for any person individually. Consistent application of this standard to taxation and property legislation, and so on, is calculated to give the best results from any one man's point of view. Therefore by adopting this principle self-interested parties have reasonable assurance that they will not lose out in the end and, in fact, will best improve their prospects.

The flaw in Edgeworth's idea is that the necessary assumptions are extremely unrealistic, especially in the case of the basic structure.[28] To

27. See F. Y. Edgeworth, *Mathematical Psychics* (London, 1888), pp. 52–56, and the first pages of "The Pure Theory of Taxation," *Economic Journal,* vol. 7 (1897). See also R. B. Brandt, *Ethical Theory* (Englewood Cliffs, N.J., Prentice Hall, Inc., 1959), pp. 376f.

28. Here I apply to Edgeworth an argument used by I. M. D. Little in his *Critique of Welfare Economics,* 2nd ed. (Oxford, The Clarendon Press, 1957), against a proposal of J. R. Hicks. See pp. 93f, 113f.

state these assumptions is to see how implausible they are. We must suppose that the effects of the decisions which make up the political process are not only more or less independent, but roughly of the same order in their social results, which cannot be very great in any event, otherwise these effects could not be independent. Moreover, it must be assumed either that men move from one social position to another in random fashion and live long enough for gains and losses to average out, or else that there is some mechanism which insures that legislation guided by the principle of utility distributes its favors evenly over time. But clearly society is not a stochastic process of this type; and some questions of social policy are much more vital than others, often causing large and enduring shifts in the institutional distribution of advantages.

Consider, for example, the case where a society is contemplating a historic change in its trade policies with foreign countries. The question is whether it shall remove long-standing tariffs on the import of agricultural products in order to obtain cheaper foodstuffs for workers in new industries. The fact that the change is justified on utilitarian grounds does not mean that it will not permanently affect the relative positions of those belonging to the landed and the industrial classes. Edgeworth's reasoning holds only when each of the many decisions has a relatively small and temporary influence on distributive shares and there is some institutional device insuring randomness. Under realistic assumptions, then, his argument can establish at best only that the principle of utility has a subordinate place as a legislative standard for lesser questions of policy. But this clearly implies that the principle fails for the main problems of social justice. The pervasive and continuing influence of our initial place in society and of our native endowments, and of the fact that the social order is one system, is what characterizes the problem of justice in the first place. We must not be enticed by mathematically attractive assumptions into pretending that the contingencies of men's social positions and the asymmetries of their situations somehow even out in the end. Rather, we must choose our conception of justice fully recognizing that this is not and cannot be the case.

It seems, then, that if the principle of average utility is to be accepted, the parties must reason from the principle of insufficient reason. They must follow what some have called the Laplacean rule for choice under uncertainty. The possibilities are identified in some natural way and each assigned the same likelihood. No general facts about society are offered to support these assignments; the parties carry on with probabilistic calculations as if information had not run out. Now I cannot discuss here the

concept of probability, but a few points should be noted.[29] First of all, it may be surprising that the meaning of probability should arise as a problem in moral philosophy, especially in the theory of justice. It is, however, the inevitable consequence of the contract doctrine which conceives of moral philosophy as part of the theory of rational choice. Considerations of probability are bound to enter in given the way in which the initial situation is defined. The veil of ignorance leads directly to the problem of choice under complete uncertainty. Of course, it is possible to regard the parties as perfect altruists and to assume that they reason as if they are certain to be in the position of each person. This interpretation of the initial situation removes the element of risk and uncertainty (§30).

In justice as fairness, however, there is no way to avoid this question entirely. The essential thing is not to allow the principles chosen to depend on special attitudes toward risk. For this reason the veil of ignorance also rules out the knowledge of these inclinations: the parties do not know whether or not they have an unusual aversion to taking chances. As far as possible the choice of a conception of justice should depend on a rational assessment of accepting risks unaffected by peculiar individual preferences for taking chances one way or the other. Of course, a social system may take advantage of these varying propensities by having institutions that permit them full play for common ends. But ideally anyway, the basic design of the system should not depend on one of these attitudes (§81). Therefore, it is not an argument for the two principles of justice that they express a peculiarly conservative point of view about taking chances in the original position. What must be shown is that given the unique features of this situation, agreeing to these principles rather than the principle of utility is rational for anyone whose aversion to uncertainty in regard to being able to secure their fundamental interests is within the normal range.

Secondly, I have simply assumed that judgments of probability, if they are to be grounds of rational decision, must have an objective basis, that is, a basis in knowledge of particular facts (or in reasonable beliefs). This evidence need not take the form of reports of relative frequencies but it should provide grounds for estimating the relative strength of the various tendencies that affect the outcome. The necessity for objective reasons is

29. William Fellner, *Probability and Profit,* pp. 210–233, contains a useful bibliography with brief commentaries. Particularly important for the recent development of the so-called Bayesian point of view is L. J. Savage, *The Foundations of Statistics* (New York, John Wiley and Sons, Inc., 1954). For a guide to the philosophical literature, see H. E. Kyburg, *Probability and Inductive Logic* (Riverside, N.J., Macmillan, 1970).

all the more urgent in view of the fundamental significance of the choice in the original position and the fact that the parties want their decision to appear well founded to others. I shall assume, therefore, to fill out the description of the original position, that the parties discount estimates of likelihoods not supported by a knowledge of particular facts and that derive largely if not solely from the principle of insufficient reason. The requirement of objective grounds does not seem to be in dispute between neo-Bayesian theorists and those adhering to more classical ideas. The controversy in this case is how far intuitive and imprecise estimates of likelihoods based on common sense and the like should be incorporated into the formal apparatus of the theory of probability rather than used in an ad hoc way to adjust the conclusions reached by methods that leave this information out of account.[30] Here neo-Bayesians have a strong case. Surely it is better when possible to use our intuitive knowledge and common sense hunches in a systematic and not in an irregular and unexplained manner. But none of this affects the contention that judgments of probability must have some objective basis in the known facts about society if they are to be rational grounds of decision in the special situation of the original position.

The last difficulty I shall mention here raises a deep problem. Although I cannot deal with it properly, it should not be passed over. The trouble arises from the peculiarity of the expectation in the final step of the reasoning for the average principle. When expectations are computed in the normal case, the utilities of the alternatives (the u_i in the expression $\Sigma p_i u_i$) are derived from a single system of preferences, namely those of the individual making the choice. The utilities represent the worth of the alternatives for this person as estimated by his scheme of values. In the present case, however, each utility is based on the preferences of a different person. There are as many distinct persons as there are utilities. Of course, it is clear that this reasoning presupposes interpersonal comparisons. But leaving aside for the moment the problem of defining these, the point to notice here is that the individual is thought to choose as if he has no aims at all which he counts as his own. He takes a chance on being any one of a number of persons complete with each individual's system of ends, abilities, and social position. We may doubt whether this expectation is a meaningful one. Since there is no one scheme of aims by which its estimates have been arrived at, it lacks the necessary unity.

30. See Fellner, *Probability and Profit,* pp. 48–67, and Luce and Raiffa, *Games and Decisions,* pp. 318–334.

To clarify this problem, let us distinguish between evaluating objective situations and evaluating aspects of the person: abilities, traits of character, and system of aims. Now from our point of view it is often easy enough to appraise another individual's situation as specified say by his social position, wealth, and the like, or by his prospects in terms of primary goods. We put ourselves in his shoes, complete with our character and preferences (not his), and take account of how our plans would be affected. We can go further. We can assess the worth to us of being in another's place with at least some of his traits and aims. Knowing our plan of life, we can decide whether it would be rational for us to have those traits and aims, and therefore advisable for us to develop and encourage them if we can. But in constructing our expectation, how are we to assess another's way of life and system of final ends? By our aims or by his? The contract argument assumes that we must decide from our own standpoint: The worth to us of the way of life of another and the realization of his ends (his total circumstances) is not, as the previously constructed expectation assumes, its worth to him. Moreover, the circumstances of justice imply that these values differ sharply. Conflicting claims arise not only because people want similar sorts of things to satisfy similar desires (for example, food and clothes for essential needs) but because their conceptions of the good differ; and while the worth to us of basic primary goods may be agreed to be comparable to their worth to others, this agreement cannot be extended to the satisfaction of our final ends. To be sure, the parties do not know their own final ends, but they do know that, in general, these ends are opposed and subject to no commonly acceptable measure. This value of someone's total circumstances to him is not the same as its value to us. Thus the expectation of the final step in the argument for the average principle of utility cannot be correct.

We may formulate the difficulty in a somewhat different way. The reasoning for the average principle must somehow define a unified expectation. Suppose, then, that the parties agree to base interpersonal comparisons on certain rules. These rules become part of the meaning of the utility principle just as the use of an index of primary goods is part of the meaning of the difference principle. Thus these comparison rules (as I shall call them) may be thought to derive, for example, from certain psychological laws that determine people's satisfaction given certain parameters such as strength of preferences and desires, natural abilities and physical attributes, private and public goods enjoyed, and so on. Individuals characterized by the same parameters are agreed to have the same

satisfaction; and so granted the acceptance of these comparison-rules, average satisfaction can be defined and the parties are assumed to maximize their expected satisfaction so understood. Thus everyone thinks of themselves as having the same deep utility function, so to speak, and regards the satisfaction that others achieve as legitimate entries into their own expectations as seen from the perspective of the original position. The same unified expectation holds for all and (using the Laplacean rule) the agreement on the principle of average utility follows.

It is crucial to note that this reasoning presupposes a particular conception of the person. The parties are conceived as having no definite highest-order interests or fundamental ends by reference to which they decide what sorts of persons they care to be. They have, as it were, no determinate character of will. They are, we might say, bare-persons: as settled by certain comparison rules, they are equally prepared to accept as defining their good whatever evaluations these rules assign to the realization of their, or anyone else's, final ends, even if these evaluations conflict with those required by their existing fundamental interests. But we have assumed that the parties do have a determinate character and will, even though the specific nature of their system of ends is unknown to them. They are, so to speak, determinate-persons: they have certain highest-order interests and fundamental ends by reference to which they would decide the kind of life and subordinate aims that are acceptable to them. It is these interests and ends, whatever they are, which they must try to protect. Since they know that the basic liberties covered by the first principle will secure these interests, they must acknowledge the two principles of justice rather than the principle of utility.

Thus to sum up: I have argued that the expectation on which the reasoning for the average principle relies is faulty on two counts. First, since there are no objective grounds in the original position for accepting equal likelihoods, or indeed, any other probability distribution, these likelihoods are merely as-if probabilities. They depend solely on the principle of insufficient reason and provide no independent reason for accepting the utility principle. Instead, the appeal to these likelihoods is, in effect, an indirect way of stipulating this principle. Second, the utilitarian argument assumes that the parties have no definite character or will, that they are not persons with determinate final interests, or a particular conception of their good, that they are concerned to protect. Thus taking both counts together, the utilitarian reasoning arrives at a purely formal expression for an expectation, but one that lacks an appropriate meaning. It is as if one continued to use probabilistic arguments and ways of making interper-

sonal comparisons long after the conditions for their legitimate use had been ruled out by the circumstances of the original position.

29. SOME MAIN GROUNDS FOR THE TWO PRINCIPLES OF JUSTICE

In this section I use the conditions of publicity and finality to give some of the main arguments for the two principles of justice. I shall rely upon the fact that for an agreement to be valid, the parties must be able to honor it under all relevant and foreseeable circumstances. There must be a rational assurance that one can carry through. The arguments I shall adduce fit under the heuristic schema suggested by the reasons for following the maximin rule. That is, they help to show that the two principles are an adequate minimum conception of justice in a situation of great uncertainty. Any further advantages that might be won by the principle of utility are highly problematical, whereas the hardship if things turn out badly are intolerable. It is at this point that the concept of a contract has a definite role: it suggests the condition of publicity and sets limits upon what can be agreed to.

The first confirming ground for the two principles can be explained in terms of what I earlier referred to as the strains of commitment. I said (§25) that the parties have a capacity for justice in the sense that they can be assured that their undertaking is not in vain. Assuming that they have taken everything into account, including the general facts of moral psychology, they can rely on one another to adhere to the principles adopted. Thus they consider the strains of commitment. They cannot enter into agreements that may have consequences they cannot accept. They will avoid those that they can adhere to only with great difficulty. Since the original agreement is final and made in perpetuity, there is no second chance. In view of the serious nature of the possible consequences, the question of the burden of commitment is especially acute. A person is choosing once and for all the standards which are to govern his life prospects. Moreover, when we enter an agreement we must be able to honor it even should the worst possibilities prove to be the case. Otherwise we have not acted in good faith. Thus the parties must weigh with care whether they will be able to stick by their commitment in all circumstances. Of course, in answering this question they have only a general knowledge of human psychology to go on. But this information is enough to tell which conception of justice involves the greater stress.

In this respect the two principles of justice have a definite advantage. Not only do the parties protect their basic rights but they insure themselves against the worst eventualities. They run no chance of having to acquiesce in a loss of freedom over the course of their life for the sake of a greater good enjoyed by others, an undertaking that in actual circumstances they might not be able to keep. Indeed, we might wonder whether such an agreement can be made in good faith at all. Compacts of this sort exceed the capacity of human nature. How can the parties possibly know, or be sufficiently sure, that they can keep such an agreement? Certainly they cannot base their confidence on a general knowledge of moral psychology. To be sure, any principle chosen in the original position may require a large sacrifice for some. The beneficiaries of clearly unjust institutions (those founded on principles which have no claim to acceptance) may find it hard to reconcile themselves to the changes that will have to be made. But in this case they will know that they could not have maintained their position anyway. In any case, the two principles of justice provide an alternative. If the only possible candidates all involved similar risks, the problem of the strains of commitment would have to be waived. This is not the case, and judged in this light the two principles seem distinctly superior.

A second consideration invokes the condition of publicity as well as that of the constraints on agreements. I shall present the argument in terms of the question of psychological stability. Earlier I stated that a strong point in favor of a conception of justice is that it generates its own support. When the basic structure of society is publicly known to satisfy its principles for an extended period of time, those subject to these arrangements tend to develop a desire to act in accordance with these principles and to do their part in institutions which exemplify them. A conception of justice is stable when the public recognition of its realization by the social system tends to bring about the corresponding sense of justice. Now whether this happens depends, of course, on the laws of moral psychology and the availability of human motives. I shall discuss these matters later on (§§75–76). At the moment we may observe that the principle of utility seems to require a greater identification with the interests of others than the two principles of justice. Thus the latter will be a more stable conception to the extent that this identification is difficult to achieve. When the two principles are satisfied, each person's basic liberties are secured and there is a sense defined by the difference principle in which everyone is benefited by social cooperation. Therefore we can explain the acceptance of the social system and the principles it satisfies

by the psychological law that persons tend to love, cherish, and support whatever affirms their own good. Since everyone's good is affirmed, all acquire inclinations to uphold the scheme.

When the principle of utility is satisfied, however, there is no such assurance that everyone benefits. Allegiance to the social system may demand that some, particularly the less favored, should forgo advantages for the sake of the greater good of the whole. Thus the scheme will not be stable unless those who must make sacrifices strongly identify with interests broader than their own. But this is not easy to bring about. The sacrifices in question are not those asked in times of social emergency when all or some must pitch in for the common good. The principles of justice apply to the basic structure of the social system and to the determination of life prospects. What the principle of utility asks is precisely a sacrifice of these prospects. Even when we are less fortunate, we are to accept the greater advantages of others as a sufficient reason for lower expectations over the whole course of our life. This is surely an extreme demand. In fact, when society is conceived as a system of cooperation designed to advance the good of its members, it seems quite incredible that some citizens should be expected, on the basis of political principles, to accept still lower prospects of life for the sake of others. It is evident then why utilitarians should stress the role of sympathy in moral learning and the central place of benevolence among the moral virtues. Their conception of justice is threatened with instability unless sympathy and benevolence can be widely and intensely cultivated. Looking at the question from the standpoint of the original position, the parties would reject the principle of utility and adopt the more realistic idea of designing the social order on a principle of reciprocal advantage. We need not suppose, of course, that in everyday life persons never make substantial sacrifices for one another, since moved by affection and ties of sentiment they often do. But such actions are not demanded as a matter of justice by the basic structure of society.

Furthermore, the public recognition of the two principles gives greater support to men's self-respect and this in turn increases the effectiveness of social cooperation. Both effects are reasons for agreeing to these principles. It is clearly rational for men to secure their self-respect. A sense of their own worth is necessary if they are to pursue their conception of the good with satisfaction and to take pleasure in its fulfillment. Self-respect is not so much a part of any rational plan of life as the sense that one's plan is worth carrying out. Now our self-respect normally depends upon the respect of others. Unless we feel that our endeavors are respected by

them, it is difficult if not impossible for us to maintain the conviction that our ends are worth advancing (§67). Hence for this reason the parties would accept the natural duty of mutual respect which asks them to treat one another civilly and to be willing to explain the grounds of their actions, especially when the claims of others are overruled (§51). Moreover, one may assume that those who respect themselves are more likely to respect each other and conversely. Self-contempt leads to contempt of others and threatens their good as much as envy does. Self-respect is reciprocally self-supporting.

Thus a desirable feature of a conception of justice is that it should publicly express men's respect for one another. In this way they insure a sense of their own value. Now the two principles achieve this end. For when society follows these principles, everyone's good is included in a scheme of mutual benefit and this public affirmation in institutions of each man's endeavors supports men's self-esteem. The establishment of equal liberty and the operation of the difference principle are bound to have this effect. The two principles are equivalent, as I have remarked, to an undertaking to regard the distribution of natural abilities in some respects as a collective asset so that the more fortunate are to benefit only in ways that help those who have lost out (§17). I do not say that the parties are moved by the ethical propriety of this idea. But there are reasons for them to accept this principle. For by arranging inequalities for reciprocal advantage and by abstaining from the exploitation of the contingencies of nature and social circumstance within a framework of equal liberties, persons express their respect for one another in the very constitution of their society. In this way they insure their self-respect as it is rational for them to do.

Another way of putting this is to say that the principles of justice manifest in the basic structure of society men's desire to treat one another not as means only but as ends in themselves. I cannot examine Kant's view here.[31] Instead I shall freely interpret it in the light of the contract doctrine. The notion of treating men as ends in themselves and never as only a means obviously needs an explanation. How can we always treat everyone as an end and never as a means only? Certainly we cannot say that it comes to treating everyone by the same general principles, since this interpretation makes the concept equivalent to formal justice. On the contract interpretation treating men as ends in themselves implies at the

31. See *The Foundations of the Metaphysics of Morals,* pp. 427–430 of vol. IV of *Kants Gesammelten Schriften,* Preussische Akademie der Wissenschaften (Berlin, 1913), where the second formulation of the categorical imperative is introduced.

very least treating them in accordance with the principles to which they would consent in an original position of equality. For in this situation men have equal representation as moral persons who regard themselves as ends and the principles they accept will be rationally designed to protect the claims of their person. The contract view as such defines a sense in which men are to be treated as ends and not as means only.

But the question arises whether there are substantive principles which convey this idea. If the parties wish to express this notion visibly in the basic structure of their society in order to secure each man's rational interest in his self-respect, which principles should they choose? Now it seems that the two principles of justice achieve this aim: for all have equal basic liberties and the difference principle interprets the distinction between treating men as a means only and treating them also as ends in themselves. To regard persons as ends in themselves in the basic design of society is to agree to forgo those gains which do not contribute to everyone's expectations. By contrast, to regard persons as means is to be prepared to impose on those already less favored still lower prospects of life for the sake of the higher expectations of others. Thus we see that the difference principle, which at first appears rather extreme, has a reasonable interpretation. If we further suppose that social cooperation among those who respect each other and themselves as manifest in their institutions is likely to be more effective and harmonious, the general level of expectations, assuming we could estimate it, may be higher when the two principles of justice are satisfied than one might otherwise have thought. The advantage of the principle of utility in this respect is no longer so clear.

The principle of utility presumably requires some who are less fortunate to accept even lower life prospects for the sake of others. To be sure, it is not necessary that those having to make such sacrifices rationalize this demand by having a lesser appreciation of their own worth. It does not follow from the utilitarian doctrine that it is because their aims are trivial or unimportant that some individuals' expectations are less. But the parties must consider the general facts of moral psychology. Surely it is natural to experience a loss of self-respect, a weakening of our sense of the value of accomplishing our aims, when we are already less favored. This is particularly likely to be so when social cooperation is arranged for the good of individuals. That is, those with greater advantages do not claim that they are necessary to preserve certain religious or cultural values which everyone has a duty to maintain. We are not here considering a doctrine of traditional order nor the principle of perfectionism, but

rather the principle of utility. In this instance, then, men's self-respect hinges on how they regard one another. If the parties accept the utility criterion, they will lack the support to their self-respect provided by the public commitment of others to arrange inequalities to everyone's advantage and to guarantee the basic liberties for all. In a public utilitarian society men, particularly the least advantaged, will find it more difficult to be confident of their own worth.

The utilitarian may answer that in maximizing the average utility these matters are already taken into account. If, for example, the equal liberties are necessary for men's self-respect and the average utility is higher when they are affirmed, then of course they should be established. So far so good. But the point is that we must not lose sight of the publicity condition. This requires that in maximizing the average utility we do so subject to the constraint that the utilitarian principle is publicly accepted and followed as the fundamental charter of society. What we cannot do is to raise the average utility by encouraging men to adopt and apply non-utilitarian principles of justice. If, for whatever reasons, the public recognition of utilitarianism entails some loss of self-esteem, there is no way around this drawback. It is an unavoidable cost of the utilitarian scheme given our stipulations. Thus suppose that the average utility is actually greater should the two principles of justice be publicly affirmed and realized in the basic structure. For the reasons mentioned, this may conceivably be the case. These principles would then represent the most attractive prospect, and on both lines of reasoning just examined, the two principles would be accepted. The utilitarian cannot reply that one is now really maximizing the average utility. In fact, the parties would have chosen the two principles of justice.

We should note, then, that utilitarianism, as I have defined it, is the view that the principle of utility is the correct principle for society's public conception of justice. And to show this one must argue that this criterion would be chosen in the original position. If we like, we can define a different variation of the initial situation in which the motivation assumption is that the parties want to adopt those principles that maximize average utility. The preceding remarks indicate that the two principles of justice may still be chosen. But if so, it is a mistake to call these principles—and the theory in which they appear—utilitarian. The motivation assumption by itself does not determine the character of the whole theory. In fact, the case for the principles of justice is strengthened if they would be chosen under different motivation assumptions. This indicates that the theory of justice is firmly grounded and not sensitive to slight changes in

this condition. What we want to know is which conception of justice characterizes our considered judgments in reflective equilibrium and best serves as the public moral basis of society. Unless one maintains that this conception is given by the principle of utility, one is not a utilitarian.[32]

The strains of commitment and the publicity condition, both of which we have discussed in this section, are also important. The first arises from the fact that, in general, the class of things that can be agreed to is included within, but smaller than, the class of things that can be rationally chosen. We can decide to take a chance and the same time fully intend that, should things turn out badly, we shall do what we can to retrieve our situation. But if we make an agreement, we have to accept the result; and so to give an undertaking in good faith, we must not only intend to honor it but with reason believe that we can do so. Thus the contract condition excludes a certain kind of randomizing. One cannot agree to a principle if there is a real possibility that it has any outcome that one will not be able to accept. I shall not comment further on the publicity condition except to note that it ties in with the desirability of embedding ideals in first principles (end of §26), with simplicity (§49), and with stability. The latter is examined further in what I have called the second part of the argument (§§79–82).

The form of the argument for the two principles is that the balance of reasons favors them over the principle of average utility, and assuming transitivity, over the classical doctrine as well. Thus the agreement of the parties depends on weighing various considerations. The reasoning is informal and not a proof, and there is an appeal to intuition as the basis of the theory of justice. Yet, as I have remarked (§21), when everything is tallied up, it may be clear where the balance of reasons lies. If so, then to the extent that the original position embodies reasonable conditions used in the justification of principles in everyday life, the claim that one would agree to the principles of justice is perfectly credible. Thus they can serve as a conception of justice in the public acceptance of which persons can recognize one another's good faith.

It may be helpful at this point to list some of the main grounds in favor of the two principles of justice over the principle of average utility. That

32. Thus while Brandt holds that a society's moral code is to be publicly recognized, and that the best code from a philosophical standpoint is the one that maximizes average utility, he does not maintain that the principle of utility must belong to the code itself. In fact, he denies that within the public morality the final court of appeal need be to utility. Thus by the definition in the text, his view is not utilitarian. See "Some Merits of One Form of Rule Utilitarianism," *University of Colorado Studies* (Boulder, Colo., 1967), pp. 58f.

the conditions of generality of principle, universality of application, and limited information are not sufficient by themselves to require these principles is clear from the reasoning for the utility principle (§27). Further assumptions must, therefore, be incorporated into the original position. Thus, I have assumed that the parties regard themselves as having certain fundamental interests that they must protect if they can; and that, as free persons, they have a highest-order interest in maintaining their liberty to revise and alter these ends (§26). The parties are, so to speak, persons with determinate interests rather than bare potentialities for all possible interests, even though the specific character of these interests is unknown to them. They must try to secure favorable conditions for advancing these definite ends, whatever they are (§28). The hierarchy of interests and its relation to the priority of liberty is taken up later (§§39, 82), but the general nature of the argument for the basic liberties is illustrated by the case of liberty of conscience and freedom of thought (§§33–35).

In addition, the veil of ignorance (§24) is interpreted to mean not only that the parties have no knowledge of their particular aims and ends (except what is contained in the thin theory of the good), but also that the historical record is closed to them. They do not know, and cannot enumerate, the social circumstances in which they may find themselves, or the array of techniques their society may have at its disposal. They have, therefore, no objective grounds for relying on one probability distribution rather than another, and the principle of insufficient reason cannot be invoked as a way around this limitation. These considerations, together with those derived from regarding the parties as having determinate fundamental interests, imply that the expectation constructed by the argument for the utility principle is unsound and lacks the necessary unity (§28).

30. CLASSICAL UTILITARIANISM, IMPARTIALITY, AND BENEVOLENCE

I now want to compare classical utilitarianism with the two principles of justice. As we have seen, the parties in the original position would reject the classical principle in favor of that of maximizing average utility. Since they are concerned to advance their own interests, they have no desire to maximize the total (or the net balance) of satisfactions. For similar reasons they would prefer the two principles of justice. From a contractarian point of view, then, the classical principle ranks below both of these alter-

natives. It must, therefore, have an entirely different derivation, for it is historically the most important form of utilitarianism. The great utilitarians who espoused it were certainly under no misapprehension that it would be chosen in what I have called the original position. Some of them, particularly Sidgwick, clearly recognized the average principle as an alternative and rejected it.[33] Since the classical view is closely related to the concept of the impartial sympathetic spectator, I shall look at this concept in order to clarify the intuitive basis of the traditional doctrine.

Consider the following definition reminiscent of Hume and Adam Smith. Something is right, a social system say, when an ideally rational and impartial spectator would approve of it from a general point of view should he possess all the relevant knowledge of the circumstances. A rightly ordered society is one meeting the approval of such an ideal observer.[34] Now there may be several problems with this definition, for example, whether the notions of approval and relevant knowledge can be specified without circularity. But I shall leave these questions aside. The essential point here is that there is no conflict so far between this definition and justice as fairness. For suppose we define the concept of right by saying that something is right if and only if it satisfies the principles which would be chosen in the original position to apply to things of its kind. It may well be the case that an ideally rational and impartial spectator would approve of a social system if and only if it satisfies the principles of justice which would be adopted in the contract scheme. The definitions may both be true of the same things. This possibility is not ruled out by the ideal observer definition. Since this definition makes no specific psychological assumptions about the impartial spectator, it yields no principles to account for his approvals under ideal conditions. One who accepts this definition is free to accept justice as fairness for this purpose: one simply allows that an ideal observer would approve of social systems to the extent that they satisfy the two principles of justice. There is an essential difference, then, between these two definitions of right. The impartial spectator definition makes no assumptions from which the prin-

33. *Methods of Ethics,* pp. 415f.

34. See Roderick Firth, "Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research,* vol. 12 (1952); and F. C. Sharp, *Good and Ill Will* (Chicago, University of Chicago Press, 1950), pp. 156–162. For Hume's account, see *Treatise of Human Nature,* ed. L. A. Selby-Bigge (Oxford, 1888), bk. III, pt. III, sec. I, especially pp. 574–584; and for Adam Smith, *The Theory of Moral Sentiments,* in L. A. Selby-Bigge, *British Moralists,* vol. I (Oxford, 1897), pp. 257–277. A general discussion is found in C. D. Broad, "Some Reflections on Moral-Sense Theories in Ethics," *Proceedings of the Aristotelian Society,* vol. 45 (1944–45). See also W. K. Kneale, "Objectivity in Morals," *Philosophy,* vol. 25 (1950).

ciples of right and justice may be derived.[35] It is designed instead to single out certain central features characteristic of moral discussion, the fact that we try to appeal to our considered judgments after conscientious reflection, and the like. The contractarian definition goes further: it attempts to provide a deductive basis for the principles that account for these judgments. The conditions of the initial situation and the motivation of the parties are intended to set out the necessary premises to achieve this end.

Now while it is possible to supplement the impartial spectator definition with the contract point of view, there are other ways of giving it a deductive basis. Thus suppose that the ideal observer is thought of as a perfectly sympathetic being. Then there is a natural derivation of the classical principle of utility along the following lines. An institution is right, let us say, if an ideally sympathetic and impartial spectator would approve of it more strongly than any other institution feasible in the circumstances. For simplicity we may assume, as Hume sometimes does, that approval is a special kind of pleasure which arises more or less intensely in contemplating the workings of institutions and their consequences for the happiness of those engaged in them. This special pleasure is the result of sympathy. In Hume's account it is quite literally a reproduction in our experience of the satisfactions and pleasures which we recognize to be felt by others.[36] Thus an impartial spectator experiences this pleasure in contemplating the social system in proportion to the net sum of pleasure felt by those affected by it. The strength of his approval corresponds to, or measures, the amount of satisfaction in the society surveyed. Therefore his expressions of approval will be given according to the classical principle of utility. To be sure, as Hume observes, sympathy is not a strong feeling. Not only is self-interest likely to inhibit the frame of mind in which we experience it, but self-interest tends to override its dictates in determining our actions. Yet when men do regard their institutions from a general point of view, Hume thought that sympathy is the only psychological principle at work, and it will at least guide our considered moral judgments. However weak sympathy may be, it nevertheless constitutes a common ground for bringing our moral opinions into

35. Thus Firth holds, for example, that an ideal observer has various general interests though not particular ones; and that these interests are indeed necessary if such an observer is to have any significant moral reactions. But nothing specific is said about the content of these interests that enables one to work out how the approvals and disapprovals of an ideal observer would be determined. See "Ethical Absolutism and the Ideal Observer," pp. 336–341.

36. See *A Treatise of Human Nature,* bk. II, pt. I, sec. XI, and bk. III, pt. I, sec. I, the first parts of each, and sec. VI. In the edition of L. A. Selby-Bigge, this is pp. 316–320, 575–580, and 618f.

agreement. Men's natural capacity for sympathy suitably generalized provides the perspective from which they can reach an understanding on a common conception of justice.

Thus we arrive at the following view. A rational and impartial sympathetic spectator is a person who takes up a general perspective: he assumes a position where his own interests are not at stake and he possesses all the requisite information and powers of reasoning. So situated he is equally sympathetic to the desires and satisfactions of everyone affected by the social system. Responding to the interests of each person in the same way, an impartial spectator gives free reign to his capacity for sympathetic identification by viewing each person's situation as it affects that person. Thus he imagines himself in the place of each person in turn, and when he has done this for everyone, the strength of his approval is determined by the balance of satisfactions to which he has sympathetically responded. When he has made the rounds of all the affected parties, so to speak, his approval expresses the total result. Sympathetically imagined pains cancel out sympathetically imagined pleasures, and the final intensity of approval corresponds to the net sum of positive feeling.

It is instructive to note a contrast between the features of the sympathetic spectator and the conditions defining the original position. The elements of the sympathetic spectator definition, impartiality, possession of relevant knowledge, and powers of imaginative identification, are to assure the complete and accurate response of natural sympathy. Impartiality prevents distortions of bias and self-interest; knowledge and the capacity for identification guarantee that the aspirations of others will be accurately appreciated. We can understand the point of the definition once we see that its parts are designed to give free scope to the operation of fellow feeling. In the original position, by contrast, the parties are mutually disinterested rather than sympathetic; but lacking knowledge of their natural assets or social situation, they are forced to view their arrangements in a general way. In the one case perfect knowledge and sympathetic identification result in a correct estimate of the net sum of satisfaction; in the other, mutual disinterestedness subject to a veil of ignorance leads to the two principles of justice.

Now, as I have mentioned, there is a sense in which classical utilitarianism fails to take seriously the distinction between persons (§5). The principle of rational choice for one man is taken as the principle of social choice as well. How does this view come about? It is the consequence, as we can now see, of wanting to give a deductive basis to an ideal observer definition of right, and of presuming that men's natural capacity for

sympathy provides the only means by which their moral judgments can be brought into agreement. The approvals of the impartial sympathetic spectator are adopted as the standard of justice, and this results in impersonality, in the conflation of all desires into one system of desire.[37]

From the standpoint of justice as fairness there is no reason why the persons in the original position would agree to the approvals of an impartial sympathetic spectator as the standard of justice. This agreement has all the drawbacks of the classical principle of utility to which it is equivalent. If, however, the parties are conceived as perfect altruists, that is, as persons whose desires conform to the approvals of such a spectator, then the classical principle would, of course, be adopted. The greater net balance of happiness with which to sympathize, the more a perfect altruist achieves his desire. Thus we arrive at the unexpected conclusion that while the average principle of utility is the ethic of a single rational

37. The most explicit and developed statement of this view I know of is that found in C. I. Lewis, *The Analysis of Knowledge and Valuation* (La Salle, Ill., Open Court Publishing Co., 1946). The whole of sec. 13 of ch. 18 is relevant here. Lewis says: "Value to more than one person is to be assessed as if their several experiences of value were included in that of a single person." Page 550. However, Lewis uses this idea to give an empirical account of social value; his theory of right is neither utilitarian nor empirical. J. J. C. Smart, in reply to the idea that fairness is a constraint on maximizing happiness, puts the point neatly when he asks: "if it is rational for me to choose the pain of a visit to the dentist in order to prevent the pain of toothache, why is it not rational of me to choose a pain for Jones, similar to that of my visit to the dentist, if that is the only way in which I can prevent a pain, equal to that of my toothache, for Robinson?" *An Outline of a System of Utilitarian Ethics,* p. 26. Another brief statement is in R. M. Hare, *Freedom and Reason* (Oxford, The Clarendon Press, 1963), p. 123.

Among the classical writers the conflation of all desires into one system is not to my knowledge clearly asserted. But it seems implicit in Edgeworth's comparison between "mécanique celeste" and "mécanique sociale" and in his idea that someday the latter may take its place with the former, both being founded upon one maximum principle, "the supreme pinnacle of moral as of physical science." He says: "As the movements of each particle, constrained or loose, in a material cosmos are continually subordinated to one maximum sum-total of accumulated energy, so the movements of each soul, whether selfishly isolated or linked sympathetically, may continually be realising the maximum energy of pleasure, the Divine love of the universe." *Mathematical Psychics,* p. 12. Sidgwick is always more restrained and there are only hints of the doctrine in *The Methods of Ethics.* Thus at one point he may be read to say that the notion of universal good is constructed from the goods of different individuals in the same way as the good (on the whole) of a single individual is constructed from the different goods that succeed one another in the temporal series of his conscious states (p. 382). This interpretation is confirmed by his saying later: "If, then, when any one hypothetically concentrates his attention on himself, Good is naturally and almost inevitably conceived to be pleasure, we may reasonably conclude that the Good of any number of similar beings, whatever their mutual relations may be, cannot be essentially different in quality." Page 405. Sidgwick also believed that the axiom of rational prudence is no less problematical than that of rational benevolence. We can equally well ask why we should concern ourselves about our own future feelings as about the feelings of other persons. Pages 418f. Presumably he thought the answer identical in each case: it is necessary to achieve the greatest sum of satisfaction. These remarks seem to suggest the conflation view.

individual (with no aversion to risk) who tries to maximize his own prospects, the classical doctrine is the ethic of perfect altruists. A surprising contrast indeed! By looking at these principles from the standpoint of the original position, we see that a different complex of ideas underlies them. Not only are they based upon contrary motivational assumptions, but the notion of taking chances has a part in one view yet none in the other. In the classical conception one chooses as if one will for certain live through the experiences of each individual, seriatim as Lewis says, and then sum up the result.[38] The idea of taking a chance on which person one will turn out to be does not arise. Thus even if the concept of the original position served no other purpose, it would be a useful analytic device. Although the various principles of utility may often have similar practical consequences, we can see that these conceptions derive from markedly distinct assumptions.

There is, however, a peculiar feature of perfect altruism that deserves mention. A perfect altruist can fulfill his desire only if someone else has independent, or first-order, desires. To illustrate this fact, suppose that in deciding what to do all vote to do what everyone else wants to do. Obviously nothing gets settled; in fact, there is nothing to decide. For a problem of justice to arise at least two persons must want to do something other than whatever everyone else wants to do. It is impossible, then, to assume that the parties are simply perfect altruists. They must have some separate interests which may conflict. Justice as fairness models this conflict by the assumption of mutual disinterest in the original position. While this may prove to be an oversimplification, one can develop a reasonably comprehensive conception of justice on this basis.

Some philosophers have accepted the utilitarian principle because they believed that the idea of an impartial sympathetic spectator is the correct interpretation of impartiality. Indeed, Hume thought that it offered the only perspective from which moral judgments could be made coherent and brought into line. Now moral judgments should be impartial; but there is another way to achieve this. An impartial judgment, we can say, is one rendered in accordance with the principles which would be chosen in the original position. An impartial person is one whose situation and character enable him to judge in accordance with these principles without bias or prejudice. Instead of defining impartiality from the standpoint of a sympathetic observer, we define impartiality from the standpoint of the litigants themselves. It is they who must choose their conception of jus-

38. See *The Analysis of Knowledge and Valuation,* p. 547.

tice once and for all in an original position of equality. They must decide by which principles their claims against one another are to be settled, and he who is to judge between men serves as their agent. The fault of the utilitarian doctrine is that it mistakes impersonality for impartiality.

The preceding remarks naturally lead one to ask what sort of theory of justice would result if one adopted the sympathetic spectator idea but did not characterize this spectator as conflating all desires into one system. Hume's conception provides one modus operandi for benevolence, but is it the only possibility? Now love clearly has among its main elements the desire to advance the other person's good as this person's rational self-love would require. Very often how one is to realize this desire is clear enough. The difficulty is that the love of several persons is thrown into confusion once the claims of these persons conflict. If we reject the classical doctrine, what does the love of mankind enjoin? It is quite pointless to say that one is to judge the situation as benevolence dictates. This assumes that we are wrongly swayed by self-concern. Our problem lies elsewhere. Benevolence is at sea as long as its many loves are in opposition in the persons of its many objects.

We might try out here the idea that a benevolent person is to be guided by the principles someone would choose if he knew that he is to split, so to speak, into the many members of society.[39] That is, he is to imagine that he is to divide into a plurality of persons whose life and experiences will be distinct in the usual way. Experiences and memories are to remain each person's own; and there is to be no conflation of desires and memories into those of one person. Since a single individual is literally to become many persons, there is no question of guessing which one; once again the problem of taking chances does not arise. Now knowing this (or believing it), which conception of justice would a person choose for a society comprised of these individuals? As this person would, let us suppose, love this plurality of persons as he loves himself, perhaps the principles he would choose characterize the aims of benevolence.

Leaving aside the difficulties in the idea of splitting that may arise from problems about personal identity, two things seem evident. First of all, it is still unclear what a person would decide, since the situation does not offhand provide an answer. But secondly, the two principles of justice now seem a relatively more plausible choice than the classical principle

39. This idea is found in Thomas Nagel, *The Possibility of Altruism* (Oxford, The Clarendon Press, 1970), pp. 140f.

of utility. The latter is no longer the natural preference, and this suggests that the conflation of persons into one is indeed at the root of the classical view. The reason why the situation remains obscure is that love and benevolence are second-order notions: they seek to further the good of beloved individuals that is already given. If the claims of these goods clash, benevolence is at a loss as to how to proceed, as long anyway as it treats these individuals as separate persons. These higher-order sentiments do not include principles of right to adjudicate these conflicts. Therefore a love of mankind that wishes to preserve the distinction of persons, to recognize the separateness of life and experience, will use the two principles of justice to determine its aims when the many goods it cherishes are in opposition. This is simply to say that this love is guided by what individuals themselves would consent to in a fair initial situation which gives them equal representation as moral persons. We now see why nothing would have been gained by attributing benevolence to the parties in the original position.

We must, however, distinguish between the love of mankind and the sense of justice. The difference is not that they are guided by different principles, since both include a desire to give justice. Rather, the former is manifest by the greater intensity and pervasiveness of this desire, and in a readiness to fulfill all the natural duties in addition to that of justice, and even to go beyond their requirements. The love of mankind is more comprehensive than the sense of justice and prompts to acts of supererogation, whereas the latter does not. Thus we see that the assumption of the mutual disinterestedness of the parties does not prevent a reasonable interpretation of benevolence and of the love of mankind within the framework of justice as fairness. The fact that we start out assuming that the parties are mutually disinterested and have conflicting first-order desires still allows us to construct a comprehensive account. For once the principles of right and justice are on hand, they may be used to define the moral virtues just as in any other theory. The virtues are sentiments, that is, related families of dispositions and propensities regulated by a higher-order desire, in this case a desire to act from the corresponding moral principles. Although justice as fairness begins by taking the persons in the original position as individuals, or more accurately as continuing strands, this is no obstacle to explicating the higher-order moral sentiments that serve to bind a community of persons together. In Part Three I shall return to these matters.

These remarks conclude the theoretical part of our discussion. I shall

make no attempt to summarize this long chapter. Having set out the initial arguments in favor of the two principles of justice over the two forms of utility, it is time to see how these principles apply to institutions and how well they seem to match our considered judgments. Only in this way can we become clearer about their meaning and find out whether they are an improvement over other conceptions.

PART TWO. INSTITUTIONS

CHAPTER IV. EQUAL LIBERTY

In the three chapters of Part Two my aim is to illustrate the content of the principles of justice. I shall do this by describing a basic structure that satisfies these principles and by examining the duties and obligations to which they give rise. The main institutions of this structure are those of a constitutional democracy. I do not argue that these arrangements are the only ones that are just. Rather my intention is to show that the principles of justice, which so far have been discussed in abstraction from institutional forms, define a workable political conception, and are a reasonable approximation to and extension of our considered judgments. In this chapter I begin by setting out a four-stage sequence that clarifies how the principles for institutions are to be applied. Two parts of the basic structure are briefly described and the concept of liberty defined. After this, three problems of equal liberty are discussed: equal liberty of conscience, political justice and equal political rights, and equal liberty of the person and its relation to the rule of law. I then take up the meaning of the priority of liberty, and conclude with a brief account of the Kantian interpretation of the original position.

31. THE FOUR-STAGE SEQUENCE

It is evident that some sort of framework is needed to simplify the application of the two principles of justice. For consider three kinds of judgments that a citizen has to make. First of all, he must judge the justice of legislation and social policies. But he also knows that his opinions will not always coincide with those of others, since men's judgments and beliefs are likely to differ especially when their interests are engaged. Therefore secondly, a citizen must decide which constitutional arrangements are just for reconciling conflicting opinions of justice. We may think of the political process as a machine which makes social decisions

when the views of representatives and their constituents are fed into it. A citizen will regard some ways of designing this machine as more just than others. So a complete conception of justice is not only able to assess laws and policies but it can also rank procedures for selecting which political opinion is to be enacted into law. There is still a third problem. The citizen accepts a certain constitution as just, and he thinks that certain traditional procedures are appropriate, for example, the procedure of majority rule duly circumscribed. Yet since the political process is at best one of imperfect procedural justice, he must ascertain when the enactments of the majority are to be complied with and when they can be rejected as no longer binding. In short, he must be able to determine the grounds and limits of political duty and obligation. Thus a theory of justice has to deal with at least three types of questions, and this indicates that it may be useful to think of the principles as applied in a several-stage sequence.

At this point, then, I introduce an elaboration of the original position. So far I have supposed that once the principles of justice are chosen the parties return to their place in society and henceforth judge their claims on the social system by these principles. But if several intermediate stages are imagined to take place in a definite sequence, this sequence may give us a schema for sorting out the complications that must be faced. Each stage is to represent an appropriate point of view from which certain kinds of questions are considered.[1] Thus I suppose that after the parties have adopted the principles of justice in the original position, they move to a constitutional convention. Here they are to decide upon the justice of political forms and choose a constitution: they are delegates, so to speak, to such a convention. Subject to the constraints of the principles of justice already chosen, they are to design a system for the constitutional powers of government and the basic rights of citizens. It is at this stage that they weigh the justice of procedures for coping with diverse political views. Since the appropriate conception of justice has been agreed upon, the veil of ignorance is partially lifted. The persons in the convention have, of course, no information about particular individuals: they do not know their own social position, their place in the distribution of natural attributes, or their conception of the good. But in addition to an understanding of the principles of social theory, they now know the relevant general facts about their society, that is, its natural circumstances and

1. The idea of a four-stage sequence is suggested by the United States Constitution and its history. For some remarks as to how this sequence might be interpreted theoretically and related to procedural justice, see K. J. Arrow, *Social Choice and Individual Values,* 2nd ed. (New York, John Wiley and Sons, 1963), pp. 89–91.

resources, its level of economic advance and political culture, and so on. They are no longer limited to the information implicit in the circumstances of justice. Given their theoretical knowledge and the appropriate general facts about their society, they are to choose the most effective just constitution, the constitution that satisfies the principles of justice and is best calculated to lead to just and effective legislation.[2]

At this point we need to distinguish two problems. Ideally a just constitution would be a just procedure arranged to insure a just outcome. The procedure would be the political process governed by the constitution, the outcome the body of enacted legislation, while the principles of justice would define an independent criterion for both procedure and outcome. In pursuit of this ideal of perfect procedural justice (§14), the first problem is to design a just procedure. To do this the liberties of equal citizenship must be incorporated into and protected by the constitution. These liberties include those of liberty of conscience and freedom of thought, liberty of the person, and equal political rights. The political system, which I assume to be some form of constitutional democracy, would not be a just procedure if it did not embody these liberties.

Clearly any feasible political procedure may yield an unjust outcome. In fact, there is no scheme of procedural political rules which guarantees that unjust legislation will not be enacted. In the case of a constitutional regime, or indeed of any political form, the ideal of perfect procedural justice cannot be realized. The best attainable scheme is one of imperfect procedural justice. Nevertheless some schemes have a greater tendency than others to result in unjust laws. The second problem, then, is to select from among the procedural arrangements that are both just and feasible those which are most likely to lead to a just and effective legal order. Once again this is Bentham's problem of the artificial identification of interests, only here the rules (just procedure) are to be framed to give legislation (just outcome) likely to accord with the principles of justice rather than the principle of utility. To solve this problem intelligently

2. It is important to distinguish the four-stage sequence and its conception of a constitutional convention from the kind of view of constitutional choice found in social theory and exemplified by J. M. Buchanan and Gordon Tullock, *The Calculus of Consent* (Ann Arbor, University of Michigan Press, 1963). The idea of the four-stage sequence is part of a moral theory, and does not belong to an account of the working of actual constitutions, except insofar as political agents are influenced by the conception of justice in question. In the contract doctrine, the principles of justice have already been agreed to, and our problem is to formulate a schema that will assist us in applying them. The aim is to characterize a just constitution and not to ascertain which sort of constitution would be adopted, or acquiesced in, under more or less realistic (though simplified) assumptions about political life, much less on individualistic assumptions of the kind characteristic of economic theory.

requires a knowledge of the beliefs and interests that men in the system are liable to have and of the political tactics that they will find it rational to use given their circumstances. The delegates are assumed, then, to know these things. Provided they have no information about particular individuals including themselves, the idea of the original position is not affected.

In framing a just constitution I assume that the two principles of justice already chosen define an independent standard of the desired outcome. If there is no such standard, the problem of constitutional design is not well posed, for this decision is made by running through the feasible just constitutions (given, say, by enumeration on the basis of social theory) looking for the one that in the existing circumstances will most probably result in effective and just social arrangements. Now at this point we come to the legislative stage, to take the next step in the sequence. The justice of laws and policies is to be assessed from this perspective. Proposed bills are judged from the position of a representative legislator who, as always, does not know the particulars about himself. Statutes must satisfy not only the principles of justice but whatever limits are laid down in the constitution. By moving back and forth between the stages of the constitutional convention and the legislature, the best constitution is found.

Now the question whether legislation is just or unjust, especially in connection with economic and social policies, is commonly subject to reasonable differences of opinion. In these cases judgment frequently depends upon speculative political and economic doctrines and upon social theory generally. Often the best that we can say of a law or policy is that it is at least not clearly unjust. The application of the difference principle in a precise way normally requires more information than we can expect to have and, in any case, more than the application of the first principle. It is often perfectly plain and evident when the equal liberties are violated. These violations are not only unjust but can be clearly seen to be unjust: the injustice is manifest in the public structure of institutions. But this state of affairs is comparatively rare with social and economic policies regulated by the difference principle.

I imagine then a division of labor between stages in which each deals with different questions of social justice. This division roughly corresponds to the two parts of the basic structure. The first principle of equal liberty is the primary standard for the constitutional convention. Its main requirements are that the fundamental liberties of the person and liberty of conscience and freedom of thought be protected and that the political

process as a whole be a just procedure. Thus the constitution establishes a secure common status of equal citizenship and realizes political justice. The second principle comes into play at the stage of the legislature. It dictates that social and economic policies be aimed at maximizing the long-term expectations of the least advantaged under conditions of fair equality of opportunity, subject to the equal liberties being maintained. At this point the full range of general economic and social facts is brought to bear. The second part of the basic structure contains the distinctions and hierarchies of political, economic, and social forms which are necessary for efficient and mutually beneficial social cooperation. Thus the priority of the first principle of justice to the second is reflected in the priority of the constitutional convention to the legislative stage.

The last stage is that of the application of rules to particular cases by judges and administrators, and the following of rules by citizens generally. At this stage everyone has complete access to all the facts. No limits on knowledge remain since the full system of rules has now been adopted and applies to persons in virtue of their characteristics and circumstances. However, it is not from this standpoint that we are to decide the grounds and limits of political duty and obligation. This third type of problem belongs to partial compliance theory, and its principles are discussed from the point of view of the original position after those of ideal theory have been chosen (§39). Once these are on hand, we can view our particular situation from the perspective of the last stage, as for example in the cases of civil disobedience and conscientious refusal (§§57–59).

The availability of knowledge in the four-stage sequence is roughly as follows. Let us distinguish between three kinds of facts: the first principles of social theory (and other theories when relevant) and their consequences; general facts about society, such as its size and level of economic advance, its institutional structure and natural environment, and so on; and finally, particular facts about individuals such as their social position, natural attributes, and peculiar interests. In the original position the only particular facts known to the parties are those that can be inferred from the circumstances of justice. While they know the first principles of social theory, the course of history is closed to them; they have no information about how often society has taken this or that form, or which kinds of societies presently exist. In the next stages, however, the general facts about their society are made available to them but not the particularities of their own condition. Limitations on knowledge can be relaxed since the principles of justice are already chosen. The flow of information is determined at each stage by what is required in order to apply these

principles intelligently to the kind of question of justice at hand, while at the same time any knowledge that is likely to give rise to bias and distortion and to set men against one another is ruled out. The notion of the rational and impartial application of principles defines the kind of knowledge that is admissible. At the last stage, clearly, there are no reasons for the veil of ignorance in any form, and all restrictions are lifted.

It is essential to keep in mind that the four-stage sequence is a device for applying the principles of justice. This scheme is part of the theory of justice as fairness and not an account of how constitutional conventions and legislatures actually proceed. It sets out a series of points of view from which the different problems of justice are to be settled, each point of view inheriting the constraints adopted at the preceding stages. Thus a just constitution is one that rational delegates subject to the restrictions of the second stage would adopt for their society. And similarly just laws and policies are those that would be enacted at the legislative stage. Of course, this test is often indeterminate: it is not always clear which of several constitutions, or economic and social arrangements, would be chosen. But when this is so, justice is to that extent likewise indeterminate. Institutions within the permitted range are equally just, meaning that they could be chosen; they are compatible with all the constraints of the theory. Thus on many questions of social and economic policy we must fall back upon a notion of quasi-pure procedural justice: laws and policies are just provided that they lie within the allowed range, and the legislature, in ways authorized by a just constitution, has in fact enacted them. This indeterminacy in the theory of justice is not in itself a defect. It is what we should expect. Justice as fairness will prove a worthwhile theory if it defines the range of justice more in accordance with our considered judgments than do existing theories, and if it singles out with greater sharpness the graver wrongs a society should avoid.

32. THE CONCEPT OF LIBERTY

In discussing the application of the first principle of justice I shall try to bypass the dispute about the meaning of liberty that has so often troubled this topic. The controversy between the proponents of negative and positive liberty as to how freedom should be defined is one I shall leave aside. I believe that for the most part this debate is not concerned with definitions at all, but rather with the relative values of the several liberties when they come into conflict. Thus one might want to maintain, as Constant

did, that the so-called liberty of the moderns is of greater value than the liberty of the ancients. While both sorts of freedom are deeply rooted in human aspirations, freedom of thought and liberty of conscience, freedom of the person and the civil liberties, ought not to be sacrificed to political liberty, to the freedom to participate equally in political affairs.[3] This question is clearly one of substantive political philosophy, and a theory of right and justice is required to answer it. Questions of definition can have at best but an ancillary role.

Therefore I shall simply assume that any liberty can be explained by a reference to three items: the agents who are free, the restrictions or limitations which they are free from, and what it is that they are free to do or not to do. Complete explanations of liberty provide the relevant information about these three things.[4] Very often certain matters are clear from the context and a full explanation is unnecessary. The general description of a liberty, then, has the following form: this or that person (or persons) is free (or not free) from this or that constraint (or set of constraints) to do (or not to do) so and so. Associations as well as natural persons may be free or not free, and constraints may range from duties and prohibitions defined by law to the coercive influences arising from public opinion and social pressure. For the most part I shall discuss liberty in connection with constitutional and legal restrictions. In these cases liberty is a certain structure of institutions, a certain system of public rules defining rights and duties. Set in this background, persons are at liberty to do something when they are free from certain constraints either to do it or not to do it and when their doing it or not doing it is protected from interference by other persons. If, for example, we consider liberty of conscience as defined by law, then individuals have this basic liberty when they are free to pursue their moral, philosophical, or religious interests without legal restrictions requiring them to engage or not to engage in any particular form of religious or other practice, and when other men have a legal duty not to interfere. A rather intricate complex of rights and duties characterizes any particular basic liberty. Not only must it be permissible for

3. See Constant's essay *Ancient and Modern Liberty* (1819). His ideas on this are discussed by Guido de Ruggiero, *The History of European Liberalism,* trans. R. G. Collingwood (Oxford, The Clarendon Press, 1927), pp. 159–164, 167–169. For a general discussion, see Isaiah Berlin, *Four Essays on Liberty* (London, Oxford University Press, 1969), esp. the third essay and pp. xxxvii–lxiii of the introduction; and G. G. MacCallum, "Negative and Positive Freedom," *Philosophical Review,* vol. 76 (1967).

4. Here I follow MacCallum, "Negative and Positive Freedom." See further Felix Oppenheim, *Dimensions of Freedom* (New York, St. Martin's Press, 1961), esp. pp. 109–118, 132–134, where a notion of social freedom is also triadically defined.

individuals to do or not to do something, but government and other persons must have a legal duty not to obstruct. I shall not delineate these rights and duties in any detail, but shall suppose that we understand their nature well enough for our purposes.

Several brief clarifications. First of all, one must keep in mind that the basic liberties are to be assessed as a whole, as one system. The worth of one such liberty normally depends upon the specification of the other liberties. Second, I assume that under reasonably favorable conditions there is always a way of defining these liberties so that the most central applications of each can be simultaneously secured and the most fundamental interests protected. Or at least that this is possible provided the two principles and their associated priorities are consistently adhered to. Finally, given such a specification of the basic liberties, it is assumed to be clear for the most part whether an institution or law actually restricts a basic liberty or merely regulates it. For example, certain rules of order are necessary for regulating discussion; without the acceptance of reasonable procedures of inquiry and debate, freedom of speech loses its value. On the other hand, a prohibition against holding or arguing for certain religious, moral, or political views is a restriction of liberty and must be judged accordingly.[5] Thus as delegates in a constitutional convention, or as members of a legislature, the parties must decide how the various liberties are to be specified so as to give the best total system of liberty. They must note the distinction between regulation and restriction, but at many points they will have to balance one basic liberty against another; for example, freedom of speech against the right to a fair trial. The best arrangement of the several liberties depends upon the totality of limitations to which they are subject.

While the equal liberties may, therefore, be restricted, these limits are subject to certain criteria expressed by the meaning of equal liberty and the serial order of the two principles of justice. Offhand there are two ways of contravening the first principle. Liberty is unequal as when one class of persons has a greater liberty than another, or liberty is less extensive than it should be. Now all the liberties of equal citizenship must be the same for each member of society. Nevertheless some of the equal liberties may be more extensive than others, assuming that their extensions can be compared. More realistically, if it is supposed that at best each liberty can be measured on its own scale, then the various liberties

5. See Alexander Meiklejohn, *Free Speech and Its Relation to Self-Government* (New York, Harper and Brothers, 1948), ch. I, sec. 6.

can be broadened or narrowed according to how they affect one another. A basic liberty covered by the first principle can be limited only for the sake of liberty itself, that is, only to insure that the same liberty or a different basic liberty is properly protected and to adjust the one system of liberties in the best way. The adjustment of the complete scheme of liberty depends solely upon the definition and extent of the particular liberties. Of course, this scheme is always to be assessed from the standpoint of the representative equal citizen. From the perspective of the constitutional convention or the legislative stage (as appropriate) we are to ask which system it would be rational for him to prefer.

A final point. The inability to take advantage of one's rights and opportunities as a result of poverty and ignorance, and a lack of means generally, is sometimes counted among the constraints definitive of liberty. I shall not, however, say this, but rather I shall think of these things as affecting the worth of liberty, the value to individuals of the rights that the first principle defines. With this understanding, and assuming that the total system of basic liberty is drawn up in the manner just explained, we may note that the two-part basic structure allows a reconciliation of liberty and equality. Thus liberty and the worth of liberty are distinguished as follows: liberty is represented by the complete system of the liberties of equal citizenship, while the worth of liberty to persons and groups depends upon their capacity to advance their ends within the framework the system defines. Freedom as equal liberty is the same for all; the question of compensating for a lesser than equal liberty does not arise. But the worth of liberty is not the same for everyone. Some have greater authority and wealth, and therefore greater means to achieve their aims. The lesser worth of liberty is, however, compensated for, since the capacity of the less fortunate members of society to achieve their aims would be even less were they not to accept the existing inequalities whenever the difference principle is satisfied. But compensating for the lesser worth of freedom is not to be confused with making good an unequal liberty. Taking the two principles together, the basic structure is to be arranged to maximize the worth to the least advantaged of the complete scheme of equal liberty shared by all. This defines the end of social justice.

These remarks about the concept of liberty are unhappily abstract. At this stage it would serve no purpose to classify systematically the various liberties. Instead I shall assume that we have a clear enough idea of the distinctions between them, and that in the course of taking up various cases these matters will gradually fall into place. In the next sections I

discuss the first principle of justice in connection with liberty of conscience and freedom of thought, political liberty, and liberty of the person as protected by the rule of law. These applications provide an occasion to clarify the meaning of the equal liberties and to present further grounds for the first principle. Moreover, each case illustrates the use of the criteria for limiting and adjusting the various freedoms and thereby exemplifies the meaning of the priority of liberty. It must, however, be emphasized that the account of the basic liberties is not offered as a precise criterion that determines when we are justified in restricting a liberty, whether basic or otherwise. There is no way to avoid some reliance on our sense of balance and judgment. As always, the aim is to formulate a conception of justice that, however much it may call upon our intuitive capacities, helps to make our considered judgments of justice converge (§8). The various priority rules are to further this end by singling out certain fundamental structural features of one's moral view.

33. EQUAL LIBERTY OF CONSCIENCE

In the preceding chapter I remarked that one of the attractive features of the principles of justice is that they guarantee a secure protection for the equal liberties. In the next several sections I wish to examine the argument for the first principle in more detail by considering the grounds for freedom of conscience.[6] So far, while it has been supposed that the parties represent continuing lines of claims and care for their immediate descendants, this feature has not been stressed. Nor have I emphasized that the parties must assume that they may have moral, religious, or philosophical interests which they cannot put in jeopardy unless there is no alternative. One might say that they regard themselves as having moral or religious

6. The notion of equal right is, of course, well known in one form or another and appears in numerous analyses of justice even where the writers differ widely on other matters. Thus if the principle of an equal right to freedom is commonly associated with Kant—see *The Metaphysical Elements of Justice,* trans. John Ladd (New York, The Library of Liberal Arts, 1965), pp. 43–45—it may be claimed that it can also be found in J. S. Mill's *On Liberty* and elsewhere in his writings, and in those of many other liberal thinkers. H. L. A. Hart has argued for something like it in "Are There Any Natural Rights?" *Philosophical Review,* vol. 64 (1955); and similarly Richard Wollheim in the symposium "Equality," *Proceedings of the Aristotelian Society,* vol. 56 (1955–1956). The principle of equal liberty as I shall use it may acquire, though, special features in view of the theory of which it is a part. In particular, it enjoins a certain structure of institutions to be departed from only as the priority rules allow (§39). It is far removed from a principle of equal consideration, since the intuitive idea is to generalize the principle of religious toleration to a social form, thereby arriving at equal liberty in public institutions.

obligations which they must keep themselves free to honor. Of course, from the standpoint of justice as fairness, these obligations are self-imposed; they are not bonds laid down by this conception of justice. The point is rather that the persons in the original position are not to view themselves as single isolated individuals. To the contrary, they assume that they have interests which they must protect as best they can and that they have ties with certain members of the next generation who will also make similar claims. Once the parties consider these matters, the case for the principles of justice is very much strengthened, as I shall now try to show.

The question of equal liberty of conscience is settled. It is one of the fixed points of our considered judgments of justice. But precisely because of this fact it illustrates the nature of the argument for the principle of equal liberty. The reasoning in this case can be generalized to apply to other freedoms, although not always with the same force. Turning then to liberty of conscience, it seems evident that the parties must choose principles that secure the integrity of their religious and moral freedom. They do not know, of course, what their religious or moral convictions are, or what is the particular content of their moral or religious obligations as they interpret them. Indeed, they do not know that they think of themselves as having such obligations. The possibility that they do suffices for the argument, although I shall make the stronger assumption. Further, the parties do not know how their religious or moral view fares in their society, whether, for example, it is in the majority or the minority. All they know is that they have obligations which they interpret in this way. The question they are to decide is which principle they should adopt to regulate the liberties of citizens in regard to their fundamental religious, moral, and philosophical interests.

Now it seems that equal liberty of conscience is the only principle that the persons in the original position can acknowledge. They cannot take chances with their liberty by permitting the dominant religious or moral doctrine to persecute or to suppress others if it wishes. Even granting (what may be questioned) that it is more probable than not that one will turn out to belong to the majority (if a majority exists), to gamble in this way would show that one did not take one's religious or moral convictions seriously, or highly value the liberty to examine one's beliefs. Nor on the other hand, could the parties consent to the principle of utility. In this case their freedom would be subject to the calculus of social interests and they would be authorizing its restriction if this would lead to a greater net balance of satisfaction. Of course, as we have seen, a utilitarian may

try to argue from the general facts of social life that when properly carried out the computation of advantages never justifies such limitations, at least under reasonably favorable conditions of culture. But even if the parties were persuaded of this, they might as well guarantee their freedom straightway by adopting the principle of equal liberty. There is nothing gained by not doing so, and to the extent that the outcome of the actuarial calculation is unclear a great deal may be lost. Indeed, if we give a realistic interpretation to the general knowledge available to the parties (see the end of §26), they are forced to reject the utilitarian principle. These considerations have all the more force in view of the complexity and vagueness of these calculations (if we can so describe them) as they are bound to be made in practice.

Moreover, the initial agreement on the principle of equal liberty is final. An individual recognizing religious and moral obligations regards them as binding absolutely in the sense that he cannot qualify his fulfillment of them for the sake of greater means for promoting his other interests. Greater economic and social benefits are not a sufficient reason for accepting less than an equal liberty. It seems possible to consent to an unequal liberty only if there is a threat of coercion which it is unwise to resist from the standpoint of liberty itself. For example, the situation may be one in which a person's religion or his moral view will be tolerated provided that he does not protest, whereas claiming an equal liberty will bring greater repression that cannot be effectively opposed. But from the perspective of the original position there is no way of ascertaining the relative strength of various doctrines and so these considerations do not arise. The veil of ignorance leads to an agreement on the principle of equal liberty; and the strength of religious and moral obligations as men interpret them seems to require that the two principles be put in serial order, at least when applied to freedom of conscience.

It may be said against the principle of equal liberty that religious sects, say, cannot acknowledge any principle at all for limiting their claims on one another. The duty to religious and divine law being absolute, no understanding among persons of different faiths is permissible from a religious point of view. Certainly men have often acted as if they held this doctrine. It is unnecessary, however, to argue against it. It suffices that if any principle can be agreed to, it must be that of equal liberty. A person may indeed think that others ought to recognize the same beliefs and first principles that he does, and that by not doing so they are grievously in error and miss the way to their salvation. But an understanding of reli-

gious obligation and of philosophical and moral first principles shows that we cannot expect others to acquiesce in an inferior liberty. Much less can we ask them to recognize us as the proper interpreter of their religious duties or moral obligations.

We should now observe that these reasons for the first principle receive further support once the parties' concern for the next generation is taken into account. Since they have a desire to obtain similar liberties for their descendants, and these liberties are also secured by the principle of equal liberty, there is no conflict of interests between generations. Moreover, the next generation could object to the choice of this principle only if the prospects offered by some other conception, say that of utility or perfection, were so attractive that the persons in the original position must not have properly considered their descendants when they rejected it. We can express this by noting that were a father, for example, to assert that he would accept the principle of equal liberty, a son could not object that were he (the father) to do so he would be neglecting his (the son's) interests. The advantages of the other principles are not this great and appear in fact uncertain and conjectural. The father could reply that when the choice of principles affects the liberty of others, the decision must, if possible, seem reasonable and responsible to them once they come of age. Those who care for others must choose for them in the light of what they will want whatever else they want once they reach maturity. Therefore following the account of primary goods, the parties presume that their descendants will want their liberty protected.

At this point we touch upon the principle of paternalism that is to guide decisions taken on behalf of others (§39). We must choose for others as we have reason to believe they would choose for themselves if they were at the age of reason and deciding rationally. Trustees, guardians, and benefactors are to act in this way, but since they usually know the situation and interests of their wards and beneficiaries, they can often make accurate estimates as to what is or will be wanted. The persons in the original position, however, are prevented from knowing any more about their descendants than they do about themselves, and so in this case too they must rely upon the theory of primary goods. Thus the father can say that he would be irresponsible if he were not to guarantee the rights of his descendants by adopting the principle of equal liberty. From the perspective of the original position, he must assume that this is what they will come to recognize as for their good.

I have tried to show, by taking liberty of conscience as an example,

how justice as fairness provides strong arguments for equal liberty. The same kind of reasoning applies, I believe, in other cases, though it is not always so convincing. I do not deny, however, that persuasive arguments for liberty are forthcoming on other views. As understood by Mill, the principle of utility often supports freedom. Mill defines the concept of value by reference to the interests of man as a progressive being. By this idea he means the interests men would have and the activities they would rather pursue under conditions encouraging freedom of choice. He adopts, in effect, a choice criterion of value: one activity is better than another if it is preferred by those who are capable of both and who have experienced each of them under circumstances of liberty.[7]

Using this principle Mill adduces essentially three grounds for free institutions. For one thing, they are required to develop men's capacities and powers, to arouse strong and vigorous natures. Unless their abilities are intensely cultivated and their natures enlivened, men will not be able to engage in and to experience the valuable activities of which they are capable. Secondly, the institutions of liberty and the opportunity for experience which they allow are necessary, at least to some degree, if men's preferences among different activities are to be rational and informed. Human beings have no other way of knowing what things they can do and which of them are most rewarding. Thus if the pursuit of value, estimated in terms of the progressive interests of mankind, is to be rational, that is, guided by a knowledge of human capacities and well-formed preferences, certain freedoms are indispensable. Otherwise society's attempt to follow the principle of utility proceeds blindly. The suppression of liberty is always likely to be irrational. Even if the general capacities of mankind were known (as they are not), each person has still to find himself, and for this freedom is a prerequisite. Finally, Mill believes that human beings prefer to live under institutions of liberty. Historical experience shows that men desire to be free whenever they have not resigned themselves to apathy and despair; whereas those who are free never want to abdicate their liberty. Although men may complain of the burdens of freedom and culture, they have an overriding desire to determine how they shall live and to settle their own affairs. Thus by Mill's choice criterion, free insti-

7. Mill's definition of utility as grounded on the permanent interests of man as a progressive being is in *On Liberty,* ch. I, par. 11. Originally I read the passage as "the permanent interests of a man," following a number of editions. I am grateful to David Spitz for telling me that Mill almost certainly wrote "man" and not "a man," and therefore the later variant, stemming from an early low-priced edition, is very probably a typesetter's error. I have revised the text accordingly. For the choice criterion of value, see *Utilitarianism,* ch. II, pars. 2–10. I heard this interpretation stated by G. A. Paul (1953) and am indebted to his remarks.

tutions have value in themselves as basic aspects of rationally preferred forms of life.[8]

These are certainly forceful arguments and under some circumstances anyway they might justify many if not most of the equal liberties. They clearly guarantee that in favorable conditions a considerable degree of liberty is a precondition of the rational pursuit of value. But even Mill's contentions, as cogent as they are, will not, it seems, justify an equal liberty for all. We still need analogues of familiar utilitarian assumptions. One must suppose a certain similarity among individuals, say their equal capacity for the activities and interests of men as progressive beings, and in addition a principle of the diminishing marginal value of basic rights when assigned to individuals. In the absence of these presumptions the advancement of human ends may be compatible with some persons' being oppressed, or at least granted but a restricted liberty. Whenever a society sets out to maximize the sum of intrinsic value or the net balance of the satisfaction of interests, it is liable to find that the denial of liberty for some is justified in the name of this single end. The liberties of equal citizenship are insecure when founded upon teleological principles. The argument for them relies upon precarious calculations as well as controversial and uncertain premises.

Moreover, nothing is gained by saying that persons are of equal intrinsic value unless this is simply a way of using the standard assumptions as if they were part of the principle of utility. That is, one applies this principle as if these assumptions were true. Doing this certainly has the merit of recognizing that we have more confidence in the principle of equal liberty than in the truth of the premises from which a perfectionist or utilitarian view would derive it. The grounds for this confidence, according to the contract view, is that the equal liberties have a different basis altogether. They are not a way of maximizing the sum of intrinsic value or of achieving the greatest net balance of satisfaction. The notion of maximizing a sum of value by adjusting the rights of individuals does not arise. Rather these rights are assigned to fulfill the principles of cooperation that citizens would acknowledge when each is fairly represented as a moral person. The conception defined by these principles is not that of maximizing anything, except in the vacuous sense of best meeting the requirements of justice, all things considered.

8. These three grounds are found in *On Liberty,* ch. III. They are not to be confused with the reasons Mill gives elsewhere, in ch. II for example, which urge the beneficial effects of free institutions.

34. TOLERATION AND THE COMMON INTEREST

Justice as fairness provides, as we have now seen, strong arguments for an equal liberty of conscience. I shall assume that these arguments can be generalized in suitable ways to support the principle of equal liberty. Therefore the parties have good grounds for adopting this principle. It is obvious that these considerations are also important in making the case for the priority of liberty. From the perspective of the constitutional convention these arguments lead to the choice of a regime guaranteeing moral liberty and freedom of thought and belief, and of religious practice, although these may be regulated as always by the state's interest in public order and security. The state can favor no particular religion and no penalties or disabilities may be attached to any religious affiliation or lack thereof. The notion of a confessional state is rejected. Instead, particular associations may be freely organized as their members wish, and they may have their own internal life and discipline subject to the restriction that their members have a real choice of whether to continue their affiliation. The law protects the right of sanctuary in the sense that apostasy is not recognized, much less penalized, as a legal offense, any more than is having no religion at all. In these ways the state upholds moral and religious liberty.

Liberty of conscience is limited, everyone agrees, by the common interest in public order and security. This limitation itself is readily derivable from the contract point of view. First of all, acceptance of this limitation does not imply that public interests are in any sense superior to moral and religious interests; nor does it require that government view religious matters as things indifferent or claim the right to suppress philosophical beliefs whenever they conflict with affairs of state. The government has no authority to render associations either legitimate or illegitimate any more than it has this authority in regard to art and science. These matters are simply not within its competence as defined by a just constitution. Rather, given the principles of justice, the state must be understood as the association consisting of equal citizens. It does not concern itself with philosophical and religious doctrine but regulates individuals' pursuit of their moral and spiritual interests in accordance with principles to which they themselves would agree in an initial situation of equality. By exercising its powers in this way the government acts as the citizens' agent and satisfies the demands of their public conception of justice. Therefore the notion of the omnicompetent laicist state is also denied, since from the principles of justice it follows that government has

neither the right nor the duty to do what it or a majority (or whatever) wants to do in questions of morals and religion. Its duty is limited to underwriting the conditions of equal moral and religious liberty.

Granting all this, it now seems evident that, in limiting liberty by reference to the common interest in public order and security, the government acts on a principle that would be chosen in the original position. For in this position each recognizes that the disruption of these conditions is a danger for the liberty of all. This follows once the maintenance of public order is understood as a necessary condition for everyone's achieving his ends whatever they are (provided they lie within certain limits) and for his fulfilling his interpretation of his moral and religious obligations. To restrain liberty of conscience at the boundary, however inexact, of the state's interest in public order is a limit derived from the principle of the common interest, that is, the interest of the representative equal citizen. The government's right to maintain public order and security is an enabling right, a right which the government must have if it is to carry out its duty of impartially supporting the conditions necessary for everyone's pursuit of his interests and living up to his obligations as he understands them.

Furthermore, liberty of conscience is to be limited only when there is a reasonable expectation that not doing so will damage the public order which the government should maintain. This expectation must be based on evidence and ways of reasoning acceptable to all. It must be supported by ordinary observation and modes of thought (including the methods of rational scientific inquiry where these are not controversial) which are generally recognized as correct. Now this reliance on what can be established and known by everyone is itself founded on the principles of justice. It implies no particular metaphysical doctrine or theory of knowledge. For this criterion appeals to what everyone can accept. It represents an agreement to limit liberty only by reference to a common knowledge and understanding of the world. Adopting this standard does not infringe upon anyone's equal freedom. On the other hand, a departure from generally recognized ways of reasoning would involve a privileged place for the views of some over others, and a principle which permitted this could not be agreed to in the original position. Furthermore, in holding that the consequences for the security of public order should not be merely possible or in certain cases even probable, but reasonably certain or imminent, there is again no implication of a particular philosophical theory. Rather this requirement expresses the high place which must be accorded to liberty of conscience and freedom of thought.

We may note at this point an analogy with the method of making interpersonal comparisons of well-being. These are founded on the index of primary goods that one may reasonably expect (§15), primary goods being those which everyone is presumed to want. This basis of comparison is one to which the parties can agree for the purposes of social justice. It does not require subtle estimates of men's capacity for happiness, much less of the relative worth of their plans of life. We need not question the meaningfulness of these notions; but they are inappropriate for designing just institutions. Similarly, the parties consent to publicly recognized criteria to determine what counts as evidence that their equal liberty is pursued in ways injurious to the common interest in public order and to the liberty of others. These principles of evidence are adopted for the aims of justice; they are not intended to apply to all questions of meaning and truth. How far they are valid in philosophy and science is a separate matter.

The characteristic feature of these arguments for liberty of conscience is that they are based solely on a conception of justice. Toleration is not derived from practical necessities or reasons of state. Moral and religious freedom follows from the principle of equal liberty; and assuming the priority of this principle, the only ground for denying the equal liberties is to avoid an even greater injustice, an even greater loss of liberty. Moreover, the argument does not rely on any special metaphysical or philosophical doctrine. It does not presuppose that all truths can be established by ways of thought recognized by common sense; nor does it hold that everything is, in some definable sense, a logical construction out of what can be observed or evidenced by rational scientific inquiry. The appeal is indeed to common sense, to generally shared ways of reasoning and plain facts accessible to all, but it is framed in such a way as to avoid these larger presumptions. Nor, on the other hand, does the case for liberty imply skepticism in philosophy or indifference to religion. Perhaps arguments for liberty of conscience can be given that have one or more of these doctrines as a premise. There is no reason to be surprised at this, since different arguments can have the same conclusion. But we need not pursue this question. The case for liberty is at least as strong as its strongest argument; the weak and fallacious ones are best forgotten. Those who would deny liberty of conscience cannot justify their action by condemning philosophical skepticism and indifference to religion, nor by appealing to social interests and affairs of state. The limitation of liberty is justified only when it is necessary for liberty itself, to prevent an invasion of freedom that would be still worse.

The parties in the constitutional convention, then, must choose a constitution that guarantees an equal liberty of conscience regulated solely by forms of argument generally accepted, and limited only when such argument establishes a reasonably certain interference with the essentials of public order. Liberty is governed by the necessary conditions for liberty itself. Now by this elementary principle alone many grounds of intolerance accepted in past ages are mistaken. Thus, for example, Aquinas justified the death penalty for heretics on the ground that it is a far graver matter to corrupt the faith, which is the life of the soul, than to counterfeit money which sustains life. So if it is just to put to death forgers and other criminals, heretics may a fortiori be similarly dealt with.[9] But the premises on which Aquinas relies cannot be established by modes of reasoning commonly recognized. It is a matter of dogma that faith is the life of the soul and that the suppression of heresy, that is, departures from ecclesiastical authority, is necessary for the safety of souls.

Again, the reasons given for limited toleration often run afoul of this principle. Thus Rousseau thought that people would find it impossible to live in peace with those whom they regarded as damned, since to love them would be to hate God who punishes them. He believed that those who regard others as damned must either torment or convert them, and therefore sects preaching this conviction cannot be trusted to preserve civil peace. Rousseau would not, then, tolerate those religions which say that outside the church there is no salvation.[10] But the consequences of such dogmatic belief which Rousseau conjectures are not borne out by experience. A priori psychological argument, however plausible, is not sufficient to abandon the principle of toleration, since justice holds that the disturbance to public order and to liberty itself must be securely established by common experience. There is, however, an important difference between Rousseau and Locke, who advocated a limited toleration, and Aquinas and the Protestant Reformers who did not.[11] Locke and Rousseau limited liberty on the basis of what they supposed were clear and evident consequences for the public order. If Catholics and atheists were not to be tolerated it was because it seemed evident that such per-

9. *Summa Theologica,* II–II, q. 11, art. 3.

10. *The Social Contract,* bk. IV, ch. VIII.

11. For the views of the Protestant Reformers, see J. E. E. D. (Lord) Acton, "The Protestant Theory of Persecution" in *The History of Freedom and Other Essays* (London, Macmillan, 1907). For Locke, see *A Letter Concerning Toleration,* included along with *The Second Treatise of Government,* ed. J. W. Gough (Oxford, Basil Blackwell, 1946), pp. 156–158.

sons could not be relied upon to observe the bonds of civil society. Presumably a greater historical experience and a knowledge of the wider possibilities of political life would have convinced them that they were mistaken, or at least that their contentions were true only under special circumstances. But with Aquinas and the Protestant Reformers the grounds of intolerance are themselves a matter of faith, and this difference is more fundamental than the limits actually drawn to toleration. For when the denial of liberty is justified by an appeal to public order as evidenced by common sense, it is always possible to urge that the limits have been drawn incorrectly, that experience does not in fact justify the restriction. Where the suppression of liberty is based upon theological principles or matters of faith, no argument is possible. The one view recognizes the priority of principles which would be chosen in the original position whereas the other does not.

35. TOLERATION OF THE INTOLERANT

Let us now consider whether justice requires the toleration of the intolerant, and if so under what conditions. There are a variety of situations in which this question arises. Some political parties in democratic states hold doctrines that commit them to suppress the constitutional liberties whenever they have the power. Again, there are those who reject intellectual freedom but who nevertheless hold positions in the university. It may appear that toleration in these cases is inconsistent with the principles of justice, or at any rate not required by them. I shall discuss the matter in connection with religious toleration. With appropriate alterations the argument can be extended to these other instances.

Several questions should be distinguished. First, there is the question whether an intolerant sect has any title to complain if it is not tolerated; second, under what conditions tolerant sects have a right not to tolerate those which are intolerant; and last, when they have the right not to tolerate them, for what ends it should be exercised. Beginning with the first question, it seems that an intolerant sect has no title to complain when it is denied an equal liberty. At least this follows if it is assumed that one has no title to object to the conduct of others that is in accordance with principles one would use in similar circumstances to justify one's actions toward them. A person's right to complain is limited to violations of principles he acknowledges himself. A complaint is a protest addressed to another in good faith. It claims a violation of a principle that both

parties accept. Now, to be sure, an intolerant man will say that he acts in good faith and that he does not ask anything for himself that he denies to others. His view, let us suppose, is that he is acting on the principle that God is to be obeyed and the truth accepted by all. This principle is perfectly general and by acting on it he is not making an exception in his own case. As he sees the matter, he is following the correct principle which others reject.

The reply to this defense is that, from the standpoint of the original position, no particular interpretation of religious truth can be acknowledged as binding upon citizens generally; nor can it be agreed that there should be one authority with the right to settle questions of theological doctrine. Each person must insist upon an equal right to decide what his religious obligations are. He cannot give up this right to another person or institutional authority. In fact, a man exercises his liberty in deciding to accept another as an authority even when he regards this authority as infallible, since in doing this he in no way abandons his equal liberty of conscience as a matter of constitutional law. For this liberty as secured by justice is imprescriptible: a person is always free to change his faith and this right does not depend upon his having exercised his powers of choice regularly or intelligently. We may observe that men's having an equal liberty of conscience is consistent with the idea that all men ought to obey God and accept the truth. The problem of liberty is that of choosing a principle by which the claims men make on one another in the name of their religion are to be regulated. Granting that God's will should be followed and the truth recognized does not as yet define a principle of adjudication. From the fact that God's intention is to be complied with, it does not follow that any person or institution has authority to interfere with another's interpretation of his religious obligations. This religious principle justifies no one in demanding in law or politics a greater liberty for himself. The only principles which authorize claims on institutions are those that would be chosen in the original position.

Let us suppose, then, that an intolerant sect has no title to complain of intolerance. We still cannot say that tolerant sects have the right to suppress them. For one thing, others may have a right to complain. They may have this right not as a right to complain on behalf of the intolerant, but simply as a right to object whenever a principle of justice is violated. For justice is infringed whenever equal liberty is denied without sufficient reason. The question, then, is whether being intolerant of another is grounds enough for limiting someone's liberty. To simplify things, assume that the tolerant sects have the right not to tolerate the intolerant in

at least one circumstance, namely, when they sincerely and with reason believe that intolerance is necessary for their own security. This right follows readily enough since, as the original position is defined, each would agree to the right of self-preservation. Justice does not require that men must stand idly by while others destroy the basis of their existence. Since it can never be to men's advantage, from a general point of view, to forgo the right of self-protection, the only question, then, is whether the tolerant have a right to curb the intolerant when they are of no immediate danger to the equal liberties of others.

Suppose that, in some way or other, an intolerant sect comes to exist within a well-ordered society accepting the two principles of justice. How are the citizens of this society to act in regard to it? Now certainly they should not suppress it simply because the members of the intolerant sect could not complain were they to do so. Rather, since a just constitution exists, all citizens have a natural duty of justice to uphold it. We are not released from this duty whenever others are disposed to act unjustly. A more stringent condition is required: there must be some considerable risks to our own legitimate interests. Thus just citizens should strive to preserve the constitution with all its equal liberties as long as liberty itself and their own freedom are not in danger. They can properly force the intolerant to respect the liberty of others, since a person can be required to respect the rights established by principles that he would acknowledge in the original position. But when the constitution itself is secure, there is no reason to deny freedom to the intolerant.

The question of tolerating the intolerant is directly related to that of the stability of a well-ordered society regulated by the two principles. We can see this as follows. It is from the position of equal citizenship that persons join the various religious associations, and it is from this position that they should conduct their discussions with one another. Citizens in a free society should not think one another incapable of a sense of justice unless this is necessary for the sake of equal liberty itself. If an intolerant sect appears in a well-ordered society, the others should keep in mind the inherent stability of their institutions. The liberties of the intolerant may persuade them to a belief in freedom. This persuasion works on the psychological principle that those whose liberties are protected by and who benefit from a just constitution will, other things equal, acquire an allegiance to it over a period of time (§72). So even if an intolerant sect should arise, provided that it is not so strong initially that it can impose its will straightway, or does not grow so rapidly that the psychological principle has no time to take hold, it will tend to lose its intolerance and

accept liberty of conscience. This is the consequence of the stability of just institutions, for stability means that when tendencies to injustice arise other forces will be called into play that work to preserve the justice of the whole arrangement. Of course, the intolerant sect may be so strong initially or growing so fast that the forces making for stability cannot convert it to liberty. This situation presents a practical dilemma which philosophy alone cannot resolve. Whether the liberty of the intolerant should be limited to preserve freedom under a just constitution depends on the circumstances. The theory of justice only characterizes the just constitution, the end of political action by reference to which practical decisions are to be made. In pursuing this end the natural strength of free institutions must not be forgotten, nor should it be supposed that tendencies to depart from them go unchecked and always win out. Knowing the inherent stability of a just constitution, members of a well-ordered society have the confidence to limit the freedom of the intolerant only in the special cases when it is necessary for preserving equal liberty itself.

The conclusion, then, is that while an intolerant sect does not itself have title to complain of intolerance, its freedom should be restricted only when the tolerant sincerely and with reason believe that their own security and that of the institutions of liberty are in danger. The tolerant should curb the intolerant only in this case. The leading principle is to establish a just constitution with the liberties of equal citizenship. The just should be guided by the principles of justice and not by the fact that the unjust cannot complain. Finally, it should be noted that even when the freedom of the intolerant is limited to safeguard a just constitution, this is not done in the name of maximizing liberty. The liberties of some are not suppressed simply to make possible a greater liberty for others. Justice forbids this sort of reasoning in connection with liberty as much as it does in regard to the sum of advantages. It is only the liberty of the intolerant which is to be limited, and this is done for the sake of equal liberty under a just constitution the principles of which the intolerant themselves would acknowledge in the original position.

The argument in this and the preceding sections suggests that the adoption of the principle of equal liberty can be viewed as a limiting case. Even though their differences are profound and no one knows how to reconcile them by reason, men can, from the standpoint of the original position, still agree on this principle if they can agree on any principle at all. This idea which arose historically with religious toleration can be extended to other instances. Thus we can suppose that the persons in the original position know that they have moral convictions although, as the

veil of ignorance requires, they do not know what these convictions are. They understand that the principles they acknowledge are to override these beliefs when there is a conflict; but otherwise they need not revise their opinions nor give them up when these principles do not uphold them. In this way the principles of justice can adjudicate between opposing moralities just as they regulate the claims of rival religions. Within the framework that justice establishes, moral conceptions with different principles, or conceptions representing a different balancing of the same principles, may be adopted by various parts of society. What is essential is that when persons with different convictions make conflicting demands on the basic structure as a matter of political principle, they are to judge these claims by the principles of justice. The principles that would be chosen in the original position are the kernel of political morality. They not only specify the terms of cooperation between persons but they define a pact of reconciliation between diverse religions and moral beliefs, and the forms of culture to which they belong. If this conception of justice now seems largely negative, we shall see that it has a happier side.

36. POLITICAL JUSTICE AND THE CONSTITUTION

I now wish to consider political justice, that is, the justice of the constitution, and to sketch the meaning of equal liberty for this part of the basic structure. Political justice has two aspects arising from the fact that a just constitution is a case of imperfect procedural justice. First, the constitution is to be a just procedure satisfying the requirements of equal liberty; and second, it is to be framed so that of all the feasible just arrangements, it is the one more likely than any other to result in a just and effective system of legislation. The justice of the constitution is to be assessed under both headings in the light of what circumstances permit, these assessments being made from the standpoint of the constitutional convention.

The principle of equal liberty, when applied to the political procedure defined by the constitution, I shall refer to as the principle of (equal) participation. It requires that all citizens are to have an equal right to take part in, and to determine the outcome of, the constitutional process that establishes the laws with which they are to comply. Justice as fairness begins with the idea that where common principles are necessary and to everyone's advantage, they are to be worked out from the viewpoint of a suitably defined initial situation of equality in which each person is fairly

represented. The principle of participation transfers this notion from the original position to the constitution as the highest-order system of social rules for making rules. If the state is to exercise a final and coercive authority over a certain territory, and if it is in this way to affect permanently men's prospects in life, then the constitutional process should preserve the equal representation of the original position to the degree that this is practicable.

For the time being I assume that a constitutional democracy can be arranged so as to satisfy the principle of participation. But we need to know more exactly what this principle requires under favorable circumstances, when taken to the limit so to speak. These requirements are, of course, familiar, comprising what Constant called the liberty of the ancients in contrast to the liberty of the moderns. Nevertheless, it is worthwhile to see how these liberties fall under the principle of participation. The adjustments that need to be made to existing conditions, and the reasoning that regulates these compromises, I discuss in the following section.

We may begin by recalling certain elements of a constitutional regime. First of all, the authority to determine basic social policies resides in a representative body selected for limited terms by and ultimately accountable to the electorate. This representative body has more than a purely advisory capacity. It is a legislature with lawmaking powers and not simply a forum of delegates from various sectors of society to which the executive explains its actions and discerns the movements of public sentiment. Nor are political parties mere interest groups petitioning the government on their own behalf; instead, to gain enough support to win office, they must advance some conception of the public good. The constitution may, of course, circumscribe the legislature in numerous respects; and constitutional norms define its actions as a parliamentary body. But in due course a firm majority of the electorate is able to achieve its aims, by constitutional amendment if necessary.

All sane adults, with certain generally recognized exceptions, have the right to take part in political affairs, and the precept one elector one vote is honored as far as possible. Elections are fair and free, and regularly held. Sporadic and unpredictable tests of public sentiment by plebiscite or other means, or at such times as may suit the convenience of those in office, do not suffice for a representative regime. There are firm constitutional protections for certain liberties, particularly freedom of speech and assembly, and liberty to form political associations. The principle of loyal opposition is recognized, the clash of political beliefs, and of the interests

and attitudes that are likely to influence them, are accepted as a normal condition of human life. A lack of unanimity is part of the circumstances of justice, since disagreement is bound to exist even among honest men who desire to follow much the same political principles. Without the conception of loyal opposition, and an attachment to constitutional rules which express and protect it, the politics of democracy cannot be properly conducted or long endure.

Three points concerning the equal liberty defined by the principle of participation call for discussion: its meaning, its extent, and the measures that enhance its worth. Starting with the question of meaning, the precept of one elector one vote implies, when strictly adhered to, that each vote has approximately the same weight in determining the outcome of elections. And this in turn requires, assuming single member territorial constituencies, that members of the legislature (with one vote each) represent the same number of electors. I shall also suppose that the precept necessitates that legislative districts be drawn up under the guidance of certain general standards specified in advance by the constitution and applied as far as possible by an impartial procedure. These safeguards are needed to prevent gerrymandering, since the weight of the vote can be as much affected by feats of gerrymander as by districts of disproportionate size. The requisite standards and procedures are to be adopted from the standpoint of the constitutional convention in which no one has the knowledge that is likely to prejudice the design of constituencies. Political parties cannot adjust boundaries to their advantage in the light of voting statistics; districts are defined by means of criteria already agreed to in the absence of this sort of information. Of course, it may be necessary to introduce certain random elements, since the criteria for designing constituencies are no doubt to some extent arbitrary. There may be no other fair way to deal with these contingencies.[12]

The principle of participation also holds that all citizens are to have an equal access, at least in the formal sense, to public office. Each is eligible to join political parties, to run for elective positions, and to hold places of authority. To be sure, there may be qualifications of age, residency, and so on. But these are to be reasonably related to the tasks of office; presumably these restrictions are in the common interest and do not discriminate unfairly among persons or groups in the sense that they fall evenly on everyone in the normal course of life.

12. For a discussion of this problem, see W. S. Vickrey, "On the Prevention of Gerrymandering," *Political Science Quarterly,* vol. 76 (1961).

The second point concerning equal political liberty is its extent. How broadly are these liberties to be defined? Offhand it is not clear what extent means here. Each of the political liberties can be more or less widely defined. Somewhat arbitrarily, but nevertheless in accordance with tradition, I shall assume that the main variation in the extent of equal political liberty lies in the degree to which the constitution is majoritarian. The definition of the other liberties I take to be more or less fixed. Thus the most extensive political liberty is established by a constitution that uses the procedure of so-called bare majority rule (the procedure in which a minority can neither override nor check a majority) for all significant political decisions unimpeded by any constitutional constraints. Whenever the constitution limits the scope and authority of majorities, either by requiring a greater plurality for certain types of measures, or by a bill of rights restricting the powers of the legislature, and the like, equal political liberty is less extensive. The traditional devices of constitutionalism—bicameral legislature, separation of powers mixed with checks and balances, a bill of rights with judicial review—limit the scope of the principle of participation. I assume, however, that these arrangements are consistent with equal political liberty provided that similar restrictions apply to everyone and that the constraints introduced are likely over time to fall evenly upon all sectors of society. And this seems probable if the fair value of political liberty is maintained. The main problem, then, is how extensive equal participation should be. This question I leave aside for the next section.

Turning now to the worth of political liberty, the constitution must take steps to enhance the value of the equal rights of participation for all members of society. It must underwrite a fair opportunity to take part in and to influence the political process. The distinction here is analogous to that made before (§12): ideally, those similarly endowed and motivated should have roughly the same chance of attaining positions of political authority irrespective of their economic and social class. But how is this fair value of these liberties to be secured?

We may take for granted that a democratic regime presupposes freedom of speech and assembly, and liberty of thought and conscience. These institutions are not only required by the first principle of justice but, as Mill argued, they are necessary if political affairs are to be conducted in a rational fashion. While rationality is not guaranteed by these arrangements, in their absence the more reasonable course seems sure to be rejected in favor of policies sought by special interests. If the public forum is to be free and open to all, and in continuous session, everyone

should be able to make use of it. All citizens should have the means to be informed about political issues. They should be in a position to assess how proposals affect their well-being and which policies advance their conception of the public good. Moreover, they should have a fair chance to add alternative proposals to the agenda for political discussion.[13] The liberties protected by the principle of participation lose much of their value whenever those who have greater private means are permitted to use their advantages to control the course of public debate. For eventually these inequalities will enable those better situated to exercise a larger influence over the development of legislation. In due time they are likely to acquire a preponderant weight in settling social questions, at least in regard to those matters upon which they normally agree, which is to say in regard to those things that support their favored circumstances.

Compensating steps must, then, be taken to preserve the fair value for all of the equal political liberties. A variety of devices can be used. For example, in a society allowing private ownership of the means of production, property and wealth must be kept widely distributed and government monies provided on a regular basis to encourage free public discussion. In addition, political parties are to be made independent from private economic interests by allotting them sufficient tax revenues to play their part in the constitutional scheme. (Their subventions might, for example, be based by some rule on the number of votes received in the last several elections, and the like.) What is necessary is that political parties be autonomous with respect to private demands, that is, demands not expressed in the public forum and argued for openly by reference to a conception of the public good. If society does not bear the costs of organization, and party funds need to be solicited from the more advantaged social and economic interests, the pleadings of these groups are bound to receive excessive attention. And this is all the more likely when the less favored members of society, having been effectively prevented by their lack of means from exercising their fair degree of influence, withdraw into apathy and resentment.

Historically one of the main defects of constitutional government has been the failure to insure the fair value of political liberty. The necessary corrective steps have not been taken, indeed, they never seem to have been seriously entertained. Disparities in the distribution of property and

13. See R. A. Dahl, *A Preface to Democratic Theory* (Chicago, University of Chicago Press, 1956), pp. 67–75, for a discussion of the conditions necessary to achieve political equality.

wealth that far exceed what is compatible with political equality have generally been tolerated by the legal system. Public resources have not been devoted to maintaining the institutions required for the fair value of political liberty. Essentially the fault lies in the fact that the democratic political process is at best regulated rivalry; it does not even in theory have the desirable properties that price theory ascribes to truly competitive markets. Moreover, the effects of injustices in the political system are much more grave and long lasting than market imperfections. Political power rapidly accumulates and becomes unequal; and making use of the coercive apparatus of the state and its law, those who gain the advantage can often assure themselves of a favored position. Thus inequities in the economic and social system may soon undermine whatever political equality might have existed under fortunate historical conditions. Universal suffrage is an insufficient counterpoise; for when parties and elections are financed not by public funds but by private contributions, the political forum is so constrained by the wishes of the dominant interests that the basic measures needed to establish just constitutional rule are seldom properly presented. These questions, however, belong to political sociology.[14] I mention them here as a way of emphasizing that our discussion is part of the theory of justice and must not be mistaken for a theory of the political system. We are in the way of describing an ideal arrangement, comparison with which defines a standard for judging actual institutions, and indicates what must be maintained to justify departures from it.

By way of summing up the account of the principle of participation, we can say that a just constitution sets up a form of fair rivalry for political office and authority. By presenting conceptions of the public good and policies designed to promote social ends, rival parties seek the citizens' approval in accordance with just procedural rules against a background of freedom of thought and assembly in which the fair value of political liberty is assured. The principle of participation compels those in authority to be responsive to the felt interests of the electorate. Representatives are not, to be sure, mere agents of their constituents, since they have a certain discretion and they are expected to exercise their judgment in enacting legislation. In a well-ordered society they must, nevertheless, represent their constituents in the substantive sense: they must seek first to pass just and effective legislation, since this is a citizen's first interest

14. My remarks draw upon F. H. Knight, *The Ethics of Competition and Other Essays* (New York, Harper and Brothers, 1935), pp. 293–305.

in government, and secondly, they must further their constituents' other interests insofar as these are consistent with justice.[15] The principles of justice are among the main criteria to be used in judging a representative's record and the reasons he gives in defense of it. Since the constitution is the foundation of the social structure, the highest-order system of rules that regulates and controls other institutions, everyone has the same access to the political procedure that it sets up. When the principle of participation is satisfied, all have the common status of equal citizen.

Finally, to avoid misunderstanding, it should be kept in mind that the principle of participation applies to institutions. It does not define an ideal of citizenship; nor does it lay down a duty requiring all to take an active part in political affairs. The duties and obligations of individuals are a separate question that I shall discuss later (see Chapter VI). What is essential is that the constitution should establish equal rights to engage in public affairs and that measures be taken to maintain the fair value of these liberties. In a well-governed state only a small fraction of persons may devote much of their time to politics. There are many other forms of human good. But this fraction, whatever its size, will most likely be drawn more or less equally from all sectors of society. The many communities of interests and centers of political life will have their active members who look after their concerns.

37. LIMITATIONS ON THE PRINCIPLE OF PARTICIPATION

It is evident from the preceding account of the principle of participation that there are three ways to limit its application. The constitution may define a more or a less extensive freedom of participation; it may allow inequalities in political liberties; and greater or smaller social resources may be devoted to insuring the worth of these freedoms to the representative citizen. I shall discuss these kinds of limitations in order, all with a view to clarifying the meaning of the priority of liberty.

The extent of the principle of participation is defined as the degree to which the procedure of (bare) majority rule is restricted by the mechanisms of constitutionalism. These devices serve to limit the scope of majority rule, the kinds of matters on which majorities have final authority, and the speed with which the aims of the majority are put into effect.

15. See H. F. Pitkin, *The Concept of Representation* (Berkeley, University of California Press, 1967), pp. 221–225, for a discussion of representation to which I am indebted.

A bill of rights may remove certain liberties from majority regulation altogether, and the separation of powers with judicial review may slow down the pace of legislative change. The question, then, is how these mechanisms might be justified consistent with the two principles of justice. We are not to ask whether these devices are in fact justified, but what kind of an argument for them is required.

To begin with, however, we should observe that the limits on the extent of the principle of participation are assumed to fall equally upon everyone. For this reason these restrictions are easier to justify than unequal political liberties. If all could have a greater liberty, at least each loses equally, other things the same; and if this lesser liberty is unnecessary and not imposed by some human agency, the scheme of liberty is to this degree irrational rather than unjust. Unequal liberty, as when the precept one man one vote is violated, is another matter and immediately raises a question of justice.

Supposing for the time being that the constraints on majority rule bear equally on all citizens, the justification for the devices of constitutionalism is that they presumably protect the other freedoms. The best arrangement is found by noting the consequences for the complete system of liberty. The intuitive idea here is straightforward. We have said that the political process is a case of imperfect procedural justice. A constitution that restricts majority rule by the various traditional devices is thought to lead to a more just body of legislation. Since the majority principle must as a practical necessity be relied upon to some degree, the problem is to find which constraints work best in given circumstances to further the ends of liberty. Of course, these matters lie outside the theory of justice. We need not consider which if any of the constitutional mechanisms is effective in achieving its aim, or how far its successful working presupposes certain underlying social conditions. The relevant point is that to justify these restrictions one must maintain that from the perspective of the representative citizen in the constitutional convention the less extensive freedom of participation is sufficiently outweighed by the greater security and extent of the other liberties. Unlimited majority rule is often thought to be hostile to these liberties. Constitutional arrangements compel a majority to delay putting its will into effect and force it to make a more considered and deliberate decision. In this and other ways procedural constraints are said to mitigate the defects of the majority principle. The justification appeals to a greater equal liberty. At no point is there a reference to compensating economic and social benefits.

One of the tenets of classical liberalism is that the political liberties are

of less intrinsic importance than liberty of conscience and freedom of the person. Should one be forced to choose between the political liberties and all the others, the governance of a good sovereign who recognized the latter and who upheld the rule of law would be far preferable. On this view, the chief merit of the principle of participation is to insure that the government respects the rights and welfare of the governed.[16] Fortunately, however, we do not often have to assess the relative total importance of the different liberties. Usually the way to proceed is to apply the principle of equal advantage in adjusting the complete system of freedom. We are not called upon either to abandon the principle of participation entirely or to allow it unlimited sway. Instead, we should narrow or widen its extent up to the point where the danger to liberty from the marginal loss in control over those holding political power just balances the security of liberty gained by the greater use of constitutional devices. The decision is not an all or nothing affair. It is a question of weighing against one another small variations in the extent and definition of the different liberties. The priority of liberty does not exclude marginal exchanges within the system of freedom. Moreover, it allows although it does not require that some liberties, say those covered by the principle of participation, are less essential in that their main role is to protect the remaining freedoms. Different opinions about the value of the liberties will, of course, affect how different persons think the full scheme of freedom should be arranged. Those who place a higher worth on the principle of participation will be prepared to take greater risks with the freedoms of the person, say, in order to give political liberty a larger place. Ideally these conflicts will not occur and it should be possible, under favorable conditions anyway, to find a constitutional procedure that allows a sufficient scope for the value of participation without jeopardizing the other liberties.

It is sometimes objected to majority rule that, however circumscribed, it fails to take account of the intensity of desire, since the larger part may override the strong feelings of a minority. This criticism rests upon the mistaken view that the intensity of desire is a relevant consideration in enacting legislation (see §54). To the contrary, whenever questions of justice are raised, we are not to go by the strength of feeling but must aim instead for the greater justice of the legal order. The fundamental criterion for judging any procedure is the justice of its likely results. A similar reply may be given to the propriety of majority rule when the vote is

16. See Isaiah Berlin, *Four Essays on Liberty,* pp. 130, 165.

rather evenly divided. Everything depends on the probable justice of the outcome. If the various sectors of society have reasonable confidence in one another and share a common conception of justice, the rule by bare majorities may succeed fairly well. To the extent that this underlying agreement is lacking, the majority principle becomes more difficult to justify because it is less probable that just policies will be followed. There may, however, be no procedures that can be relied upon once distrust and enmity pervade society. I do not wish to pursue these matters further. I mention these familiar points about majority rule only to emphasize that the test of constitutional arrangements is always the overall balance of justice. Where issues of justice are involved, the intensity of desires should not be taken into account. Of course, as things are, legislators must reckon with strong public feelings. Men's sense of outrage however irrational will set boundaries upon what is politically attainable; and popular views will affect the strategies of enforcement within these limits. But questions of strategy are not to be confused with those of justice. If a bill of rights guaranteeing liberty of conscience and freedom of thought and assembly would be effective, then it should be adopted. Whatever the depth of feeling against them, these rights should if possible be made to stand. The force of opposing attitudes has no bearing on the question of right but only on the feasibility of arrangements of liberty.

The justification of unequal political liberty proceeds in much the same way. One takes up the point of view of the representative citizen in the constitutional convention and assesses the total system of freedom as it looks to him. But in this case there is an important difference. We must now reason from the perspective of those who have the lesser political liberty. An inequality in the basic structure must always be justified to those in the disadvantaged position. This holds whatever the primary social good and especially for liberty. Therefore the priority rule requires us to show that the inequality of right would be accepted by the less favored in return for the greater protection of their other liberties that results from this restriction.

Perhaps the most obvious political inequality is the violation of the precept one person one vote. Yet until recent times most writers rejected equal universal suffrage. Indeed, persons were not regarded as the proper subjects of representation at all. Often it was interests that were to be represented, with Whig and Tory differing as to whether the interest of the rising middle class should be given a place alongside the landed and

ecclesiastical interests. For others it is regions that are to be represented, or forms of culture, as when one speaks of the representation of the agricultural and urban elements of society. At the first sight, these kinds of representation appear unjust. How far they depart from the precept one person one vote is a measure of their abstract injustice, and indicates the strength of the countervailing reasons that must be forthcoming.[17]

Now it frequently turns out that those who oppose equal political liberty put forward justifications of the required form. They are at least prepared to argue that political inequality is to the benefit of those with the lesser liberty. Consider as an illustration Mill's view that persons with greater intelligence and education should have extra votes in order that their opinions may have a greater influence.[18] Mill believed that in this case plural voting accords with the natural order of human life, for whenever persons conduct a common enterprise in which they have a joint interest, they recognize that while all should have a voice, the say of everyone need not be equal. The judgment of the wiser and more knowledgeable should have a superior weight. Such an arrangement is in the interest of each and conforms to men's sentiment of justice. National affairs are precisely such a joint concern. Although all should indeed have the vote, those with a greater capacity for the management of the public interest should have a larger say. Their influence should be great enough to protect them from the class legislation of the uneducated, but not so large as to allow them to enact class legislation in their own behalf. Ideally, those with superior wisdom and judgment should act as a constant force on the side of justice and the common good, a force that, although always weak by itself, can often tip the scale in the right direction if the larger forces cancel out. Mill was persuaded that everyone would gain from this arrangement, including those whose votes count for less. Of course, as it stands, this argument does not go beyond the general conception of justice as fairness. Mill does not state explicitly that the gain to the uneducated is to be estimated in the first instance by the larger security of their other liberties, although his reasoning suggests that he thought this to be the case. In any event, if Mill's view is to satisfy the restrictions imposed by the priority of liberty, this is how the argument must go.

I do not wish to criticize Mill's proposal. My account of it is solely for

17. See J. R. Pole, *Political Representation in England and the Origin of the American Republic* (London, Macmillan, 1966), pp. 535–537.

18. *Representative Government,* ed. R. B. McCallum, together with *On Liberty* (Oxford, Basil Blackwell, 1946). pp. 216–222. (This is much of the latter half of ch. VIII.)

purposes of illustration. His view enables one to see why political equality is sometimes regarded as less essential than equal liberty of conscience or liberty of the person. Government is assumed to aim at the common good, that is, at maintaining conditions and achieving objectives that are similarly to everyone's advantage. To the extent that this presumption holds, and some men can be identified as having superior wisdom and judgment, others are willing to trust them and to concede to their opinion a greater weight. The passengers of a ship are willing to let the captain steer the course, since they believe that he is more knowledgeable and wishes to arrive safely as much as they do. There is both an identity of interests and a noticeably greater skill and judgment in realizing it. Now the ship of state is in some ways analogous to a ship at sea; and to the extent that this is so, the political liberties are indeed subordinate to the other freedoms that, so to say, define the intrinsic good of the passengers. Admitting these assumptions, plural voting may be perfectly just.

Of course, the grounds for self-government are not solely instrumental. Equal political liberty when assured its fair value is bound to have a profound effect on the moral quality of civic life. Citizens' relations to one another are given a secure basis in the manifest constitution of society. The medieval maxim that what touches all concerns all is seen to be taken seriously and declared as the public intention. Political liberty so understood is not designed to satisfy the individual's desire for self-mastery, much less his quest for power. Taking part in political life does not make the individual master of himself, but rather gives him an equal voice along with others in settling how basic social conditions are to be arranged. Nor does it answer to the ambition to dictate to others, since each is now required to moderate his claims by what everyone is able to recognize as just. The public will to consult and to take everyone's beliefs and interests into account lays the foundations for civic friendship and shapes the ethos of political culture.

Moreover, the effect of self-government where equal political rights have their fair value is to enhance the self-esteem and the sense of political competence of the average citizen. His awareness of his own worth developed in the smaller associations of his community is confirmed in the constitution of the whole society. Since he is expected to vote, he is expected to have political opinions. The time and thought that he devotes to forming his views is not governed by the likely material return of his political influence. Rather it is an activity enjoyable in itself that leads to a larger conception of society and to the development of his intellectual

and moral faculties. As Mill observed, he is called upon to weigh interests other than his own, and to be guided by some conception of justice and the public good rather than by his own inclinations.[19] Having to explain and justify his views to others, he must appeal to principles that others can accept. Moreover, Mill adds, this education to public spirit is necessary if citizens are to acquire an affirmative sense of political duty and obligation, that is, one that goes beyond the mere willingness to submit to law and government. Without these more inclusive sentiments men become estranged and isolated in their smaller associations, and affective ties may not extend outside the family or a narrow circle of friends. Citizens no longer regard one another as associates with whom one can cooperate to advance some interpretation of the public good; instead, they view themselves as rivals, or else as obstacles to one another's ends. All of these considerations Mill and others have made familiar. They show that equal political liberty is not solely a means. These freedoms strengthen men's sense of their own worth, enlarge their intellectual and moral sensibilities, and lay the basis for a sense of duty and obligation upon which the stability of just institutions depends. The connection of these matters to human good and the sense of justice I shall leave until Part Three. There I shall try to tie these things together under the conception of the good of justice.

38. THE RULE OF LAW

I now wish to consider rights of the person as these are protected by the principle of the rule of law.[20] As before my intention is not only to relate these notions to the principles of justice but to elucidate the sense of the priority of liberty. I have already noted (§10) that the conception of formal justice, the regular and impartial administration of public rules, becomes the rule of law when applied to the legal system. One kind of unjust action is the failure of judges and others in authority to apply the

19. *Representative Government,* pp. 149–151, 209–211. (These are the end of ch. III, and at the beginning of ch. VIII.)

20. For a general discussion, see Lon Fuller, *The Morality of Law* (New Haven, Yale University Press, 1964), ch. II. The concept of principled decisions in constitutional law is considered by Herbert Wechsler, *Principles, Politics, and Fundamental Law* (Cambridge, Harvard University Press, 1961). See Otto Kirchenheimer, *Political Justice* (Princeton, Princeton University Press, 1961), and J. N. Shklar, *Legalism* (Cambridge, Harvard University Press, 1964), pt. II, for the use and abuse of judicial forms in politics. J. R. Lucas, *The Principles of Politics* (Oxford, The Clarendon Press, 1966), pp. 106–143, contains a philosophical account.

appropriate rule or to interpret it correctly. It is more illuminating in this connection to think not of gross violations exemplified by bribery and corruption, or the abuse of the legal system to punish political enemies, but rather of the subtle distortions of prejudice and bias as these effectively discriminate against certain groups in the judicial process. The regular and impartial, and in this sense fair, administration of law we may call "justice as regularity." This is a more suggestive phrase than "formal justice."

Now the rule of law is obviously closely related to liberty. We can see this by considering the notion of a legal system and its intimate connection with the precepts definitive of justice as regularity. A legal system is a coercive order of public rules addressed to rational persons for the purpose of regulating their conduct and providing the framework for social cooperation. When these rules are just they establish a basis for legitimate expectations. They constitute grounds upon which persons can rely on one another and rightly object when their expectations are not fulfilled. If the bases of these claims are unsure, so are the boundaries of men's liberties. Of course, other rules share many of these features. Rules of games and of private associations are likewise addressed to rational persons in order to give shape to their activities. Given that these rules are fair or just, then once men have entered into these arrangements and accepted the benefits that result, the obligations which thereby arise constitute a basis for legitimate expectations. What distinguishes a legal system is its comprehensive scope and its regulative powers with respect to other associations. The constitutional agencies that it defines generally have the exclusive legal right to at least the more extreme forms of coercion. The kinds of duress that private associations can employ are strictly limited. Moreover, the legal order exercises a final authority over a certain well-defined territory. It is also marked by the wide range of the activities it regulates and the fundamental nature of the interests it is designed to secure. These features simply reflect the fact that the law defines the basic structure within which the pursuit of all other activities takes place.

Given that the legal order is a system of public rules addressed to rational persons, we can account for the precepts of justice associated with the rule of law. These precepts are those that would be followed by any system of rules which perfectly embodied the idea of a legal system. This is not, of course, to say that existing laws necessarily satisfy these precepts in all cases. Rather, these maxims follow from an ideal notion which laws are expected to approximate, at least for the most part. If

deviations from justice as regularity are too pervasive, a serious question may arise whether a system of law exists as opposed to a collection of particular orders designed to advance the interests of a dictator or the ideal of a benevolent despot. Often there is no clear answer to this question. The point of thinking of a legal order as a system of public rules is that it enables us to derive the precepts associated with the principle of legality. Moreover, we can say that, other things equal, one legal order is more justly administered than another if it more perfectly fulfills the precepts of the rule of law. It will provide a more secure basis for liberty and a more effective means for organizing cooperative schemes. Yet because these precepts guarantee only the impartial and regular administration of rules, whatever these are, they are compatible with injustice. They impose rather weak constraints on the basic structure, but ones that are not by any means negligible.

Let us begin with the precept that ought implies can. This precept identifies several obvious features of legal systems. First of all, the actions which the rules of law require and forbid should be of a kind which men can reasonably be expected to do and to avoid. A system of rules addressed to rational persons to organize their conduct concerns itself with what they can and cannot do. It must not impose a duty to do what cannot be done. Secondly, the notion that ought implies can conveys the idea that those who enact laws and give orders do so in good faith. Legislators and judges, and other officials of the system, must believe that the laws can be obeyed; and they are to assume that any orders given can be carried out. Moreover, not only must the authorities act in good faith, but their good faith must be recognized by those subject to their enactments. Laws and commands are accepted as laws and commands only if it is generally believed that they can be obeyed and executed. If this is in question, the actions of authorities presumably have some other purpose than to organize conduct. Finally, this precept expresses the requirement that a legal system should recognize impossibility of performance as a defense, or at least as a mitigating circumstance. In enforcing rules a legal system cannot regard the inability to perform as irrelevant. It would be an intolerable burden on liberty if the liability to penalties was not normally limited to actions within our power to do or not to do.

The rule of law also implies the precept that similar cases be treated similarly. Men could not regulate their actions by means of rules if this precept were not followed. To be sure, this notion does not take us very far. For we must suppose that the criteria of similarity are given by the legal rules themselves and the principles used to interpret them. Never-

theless, the precept that like decisions be given in like cases significantly limits the discretion of judges and others in authority. The precept forces them to justify the distinctions that they make between persons by reference to the relevant legal rules and principles. In any particular case, if the rules are at all complicated and call for interpretation, it may be easy to justify an arbitrary decision. But as the number of cases increases, plausible justifications for biased judgments become more difficult to construct. The requirement of consistency holds of course for the interpretation of all rules and for justifications at all levels. Eventually reasoned arguments for discriminatory judgments become harder to formulate and the attempt to do so less persuasive. This precept holds also in cases of equity, that is, when an exception is to be made when the established rule works an unexpected hardship. But with this proviso: since there is no clear line separating these exceptional cases, there comes a point, as in matters of interpretation, at which nearly any difference will make a difference. In these instances, the principle of authoritative decision applies, and the weight of precedent or of the announced verdict suffices.[21]

The precept that there is no offense without a law *(Nullum crimen sine lege),* and the requirements it implies, also follow from the idea of a legal system. This precept demands that laws be known and expressly promulgated, that their meaning be clearly defined, that statutes be general both in statement and intent and not be used as a way of harming particular individuals who may be expressly named (bills of attainder), that at least the more severe offenses be strictly construed, and that penal laws should not be retroactive to the disadvantage of those to whom they apply. These requirements are implicit in the notion of regulating behavior by public rules. For if, say, statutes are not clear in what they enjoin and forbid, the citizen does not know how he is to behave. Moreover, while there may be occasional bills of attainder and retroactive enactments, these cannot be pervasive or characteristic features of the system, else it must have another purpose. A tyrant might change laws without notice, and punish (if that is the right word) his subjects accordingly, because he takes pleasure in seeing how long it takes them to figure out what the new rules are from observing the penalties he inflicts. But these rules would not be a legal system, since they would not serve to organize social behavior by providing a basis for legitimate expectations.

Finally, there are those precepts defining the notion of natural justice. These are guidelines intended to preserve the integrity of the judicial

21. See Lon Fuller, *Anatomy of the Law* (New York, The New American Library, 1969), p. 182.

process.[22] If laws are directives addressed to rational persons for their guidance, courts must be concerned to apply and to enforce these rules in an appropriate way. A conscientious effort must be made to determine whether an infraction has taken place and to impose the correct penalty. Thus a legal system must make provisions for conducting orderly trials and hearings; it must contain rules of evidence that guarantee rational procedures of inquiry. While there are variations in these procedures, the rule of law requires some form of due process: that is, a process reasonably designed to ascertain the truth, in ways consistent with the other ends of the legal system, as to whether a violation has taken place and under what circumstances. For example, judges must be independent and impartial, and no man may judge his own case. Trials must be fair and open, but not prejudiced by public clamor. The precepts of natural justice are to insure that the legal order will be impartially and regularly maintained.

Now the connection of the rule of law with liberty is clear enough. Liberty, as I have said, is a complex of rights and duties defined by institutions. The various liberties specify things that we may choose to do, if we wish, and in regard to which, when the nature of the liberty makes it appropriate, others have a duty not to interfere.[23] But if the precept of no crime without a law is violated, say by statutes, being vague and imprecise, what we are at liberty to do is likewise vague and imprecise. The boundaries of our liberty are uncertain. And to the extent that this is so, liberty is restricted by a reasonable fear of its exercise. The same sort of consequences follow if similar cases are not treated similarly, if the judicial process lacks its essential integrity, if the law does not recognize impossibility of performance as a defense, and so on. The principle of legality has a firm foundation, then, in the agreement of rational persons

22. This sense of natural justice is traditional. See H. L. A. Hart, *The Concept of Law* (Oxford, The Clarendon Press, 1961), pp. 156, 202.

23. It may be disputed whether this view holds for all rights, for example, the right to pick up an unclaimed article. See Hart in *Philosophical Review,* vol. 64, p. 179. But perhaps it is true enough for our purposes here. While some of the basic rights are similarly competition rights, as we may call them—for example, the right to participate in public affairs and to influence the political decisions taken—at the same time everyone has a duty to conduct himself in a certain way. This duty is one of fair political conduct, so to speak, and to violate it is a kind of interference. As we have seen, the constitution aims to establish a framework within which equal political rights fairly pursued and having their fair value are likely to lead to just and effective legislation. When appropriate we can interpret the statement in the text along these lines. On this point see Richard Wollheim, "Equality," *Proceedings of the Aristotelian Society,* vol. 56 (1955–1956), pp. 291ff. Put another way, the right can be redescribed as the right to try to do something under specified circumstances, these circumstances allowing for the fair rivalry of others. Unfairness becomes a characteristic form of interference.

to establish for themselves the greatest equal liberty. To be confident in the possession and exercise of these freedoms, the citizens of a well-ordered society will normally want the rule of law maintained.

We can arrive at the same conclusion in a slightly different way. It is reasonable to assume that even in a well-ordered society the coercive powers of government are to some degree necessary for the stability of social cooperation. For although men know that they share a common sense of justice and that each wants to adhere to the existing arrangements, they may nevertheless lack full confidence in one another. They may suspect that some are not doing their part, and so they may be tempted not to do theirs. The general awareness of these temptations may eventually cause the scheme to break down. The suspicion that others are not honoring their duties and obligations is increased by the fact that, in the absence of the authoritative interpretation and enforcement of the rules, it is particularly easy to find excuses for breaking them. Thus even under reasonably ideal conditions, it is hard to imagine, for example, a successful income tax scheme on a voluntary basis. Such an arrangement is unstable. The role of an authorized public interpretation of rules supported by collective sanctions is precisely to overcome this instability. By enforcing a public system of penalties government removes the grounds for thinking that others are not complying with the rules. For this reason alone, a coercive sovereign is presumably always necessary, even though in a well-ordered society sanctions are not severe and may never need to be imposed. Rather, the existence of effective penal machinery serves as men's security to one another. This proposition and the reasoning behind it we may think of as Hobbes's thesis[24] (§42).

Now in setting up such a system of sanctions the parties in a constitutional convention must weigh its disadvantages. These are of at least two kinds: one kind is the cost of maintaining the agency covered say by taxation; the other is the danger to the liberty of the representative citizen measured by the likelihood that these sanctions will wrongly interfere with his freedom. The establishment of a coercive agency is rational only if these disadvantages are less than the loss of liberty from instability. Assuming this to be so, the best arrangement is one that minimizes these hazards. It is clear that, other things equal, the dangers to liberty are less when the law is impartially and regularly administered in accordance with the principle of legality. While a coercive mechanism is necessary, it

24. See *Leviathan,* chs. 13–18. And also Howard Warrender, *The Political Philosophy of Hobbes* (Oxford, The Clarendon Press, 1957), ch. III; and D. P. Gauthier, *The Logic of Leviathan* (Oxford, The Clarendon Press, 1969), pp. 76–89.

is obviously essential to define precisely the tendency of its operations. Knowing what things it penalizes and knowing that these are within their power to do or not to do, citizens can draw up their plans accordingly. One who complies with the announced rules need never fear an infringement of his liberty.

It is clear from the preceding remarks that we need an account of penal sanctions however limited even for ideal theory. Given the normal conditions of human life, some such arrangements are necessary. I have maintained that the principles justifying these sanctions can be derived from the principle of liberty. The ideal conception shows in this case anyway how the nonideal scheme is to be set up; and this confirms the conjecture that it is ideal theory which is fundamental. We also see that the principle of responsibility is not founded on the idea that punishment is primarily retributive or denunciatory. Instead it is acknowledged for the sake of liberty itself. Unless citizens are able to know what the law is and are given a fair opportunity to take its directives into account, penal sanctions should not apply to them. This principle is simply the consequence of regarding a legal system as an order of public rules addressed to rational persons in order to regulate their cooperation, and of giving the appropriate weight to liberty. I believe that this view of responsibility enables us to explain most of the excuses and defenses recognized by the criminal law under the heading of *mens rea* and that it can serve as a guide to legal reform. However, these points cannot be pursued here.[25] It suffices to note that ideal theory requires an account of penal sanctions as a stabilizing device and indicates the manner in which this part of partial compliance theory should be worked out. In particular, the principle of liberty leads to the principle of responsibility.

The moral dilemmas that arise in partial compliance theory are also to be viewed with the priority of liberty in mind. Thus we can imagine situations of an unhappy sort in which it may be permissible to insist less strongly on the precepts of the rule of law being followed. For example, in some extreme eventualities persons might be held liable for certain offenses contrary to the precept ought implies can. Suppose that, aroused by sharp religious antagonisms, members of rival sects are collecting weapons and forming armed bands in preparation for civil strife. Confronted with this situation the government may enact a statute forbidding

25. For these matters, consult H. L. A. Hart, *Punishment and Responsibility* (Oxford, The Clarendon Press, 1968), pp. 173–183, whom I follow here.

the possession of firearms (assuming that possession is not already an offense). And the law may hold that sufficient evidence for conviction is that the weapons are found in the defendant's house or property, unless he can establish that they were put there by another. Except for this proviso, the absence of intent and knowledge of possession, and conformity to reasonable standards of care, are declared irrelevant. It is contended that these normal defenses would make the law ineffective and impossible to enforce.

Now although this statute trespasses upon the precept ought implies can it might be accepted by the representative citizen as a lesser loss of liberty, at least if the penalties imposed are not too severe. (Here I assume that imprisonment, say, is a drastic curtailment of liberty, and so the severity of the contemplated punishments must be taken into account.) Viewing the situation from the legislative stage, one may decide that the formation of paramilitary groups, which the passing of the statute may forestall, is a much greater danger to the freedom of the average citizen than being held strictly liable for the possession of weapons. Citizens may affirm the law as the lesser of two evils, resigning themselves to the fact that while they may be held guilty for things they have not done, the risks to their liberty on any other course would be worse. Since bitter dissensions exist, there is no way to prevent some injustices, as we ordinarily think of them, from occurring. All that can be done is to limit these injustices in the least unjust way.

The conclusion once again is that arguments for restricting liberty proceed from the principle of liberty itself. To some degree anyway, the priority of liberty carries over to partial compliance theory. Thus in the situation discussed the greater good of some has not been balanced against the lesser good of others. Nor has a lesser liberty been accepted for the sake of greater economic and social benefits. Rather the appeal has been to the common good in the form of the basic equal liberties of the representative citizen. Unfortunate circumstances and the unjust designs of some necessitate a much lesser liberty than that enjoyed in a well-ordered society. Any injustice in the social order is bound to take its toll; it is impossible that its consequences should be entirely canceled out. In applying the principle of legality we must keep in mind the totality of rights and duties that defines the liberties and adjust its claims accordingly. Sometimes we may be forced to allow certain breaches of its precepts if we are to mitigate the loss of freedom from social evils that cannot be removed, and to aim for the least injustice that conditions allow.

39. THE PRIORITY OF LIBERTY DEFINED

Aristotle remarks that it is a peculiarity of men that they possess a sense of the just and the unjust and that their sharing a common understanding of justice makes a polis.[26] Analogously one might say, in view of our discussion, that a common understanding of justice as fairness makes a constitutional democracy. For I have tried to show, after presenting further arguments for the first principle, that the basic liberties of a democratic regime are most firmly secured by this conception of justice. In each case the conclusions reached are familiar. My aim has been to indicate not only that the principles of justice fit our considered judgments but also that they provide the strongest arguments for freedom. By contrast teleological principles permit at best uncertain grounds for liberty, or at least for equal liberty. And liberty of conscience and freedom of thought should not be founded on philosophical or ethical skepticism, nor on indifference to religious and moral interests. The principles of justice define an appropriate path between dogmatism and intolerance on the one side, and a reductionism which regards religion and morality as mere preferences on the other. And since the theory of justice relies upon weak and widely held presumptions, it may win quite general acceptance. Surely our liberties are most firmly based when they are derived from principles that persons fairly situated with respect to one another can agree to if they can agree to anything at all.

I now wish to examine more carefully the meaning of the priority of liberty. I shall not argue here for this priority (leaving this aside until §82); instead I wish to clarify its sense in view of the preceding examples, among others. There are several priorities to be distinguished. By the priority of liberty I mean the precedence of the principle of equal liberty over the second principle of justice. The two principles are in lexical order, and therefore the claims of liberty are to be satisfied first. Until this is achieved no other principle comes into play. The priority of the right over the good, or of fair opportunity over the difference principle, is not presently our concern.

As all the previous examples illustrate, the precedence of liberty means that liberty can be restricted only for the sake of liberty itself. There are two sorts of cases. The basic liberties may either be less extensive though still equal, or they may be unequal. If liberty is less extensive, the representative citizen must find this a gain for his freedom on balance; and if

26. *Politics,* bk. I, ch. II, 1253a15.

liberty is unequal, the freedom of those with the lesser liberty must be better secured. In both instances the justification proceeds by reference to the whole system of the equal liberties. These priority rules have already been noted on a number of occasions.

There is, however, a further distinction that must be made between two kinds of circumstances that justify or excuse a restriction of liberty. First a restriction can derive from the natural limitations and accidents of human life, or from historical and social contingencies. The question of the justice of these constraints does not arise. For example, even in a well-ordered society under favorable circumstances, liberty of thought and conscience is subject to reasonable regulations and the principle of participation is restricted in extent. These constraints issue from the more or less permanent conditions of political life; others are adjustments to the natural features of the human situation, as with the lesser liberty of children. In these cases the problem is to discover the just way to meet certain given limitations.

In the second kind of case, injustice already exists, either in social arrangements or in the conduct of individuals. The question here is what is the just way to answer injustice. This injustice may, of course, have many explanations, and those who act unjustly often do so with the conviction that they pursue a higher cause. The examples of intolerant and of rival sects illustrate this possibility. But men's propensity to injustice is not a permanent aspect of community life; it is greater or less depending in large part on social institutions, and in particular on whether these are just or unjust. A well-ordered society tends to eliminate or at least to control men's inclinations to injustice (see Chapters VIII–IX), and therefore warring and intolerant sects, say, are much less likely to exist, or to be a danger, once such a society is established. How justice requires us to meet injustice is a very different problem from how best to cope with the inevitable limitations and contingencies of human life.

These two kinds of cases raise several questions. It will be recalled that strict compliance is one of the stipulations of the original position; the principles of justice are chosen on the supposition that they will be generally complied with. Any failures are discounted as exceptions (§25). By putting these principles in lexical order, the parties are choosing a conception of justice suitable for favorable conditions and assuming that a just society can in due course be achieved. Arranged in this order, the principles define then a perfectly just scheme; they belong to ideal theory and set up an aim to guide the course of social reform. But even granting the soundness of these principles for this purpose, we must still ask how

well they apply to institutions under less than favorable conditions, and whether they provide any guidance for instances of injustice. The principles and their lexical order were not acknowledged with these situations in mind and so it is possible that they no longer hold.

I shall not attempt to give a systematic answer to these questions. A few special cases are taken up later (see Chapter VI). The intuitive idea is to split the theory of justice into two parts. The first or ideal part assumes strict compliance and works out the principles that characterize a well-ordered society under favorable circumstances. It develops the conception of a perfectly just basic structure and the corresponding duties and obligations of persons under the fixed constraints of human life. My main concern is with this part of the theory. Nonideal theory, the second part, is worked out after an ideal conception of justice has been chosen; only then do the parties ask which principles to adopt under less happy conditions. This division of the theory has, as I have indicated, two rather different subparts. One consists of the principles for governing adjustments to natural limitations and historical contingencies, and the other of principles for meeting injustice.

Viewing the theory of justice as a whole, the ideal part presents a conception of a just society that we are to achieve if we can. Existing institutions are to be judged in the light of this conception and held to be unjust to the extent that they depart from it without sufficient reason. The lexical ranking of the principles specifies which elements of the ideal are relatively more urgent, and the priority rules this ordering suggests are to be applied to nonideal cases as well. Thus as far as circumstances permit, we have a natural duty to remove any injustices, beginning with the most grievous as identified by the extent of the deviation from perfect justice. Of course, this idea is extremely rough. The measure of departures from the ideal is left importantly to intuition. Still our judgment is guided by the priority indicated by the lexical ordering. If we have a reasonably clear picture of what is just, our considered convictions of justice may fall more closely into line even though we cannot formulate precisely how this greater convergence comes about. Thus while the principles of justice belong to the theory of an ideal state of affairs, they are generally relevant.

The several parts of nonideal theory may be illustrated by various examples, some of which we have discussed. One type of situation is that involving a less extensive liberty. Since there are no inequalities, but all are to have a narrower rather than a wider freedom, the question can be

assessed from the perspective of the representative equal citizen. To appeal to the interests of this representative man in applying the principles of justice is to invoke the principle of the common interest. (The common good I think of as certain general conditions that are in an appropriate sense equally to everyone's advantage.) Several of the preceding examples involve a less extensive liberty: the regulation of liberty of conscience and freedom of thought in ways consistent with public order, and the limitation on the scope of majority rule belong to this category (§§34, 37). These constraints arise from the permanent conditions of human life and therefore these cases belong to that part of nonideal theory which deals with natural limitations. The two examples of curbing the liberties of the intolerant and of restraining the violence of contending sects, since they involve injustice, belong to the partial compliance part of nonideal theory. In each of these four cases, however, the argument proceeds from the viewpoint of the representative citizen. Following the idea of the lexical ordering, the limitations upon the extent of liberty are for the sake of liberty itself and result in a lesser but still equal freedom.

The second kind of case is that of an unequal liberty. If some have more votes than others, political liberty is unequal; and the same is true if the votes of some are weighted much more heavily, or if a segment of society is without the franchise altogether. In many historical situations a lesser political liberty may have been justified. Perhaps Burke's unrealistic account of representation had an element of validity in the context of eighteenth century society.[27] If so, it reflects the fact that the various liberties are not all on a par, for while at that time unequal political liberty might conceivably have been a permissible adjustment to historical limitations, serfdom and slavery, and religious intolerance, certainly were not. These constraints do not justify the loss of liberty of conscience and the rights defining the integrity of the person. The case for certain political liberties and the rights of fair equality of opportunity is less compelling. As I noted before (§11), it may be necessary to forgo part of these freedoms when this is required to transform a less fortunate society into one in which all the basic liberties can be fully enjoyed. Under conditions that cannot be changed at present, there may be no way to institute the effective exercise of these freedoms; but if possible the more central ones should be realized first. In any case, to accept the lexical ordering of the two principles we are not forced to deny that the feasibility of the basic

27. See H. F. Pitkin, *The Concept of Representation,* ch. VIII, for an account of Burke's view.

liberties depends upon circumstances. We must, however, make sure that the course of change being followed is such that social conditions will eventually be brought about under which restrictions on these freedoms are no longer justified. Their full achievement is, so to speak, the inherent long-run tendency of a just system.

In these remarks I have assumed that it is always those with the lesser liberty who must be compensated. We are always to appraise the situation from their point of view (as seen from the constitutional convention or the legislature). Now it is this restriction that makes it practically certain that slavery and serfdom, in their familiar forms anyway, are tolerable only when they relieve even worse injustices. There may be transition cases where enslavement is better than current practice. For example, suppose that city-states that previously have not taken prisoners of war but have always put captives to death agree by treaty to hold prisoners as slaves instead. Although we cannot allow the institution of slavery on the grounds that the greater gains of some outweigh the losses to others, it may be that under these conditions, since all run the risk of capture in war, this form of slavery is less unjust than present custom. At least the servitude envisaged is not hereditary (let us suppose) and it is accepted by the free citizens of more or less equal city-states. The arrangement seems defensible as an advance on established institutions, if slaves are not treated too severely. In time it will presumably be abandoned altogether, since the exchange of prisoners of war is a still more desirable arrangement, the return of the captured members of the community being preferable to the services of slaves. But none of these considerations, however fanciful, tend in any way to justify hereditary slavery or serfdom by citing natural or historical limitations. Moreover, one cannot at this point appeal to the necessity or at least to the great advantage of these servile arrangements for the higher forms of culture. As I shall argue later, the principle of perfection would be rejected in the original position (§50).

The problem of paternalism deserves some discussion here, since it has been mentioned in the argument for equal liberty, and concerns a lesser freedom. In the original position the parties assume that in society they are rational and able to manage their own affairs. Therefore they do not acknowledge any duties to self, since this is unnecessary to further their good. But once the ideal conception is chosen, they will want to insure themselves against the possibility that their powers are undeveloped and they cannot rationally advance their interests, as in the case of children; or that through some misfortune or accident they are unable to

make decisions for their good, as in the case of those seriously injured or mentally disturbed. It is also rational for them to protect themselves against their own irrational inclinations by consenting to a scheme of penalties that may give them a sufficient motive to avoid foolish actions and by accepting certain impositions designed to undo the unfortunate consequences of their imprudent behavior. For these cases the parties adopt principles stipulating when others are authorized to act in their behalf and to override their present wishes if necessary; and this they do recognizing that sometimes their capacity to act rationally for their good may fail, or be lacking altogether.[28]

Thus the principles of paternalism are those that the parties would acknowledge in the original position to protect themselves against the weakness and infirmities of their reason and will in society. Others are authorized and sometimes required to act on our behalf and to do what we would do for ourselves if we were rational, this authorization coming into effect only when we cannot look after our own good. Paternalistic decisions are to be guided by the individual's own settled preferences and interests insofar as they are not irrational, or failing a knowledge of these, by the theory of primary goods. As we know less and less about a person, we act for him as we would act for ourselves from the standpoint of the original position. We try to get for him the things he presumably wants whatever else he wants. We must be able to argue that with the development or the recovery of his rational powers the individual in question will accept our decision on his behalf and agree with us that we did the best thing for him.

The requirement that the other person in due course accepts his condition is not, however, by any means sufficient, even if this condition is not open to rational criticism. Thus imagine two persons in full possession of their reason and will who affirm different religious or philosophical beliefs; and suppose that there is some psychological process that will convert each to the other's view, despite the fact that the process is imposed on them against their wishes. In due course, let us suppose, both will come to accept conscientiously their new beliefs. We are still not permitted to submit them to this treatment. Two further stipulations are necessary: paternalistic intervention must be justified by the evident failure or absence of reason and will; and it must be guided by the principles of

28. For a discussion of this problem see Gerald Dworkin, "Paternalism," an essay in *Morality and the Law,* ed. R. A. Wasserstrom (Belmont, Calif., Wadsworth Publishing Co., 1971), pp. 107–126.

justice and what is known about the subject's more permanent aims and preferences, or by the account of primary goods. These restrictions on the initiation and direction of paternalistic measures follow from the assumptions of the original position. The parties want to guarantee the integrity of their person and their final ends and beliefs whatever these are. Paternalistic principles are a protection against our own irrationality, and must not be interpreted to license assaults on one's convictions and character by any means so long as these offer the prospect of securing consent later on. More generally, methods of education must likewise honor these constraints (§78).

The force of justice as fairness would appear to arise from two things: the requirement that all inequalities be justified to the least advantaged, and the priority of liberty. This pair of constraints distinguishes it from intuitionism and teleological theories. Taking the preceding discussion into account, we can reformulate the first principle of justice and conjoin to it the appropriate priority rule. The changes and additions are, I believe, self-explanatory. The principle now reads as follows.

FIRST PRINCIPLE

> Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.

PRIORITY RULE

> The principles of justice are to be ranked in lexical order and therefore liberty can be restricted only for the sake of liberty. There are two cases: (a) a less extensive liberty must strengthen the total system of liberty shared by all, and (b) a less than equal liberty must be acceptable to those citizens with the lesser liberty.

It perhaps bears repeating that I have yet to give a systematic argument for the priority rule, although I have checked it out in a number of important cases. It appears to fit our considered convictions fairly well. But an argument from the standpoint of the original position I postpone until Part Three when the full force of the contract doctrine can be brought into play (§82).

40. THE KANTIAN INTERPRETATION OF JUSTICE AS FAIRNESS

For the most part I have considered the content of the principle of equal liberty and the meaning of the priority of the rights that it defines. It seems appropriate at this point to note that there is a Kantian interpretation of the conception of justice from which this principle derives. This interpretation is based upon Kant's notion of autonomy. It is a mistake, I believe, to emphasize the place of generality and universality in Kant's ethics. That moral principles are general and universal is hardly new with him; and as we have seen these conditions do not in any case take us very far. It is impossible to construct a moral theory on so slender a basis, and therefore to limit the discussion of Kant's doctrine to these notions is to reduce it to triviality. The real force of his view lies elsewhere.[29]

For one thing, he begins with the idea that moral principles are the object of rational choice. They define the moral law that men can rationally will to govern their conduct in an ethical commonwealth. Moral philosophy becomes the study of the conception and outcome of a suitably defined rational decision. This idea has immediate consequences. For once we think of moral principles as legislation for a kingdom of ends, it is clear that these principles must not only be acceptable to all but public as well. Finally Kant supposes that this moral legislation is to be agreed to under conditions that characterize men as free and equal rational beings. The description of the original position is an attempt to interpret this conception. I do not wish to argue here for this interpretation on the basis of Kant's text. Certainly some will want to read him

29. Especially to be avoided is the idea that Kant's doctrine provides at best only the general, or formal, elements for a utilitarian or indeed for any other moral conception. This idea is found in Sidgwick, *The Methods of Ethics,* 7th ed. (London, Macmillan, 1907), pp. xvii and xx of the preface; and in F. H. Bradley, *Ethical Studies,* 2nd ed. (Oxford, Clarendon Press, 1927), Essay IV; and goes back at least to Hegel. One must not lose sight of the full scope of his view and take the later works into consideration. Unfortunately, there is no commentary on Kant's moral theory as a whole; perhaps it would prove impossible to write. But the standard works of H. J. Paton, *The Categorical Imperative* (Chicago, University of Chicago Press, 1948), and L. W. Beck, *A Commentary on Kant's Critique of Practical Reason* (Chicago, University of Chicago Press, 1960), and others need to be further complemented by studies of the other writings. See here M. J. Gregor's *Laws of Freedom* (Oxford, Basil Blackwell, 1963), an account of *The Metaphysics of Morals,* and J. G. Murphy's brief *Kant: The Philosophy of Right* (London, Macmillan, 1970). Beyond this, *The Critique of Judgment, Religion within the Limits of Reason,* and the political writings cannot be neglected without distorting his doctrine. For the last, see *Kant's Political Writings,* ed. Hans Reiss and trans. H. B. Nisbet (Cambridge, The University Press, 1970).

differently. Perhaps the remarks to follow are best taken as suggestions for relating justice as fairness to the high point of the contractarian tradition in Kant and Rousseau.

Kant held, I believe, that a person is acting autonomously when the principles of his action are chosen by him as the most adequate possible expression of his nature as a free and equal rational being. The principles he acts upon are not adopted because of his social position or natural endowments, or in view of the particular kind of society in which he lives or the specific things that he happens to want. To act on such principles is to act heteronomously. Now the veil of ignorance deprives the persons in the original position of the knowledge that would enable them to choose heteronomous principles. The parties arrive at their choice together as free and equal rational persons knowing only that those circumstances obtain which give rise to the need for principles of justice.

To be sure, the argument for these principles does add in various ways to Kant's conception. For example, it adds the feature that the principles chosen are to apply to the basic structure of society; and premises characterizing this structure are used in deriving the principles of justice. But I believe that this and other additions are natural enough and remain fairly close to Kant's doctrine, at least when all of his ethical writings are viewed together. Assuming, then, that the reasoning in favor of the principles of justice is correct, we can say that when persons act on these principles they are acting in accordance with principles that they would choose as rational and independent persons in an original position of equality. The principles of their actions do not depend upon social or natural contingencies, nor do they reflect the bias of the particulars of their plan of life or the aspirations that motivate them. By acting from these principles persons express their nature as free and equal rational beings subject to the general conditions of human life. For to express one's nature as a being of a particular kind is to act on the principles that would be chosen if this nature were the decisive determining element. Of course, the choice of the parties in the original position is subject to the restrictions of that situation. But when we knowingly act on the principles of justice in the ordinary course of events, we deliberately assume the limitations of the original position. One reason for doing this, for persons who can do so and want to, is to give expression to one's nature.

The principles of justice are also analogous to categorical imperatives. For by a categorical imperative Kant understands a principle of conduct that applies to a person in virtue of his nature as a free and equal rational being. The validity of the principle does not presuppose that one

has a particular desire or aim. Whereas a hypothetical imperative by contrast does assume this: it directs us to take certain steps as effective means to achieve a specific end. Whether the desire is for a particular thing, or whether it is for something more general, such as certain kinds of agreeable feelings or pleasures, the corresponding imperative is hypothetical. Its applicability depends upon one's having an aim which one need not have as a condition of being a rational human individual. The argument for the two principles of justice does not assume that the parties have particular ends, but only that they desire certain primary goods. These are things that it is rational to want whatever else one wants. Thus given human nature, wanting them is part of being rational; and while each is presumed to have some conception of the good, nothing is known about his final ends. The preference for primary goods is derived, then, from only the most general assumptions about rationality and the conditions of human life. To act from the principles of justice is to act from categorical imperatives in the sense that they apply to us whatever in particular our aims are. This simply reflects the fact that no such contingencies appear as premises in their derivation.

We may note also that the motivational assumption of mutual disinterest parallels Kant's notion of autonomy, and gives another reason for this condition. So far this assumption has been used to characterize the circumstances of justice and to provide a clear conception to guide the reasoning of the parties. We have also seen that the concept of benevolence, being a second-order notion, would not work out well. Now we can add that the assumption of mutual disinterest is to allow for freedom in the choice of a system of final ends.[30] Liberty in adopting a conception of the good is limited only by principles that are deduced from a doctrine which imposes no prior constraints on these conceptions. Presuming mutual disinterest in the original position carries out this idea. We postulate that the parties have opposing claims in a suitably general sense. If their ends were restricted in some specific way, this would appear at the outset as an arbitrary restriction on freedom. Moreover, if the parties were conceived as altruists, or as pursuing certain kinds of pleasures, then the principles chosen would apply, as far as the argument would have shown, only to persons whose freedom was restricted to choices compatible with altruism or hedonism. As the argument now runs, the principles of justice cover all persons with rational plans of life, whatever their content, and these principles represent the appropriate restrictions on freedom. Thus it

30. For this point I am indebted to Charles Fried.

is possible to say that the constraints on conceptions of the good are the result of an interpretation of the contractual situation that puts no prior limitations on what men may desire. There are a variety of reasons, then, for the motivational premise of mutual disinterest. This premise is not only a matter of realism about the circumstances of justice or a way to make the theory manageable. It also connects up with the Kantian idea of autonomy.

There is, however, a difficulty that should be clarified. It is well expressed by Sidgwick.[31] He remarks that nothing in Kant's ethics is more striking than the idea that a man realizes his true self when he acts from the moral law, whereas if he permits his actions to be determined by sensuous desires or contingent aims, he becomes subject to the law of nature. Yet in Sidgwick's opinion this idea comes to naught. It seems to him that on Kant's view the lives of the saint and the scoundrel are equally the outcome of a free choice (on the part of the noumenal self) and equally the subject of causal laws (as a phenomenal self). Kant never explains why the scoundrel does not express in a bad life his characteristic and freely chosen selfhood in the same way that a saint expresses his characteristic and freely chosen selfhood in a good one. Sidgwick's objection is decisive, I think, as long as one assumes, as Kant's exposition may seem to allow, both that the noumenal self can choose any consistent set of principles and that acting from such principles, whatever they are, is sufficient to express one's choice as that of a free and equal rational being. Kant's reply must be that though acting on any consistent set of principles could be the outcome of a decision on the part of the noumenal self, not all such action by the phenomenal self expresses this decision as that of a free and equal rational being. Thus if a person realizes his true self by expressing it in his actions, and if he desires above all else to realize this self, then he will choose to act from principles that manifest his nature as a free and equal rational being. The missing part of the argument concerns the concept of expression. Kant did not show that acting from the moral law expresses our nature in identifiable ways that acting from contrary principles does not.

This defect is made good, I believe, by the conception of the original position. The essential point is that we need an argument showing which principles, if any, free and equal rational persons would choose and these principles must be applicable in practice. A definite answer to this ques-

31. See *The Methods of Ethics,* Appendix, "The Kantian Conception of Free Will" (reprinted from *Mind,* vol. 13, 1888), pp. 511–516, esp. p. 516.

tion is required to meet Sidgwick's objection. My suggestion is that we think of the original position as in important ways similar to the point of view from which noumenal selves see the world. The parties qua noumenal selves have complete freedom to choose whatever principles they wish; but they also have a desire to express their nature as rational and equal members of the intelligible realm with precisely this liberty to choose, that is, as beings who can look at the world in this way and express this perspective in their life as members of society. They must decide, then, which principles when consciously followed and acted upon in everyday life will best manifest this freedom in their community, most fully reveal their independence from natural contingencies and social accident. Now if the argument of the contract doctrine is correct, these principles are indeed those defining the moral law, or more exactly, the principles of justice for institutions and individuals. The description of the original position resembles the point of view of noumenal selves, of what it means to be a free and equal rational being. Our nature as such beings is displayed when we act from the principles we would choose when this nature is reflected in the conditions determining the choice. Thus men exhibit their freedom, their independence from the contingencies of nature and society, by acting in ways they would acknowledge in the original position.

Properly understood, then, the desire to act justly derives in part from the desire to express most fully what we are or can be, namely free and equal rational beings with a liberty to choose. It is for this reason, I believe, that Kant speaks of the failure to act on the moral law as giving rise to shame and not to feelings of guilt. And this is appropriate, since for him acting unjustly is acting in a manner that fails to express our nature as a free and equal rational being. Such actions therefore strike at our self-respect, our sense of our own worth, and the experience of this loss is shame (§67). We have acted as though we belonged to a lower order, as though we were a creature whose first principles are decided by natural contingencies. Those who think of Kant's moral doctrine as one of law and guilt badly misunderstand him. Kant's main aim is to deepen and to justify Rousseau's idea that liberty is acting in accordance with a law that we give to ourselves. And this leads not to a morality of austere command but to an ethic of mutual respect and self-esteem.[32]

32. See B. A. O. Williams, "The Idea of Equality," in *Philosophy, Politics and Society,* Second Series, ed. Peter Laslett and W. G. Runciman (Oxford, Basil Blackwell, 1962), pp. 115f. For confirmation of this interpretation, see Kant's remarks on moral education in *The Critique of Practical Reason,* pt. II. See also Beck, *A Commentary on Kant's Critique of Practical Reason,* pp. 233–236.

The original position may be viewed, then, as a procedural interpretation of Kant's conception of autonomy and the categorical imperative within the framework of an empirical theory. The principles regulative of the kingdom of ends are those that would be chosen in this position, and the description of this situation enables us to explain the sense in which acting from these principles expresses our nature as free and equal rational persons. No longer are these notions purely transcendent and lacking explicable connections with human conduct, for the procedural conception of the original position allows us to make these ties. Of course, I have departed from Kant's views in several respects. I cannot discuss these matters here; but two points should be noted. The person's choice as a noumenal self I have assumed to be a collective one. The force of the self's being equal is that the principles chosen must be acceptable to other selves. Since all are similarly free and rational, each must have an equal say in adopting the public principles of the ethical commonwealth. This means that as noumenal selves, everyone is to consent to these principles. Unless the scoundrel's principles would be agreed to, they cannot express this free choice, however much a single self might be of a mind to adopt them. Later I shall try to define a clear sense in which this unanimous agreement is best expressive of the nature of even a single self (§85). It in no way overrides a person's interests as the collective nature of the choice might seem to imply. But I leave this aside for the present.

Secondly, I have assumed all along that the parties know that they are subject to the conditions of human life. Being in the circumstances of justice, they are situated in the world with other men who likewise face limitations of moderate scarcity and competing claims. Human freedom is to be regulated by principles chosen in the light of these natural restrictions. Thus justice as fairness is a theory of human justice and among its premises are the elementary facts about persons and their place in nature. The freedom of pure intelligences not subject to these constraints (God and the angels) are outside the range of the theory. Kant may have meant his doctrine to apply to all rational beings as such and therefore that men's social situation in the world is to have no role in determining the first principles of justice. If so, this is another difference between justice as fairness and Kant's theory.

But the Kantian interpretation is not intended as an interpretation of Kant's actual doctrine but rather of justice as fairness. Kant's view is marked by a number of deep dualisms, in particular, the dualism between the necessary and the contingent, form and content, reason and desire, and noumena and phenomena. To abandon these dualisms as he under-

stood them is, for many, to abandon what is distinctive in his theory. I believe otherwise. His moral conception has a characteristic structure that is more clearly discernible when these dualisms are not taken in the sense he gave them but recast and their moral force reformulated within the scope of an empirical theory. What I have called the Kantian interpretation indicates how this can be done.

CHAPTER V. DISTRIBUTIVE SHARES

In this chapter I take up the second principle of justice and describe an arrangement of institutions that fulfills its requirements within the setting of a modern state. I begin by noting that the principles of justice may serve as part of a doctrine of political economy. The utilitarian tradition has stressed this application and we must see how they fare in this regard. I also emphasize that these principles have embedded in them a certain ideal of social institutions, and this fact will be of importance when we consider the values of community in Part Three. As a preparation for subsequent discussions, there are some brief comments on economic systems, the role of markets, and the like. Then I turn to the difficult problem of saving and justice between generations. The essentials are put together in an intuitive way, followed by some remarks devoted to the question of time preference and to some further cases of priority. After this I try to show that the account of distributive shares can explain the place of the common sense precepts of justice. I also examine perfectionism and intuitionism as theories of distributive justice, thus rounding out to some degree the contrast with other traditional views. Throughout the choice between a private-property economy and socialism is left open; from the standpoint of the theory of justice alone, various basic structures would appear to satisfy its principles.

41. THE CONCEPT OF JUSTICE IN POLITICAL ECONOMY

My aim in this chapter is to see how the two principles work out as a conception of political economy, that is, as standards by which to assess economic arrangements and policies, and their background institutions. (Welfare economics is often defined in the same way.[1] I do not use this

1. Welfare economics is so defined by K. J. Arrow and Tibor Scitovsky in their introduction to

name because the term "welfare" suggests that the implicit moral conception is utilitarian; the phrase "social choice" is far better although I believe its connotations are still too narrow.) A doctrine of political economy must include an interpretation of the public good which is based on a conception of justice. It is to guide the reflections of the citizen when he considers questions of economic and social policy. He is to take up the perspective of the constitutional convention or the legislative stage and ascertain how the principles of justice apply. A political opinion concerns what advances the good of the body politic as a whole and invokes some criterion for the just division of social advantages.

From the beginning I have stressed that justice as fairness applies to the basic structure of society. It is a conception for ranking social forms viewed as closed systems. Some decision concerning these background arrangements is fundamental and cannot be avoided. In fact, the cumulative effect of social and economic legislation is to specify the basic structure. Moreover, the social system shapes the wants and aspirations that its citizens come to have. It determines in part the sort of persons they want to be as well as the sort of persons they are. Thus an economic system is not only an institutional device for satisfying existing wants and needs but a way of creating and fashioning wants in the future. How men work together now to satisfy their present desires affects the desires they will have later on, the kind of persons they will be. These matters are, of course, perfectly obvious and have always been recognized. They were stressed by economists as different as Marshall and Marx.[2] Since economic arrangements have these effects, and indeed must do so, the choice of these institutions involves some view of human good and of the design of institutions to realize it. This choice must, therefore, be made on moral and political as well as on economic grounds. Considerations of efficiency are but one basis of decision and often relatively minor at that. Of course, this decision may not be openly faced; it may be made by default. We often acquiesce without thinking in the moral and political conception implicit in the status quo, or leave things to be settled by how contending social and economic forces happen to work themselves out. But political

Readings in Welfare Economics (Homewood, Ill., Richard D. Irwin, 1969), p. 1. For further discussion, see Abram Bergson, *Essays in Normative Economics* (Cambridge, Harvard University Press, 1966), pp. 35–39, 60–63, 68f; and A. K. Sen, *Collective Choice and Social Welfare* (San Francisco, Holden-Day, 1970), pp. 56–59.

2. For a discussion of this point and its consequences for political principles, see Brian Barry, *Political Argument* (London, Routledge and Kegan Paul, 1965), pp. 75–79.

economy must investigate this problem even if the conclusion reached is that it is best left to the course of events to decide.

Now it may seem at first sight that the influence of the social system upon human wants and men's view of themselves poses a decisive objection to the contract view. One might think that this conception of justice relies upon the aims of existing individuals and regulates the social order by principles that persons guided by these aims would choose. How, then, can this doctrine determine an Archimedean point from which the basic structure itself can be appraised? It might seem as if there is no alternative but to judge institutions in the light of an ideal conception of the person arrived at on perfectionist or on a priori grounds. But, as the account of the original position and its Kantian interpretation makes clear, we must not overlook the very special nature of that situation and the scope of the principles adopted there. Only the most general assumptions are made about the aims of the parties, namely, that they take an interest in primary social goods, in things that men are presumed to want whatever else they want. To be sure, the theory of these goods depends on psychological premises and these may prove incorrect. But the idea at any rate is to define a class of goods that are normally wanted as parts of rational plans of life which may include the most varied sorts of ends. To suppose, then, that the parties want these goods, and to found a conception of justice on this presumption, is not to tie it to a particular pattern of human interests as these might be generated by a particular arrangement of institutions. The theory of justice does, indeed, presuppose a theory of the good, but within wide limits this does not prejudge the choice of the sort of persons that men want to be.

Once the principles of justice are derived, however, the contract doctrine does establish certain limits on the conception of the good. These limits follow from the priority of justice over efficiency and the priority of liberty over social and economic advantages (assuming that serial order obtains). For as I remarked earlier (§6), these priorities mean that desires for things that are inherently unjust, or that cannot be satisfied except by the violation of just arrangements, have no weight. There is no value in fulfilling these wants and the social system should discourage them. Further, one must take into account the problem of stability. A just system must generate its own support. This means that it must be arranged so as to bring about in its members the corresponding sense of justice, an effective desire to act in accordance with its rules for reasons of justice. Thus the requirement of stability and the criterion of discour-

aging desires that conflict with the principles of justice put further constraints on institutions. They must be not only just but framed so as to encourage the virtue of justice in those who take part in them. In this sense, the principles of justice define a partial ideal of the person which social and economic arrangements must respect. Finally, as the argument for embedding ideals into our working principles has brought out, certain institutions are required by the two principles. They define an ideal basic structure, or the outlines of one, toward which the course of reform should evolve.

The upshot of these considerations is that justice as fairness is not at the mercy, so to speak, of existing wants and interests. It sets up an Archimedean point for assessing the social system without invoking a priori considerations. The long range aim of society is settled in its main lines irrespective of the particular desires and needs of its present members. And an ideal conception of justice is defined since institutions are to foster the virtue of justice and to discourage desires and aspirations incompatible with it. Of course, the pace of change and the particular reforms called for at any given time depend upon current conditions. But the conception of justice, the general form of a just society and the ideal of the person consistent with it are not similarly dependent. There is no place for the question whether men's desires to play the role of superior or inferior might not be so great that autocratic institutions should be accepted, or whether men's perception of the religious practices of others might not be so upsetting that liberty of conscience should not be allowed. We have no occasion to ask whether under reasonably favorable conditions the economic gains of technocratic but authoritarian institutions might be so great as to justify the sacrifice of basic freedoms. Of course, these remarks assume that the general assumptions on which the principles of justice were chosen are correct. But if they are, this sort of question is already decided by these principles. Certain institutional forms are embedded within the conception of justice. This view shares with perfectionism the feature of setting up an ideal of the person that constrains the pursuit of existing desires. In this respect justice as fairness and perfectionism are both opposed to utilitarianism.

Now it may appear that since utilitarianism makes no distinctions between the quality of desires and all satisfactions have some value, it has no criteria for choosing between systems of desires, or ideals of the person. From a theoretical point of view anyway, this is incorrect. The utilitarian can always say that given social conditions and men's interests as

they are, and taking into account how they will develop under this or that alternative institutional arrangement, encouraging one pattern of wants rather than another is likely to lead to a greater net balance (or to a higher average) of satisfaction. On this basis the utilitarian selects between ideals of the person. Some attitudes and desires, being less compatible with fruitful social cooperation, tend to reduce the total (or the average) happiness. Roughly speaking, the moral virtues are those dispositions and effective desires that can generally be relied upon to promote the greatest sum of well-being. Thus, it would be a mistake to claim that the principle of utility provides no grounds for choosing among ideals of the person, however difficult it may be to apply the principle in practice. Nevertheless, the choice does depend upon existing desires and present social circumstances and their natural continuations into the future. These initial conditions may heavily influence the conception of human good that should be encouraged. The contrast is that both justice as fairness and perfectionism establish independently an ideal conception of the person and of the basic structure so that not only are some desires and inclinations necessarily discouraged but the effect of the initial circumstances will eventually disappear. With utilitarianism we cannot be sure what will happen. Since there is no ideal embedded in its first principle, the place we start from may always influence the path we are to follow.

By way of summing up, the essential point is that despite the individualistic features of justice as fairness, the two principles of justice are not contingent upon existing desires or present social conditions. Thus we are able to derive a conception of a just basic structure, and an ideal of the person compatible with it, that can serve as a standard for appraising institutions and for guiding the overall direction of social change. In order to find an Archimedean point it is not necessary to appeal to a priori or perfectionist principles. By assuming certain general desires, such as the desire for primary social goods, and by taking as a basis the agreements that would be made in a suitably defined initial situation, we can achieve the requisite independence from existing circumstances. The original position is so characterized that unanimity is possible; the deliberations of any one person are typical of all. Moreover, the same will hold for the considered judgments of the citizens of a well-ordered society effectively regulated by the principles of justice. Everyone has a similar sense of justice and in this respect a well-ordered society is homogeneous. Political argument appeals to this moral consensus.

It may be thought that the assumption of unanimity is peculiar to

the political philosophy of idealism.[3] As it is used in the contract view, however, there is nothing characteristically idealist about the supposition of unanimity. This condition is part of the procedural conception of the original position and it represents a constraint on arguments. In this way it shapes the content of the theory of justice, the principles that are to match our considered judgments. Hume and Adam Smith likewise assume that if men were to take up a certain point of view, that of the impartial spectator, they would be led to similar convictions. A utilitarian society may also be well-ordered. For the most part the philosophical tradition, including intuitionism, has assumed that there exists some appropriate perspective from which unanimity on moral questions may be hoped for, at least among rational persons with relevantly similar and sufficient information. Or if unanimity is impossible, disparities between judgments are greatly reduced once this standpoint is adopted. Different moral theories arise from different interpretations of this point of view, of what I have called the initial situation. In this sense the idea of unanimity among rational persons is implicit throughout the tradition of moral philosophy.

What distinguishes justice as fairness is how it characterizes the initial situation, the setting in which the condition of unanimity appears. Since the original position can be given a Kantian interpretation, this conception of justice does indeed have affinities with idealism. Kant sought to give a philosophical foundation to Rousseau's idea of the general will. The theory of justice in turn tries to present a natural procedural rendering of Kant's conception of the kingdom of ends, and of the notions of autonomy and the categorical imperative (§40). In this way the underlying structure of Kant's doctrine is detached from its metaphysical surroundings so that it can be seen more clearly and presented relatively free from objection.

There is another resemblance to idealism: justice as fairness has a central place for the value of community, and how this comes about depends upon the Kantian interpretation. I discuss this topic in Part Three. The essential idea is that we want to account for the social values, for the intrinsic good of institutional, community, and associative activities, by a conception of justice that in its theoretical basis is individualistic. For reasons of clarity among others, we do not want to rely on an undefined

3. This suggestion is found in K. J. Arrow, *Social Choice and Individual Values,* 2nd ed. (New York, John Wiley and Sons, 1963), pp. 74f, 81–86.

concept of community, or to suppose that society is an organic whole with a life of its own distinct from and superior to that of all its members in their relations with one another. Thus the contractual conception of the original position is worked out first. It is reasonably simple and the problem of rational choice that it poses is relatively precise. From this conception, however individualistic it might seem, we must eventually explain the value of community. Otherwise the theory of justice cannot succeed. To accomplish this we shall need an account of the primary good of self-respect which relates it to the parts of the theory already developed. But for the time being, I shall leave these problems aside and proceed to consider some further implications of the two principles of justice for the economic aspects of the basic structure.

42. SOME REMARKS ABOUT ECONOMIC SYSTEMS

It is essential to keep in mind that our topic is the theory of justice and not economics, however elementary. We are only concerned with some moral problems of political economy. For example, I shall ask: what is the proper rate of saving over time, how should the background institutions of taxation and property be arranged, or at what level is the social minimum to be set? In asking these questions my intention is not to explain, much less to add anything to, what economic theory says about the working of these institutions. Attempting to do this here would obviously be out of place. Certain elementary parts of economic theory are brought in solely to illustrate the content of the principles of justice. If economic theory is used incorrectly or if the received doctrine is itself mistaken, I hope that for the purposes of the theory of justice no harm is done. But as we have seen, ethical principles depend upon general facts and therefore a theory of justice for the basic structure presupposes an account of these arrangements. It is necessary to make some assumptions and to spell out their consequences if we are to test moral conceptions. These assumptions are bound to be inaccurate and oversimplified, but this may not matter too much if they enable us to uncover the content of the principles of justice and we are satisfied that under a wide range of circumstances the difference principle will lead to acceptable conclusions. In short, questions of political economy are discussed simply to find out the practicable bearing of justice as fairness. I discuss these matters from the point of view of the citizen who is trying to organize his judgments concerning the justice of economic institutions.

In order to avoid misunderstandings and to indicate some of the main problems, I shall begin with a few remarks about economic systems. Political economy is importantly concerned with the public sector and the proper form of the background institutions that regulate economic activity, with taxation and the rights of property, the structure of markets, and so on. An economic system regulates what things are produced and by what means, who receives them and in return for which contributions, and how large a fraction of social resources is devoted to saving and to the provision of public goods. Ideally all of these matters should be arranged in ways that satisfy the two principles of justice. But we have to ask whether this is possible and what in particular these principles require.

To begin with, it is helpful to distinguish between two aspects of the public sector; otherwise the difference between a private-property economy and socialism is left unclear. The first aspect has to do with the ownership of the means of production. The classical distinction is that the size of the public sector under socialism (as measured by the fraction of total output produced by state-owned firms and managed either by state officials or by workers' councils) is much larger. In a private-property economy the number of publicly owned firms is presumably small and in any event limited to special cases such as public utilities and transportation.

A second quite different feature of the public sector is the proportion of total social resources devoted to public goods. The distinction between public and private goods raises a number of intricate points, but the main idea is that a public good has two characteristic features, indivisibility and publicness.[4] That is, there are many individuals, a public so to speak, who want more or less of this good, but if they are to enjoy it at all must each enjoy the same amount. The quantity produced cannot be divided up as private goods can and purchased by individuals according to their preferences for more and less. There are various kinds of public goods depending upon their degree of indivisibility and the size of the relevant public. The polar case of a public good is full indivisibility over the whole society. A standard example is the defense of the nation against (unjustified) foreign attack. All citizens must be provided with this good in the same amount; they cannot be given varying protection depending on their wishes. The consequence of indivisibility and publicness in these

4. For a discussion of public goods, see J. M. Buchanan, *The Demand and Supply of Public Goods* (Chicago, Rand McNally, 1968), esp. ch. IX. This work contains useful bibliographical appendixes to the literature.

cases is that the provision of public goods must be arranged for through the political process and not through the market. Both the amount to be produced and its financing need to be worked out by legislation. Since there is no problem of distribution in the sense that all citizens receive the same quantity, distribution costs are zero.

Various features of public goods derive from these two characteristics. First of all, there is the free-rider problem.[5] Where the public is large and includes many individuals, there is a temptation for each person to try to avoid doing his share. This is because whatever one man does his action will not significantly affect the amount produced. He regards the collective action of others as already given one way or the other. If the public good is produced his enjoyment of it is not decreased by his not making a contribution. If it is not produced his action would not have changed the situation anyway. A citizen receives the same protection from foreign invasion regardless of whether he has paid his taxes. Therefore in the polar case trade and voluntary agreements cannot be expected to develop.

It follows that arranging for and financing public goods must be taken over by the state and some binding rule requiring payment must be enforced. Even if all citizens were willing to pay their share, they would presumably do so only when they are assured that others will pay theirs as well. Thus once citizens have agreed to act collectively and not as isolated individuals taking the actions of the others as given, there is still the task of tying down the agreement. The sense of justice leads us to promote just schemes and to do our share in them when we believe that others, or sufficiently many of them, will do theirs. But in normal circumstances a reasonable assurance in this regard can only be given if there is a binding rule effectively enforced. Assuming that the public good is to everyone's advantage, and one that all would agree to arrange for, the use of coercion is perfectly rational from each man's point of view. Many of the traditional activities of government, insofar as they can be justified, can be accounted for in this way.[6] The need for the enforcement of rules by the state will still exist even when everyone is moved by the same sense of justice. The characteristic features of essential public goods necessitate collective agreements, and firm assurance must be given to all that they will be honored.

5. See Buchanan, ch. V; and also Mancur Olson, *The Logic of Collective Action* (Cambridge, Harvard University Press, 1965), chs. I and II, where the problem is discussed in connection with the theory of organizations.

6. See W. J. Baumol, *Welfare Economics and the Theory of the State* (London, Longmans, Green, 1952), chs. I, VII-IX, XII.

Another aspect of the public goods situation is that of externality. When goods are public and indivisible, their production will cause benefits and losses to others which may not be taken into account by those who arrange for these goods or who decide to produce them. Thus in the polar case, if but a part of the citizenry pays taxes to cover the expenditure on public goods, the whole society is still affected by the items provided. Yet those who agree to these levies may not consider these effects, and so the amount of public expenditure is presumably different from what it would be if all benefits and losses had been considered. The everyday cases are those where the indivisibility is partial and the public is smaller. Someone who has himself inoculated against a contagious disease helps others as well as himself; and while it may not pay him to obtain this protection, it may be worth it to the local community when all advantages are tallied up. And, of course, there are the striking cases of public harms, as when industries sully and erode the natural environment. These costs are not normally reckoned with by the market, so that the commodities produced are sold at much less than their marginal social costs. There is a divergence between private and social accounting that the market fails to register. One essential task of law and government is to institute the necessary corrections.

It is evident, then, that the indivisibility and publicness of certain essential goods, and the externalities and temptations to which they give rise, necessitate collective agreements organized and enforced by the state. That political rule is founded solely on men's propensity to self-interest and injustice is a superficial view. For even among just men, once goods are indivisible over large numbers of individuals, their actions decided upon in isolation from one another will not lead to the general good. Some collective arrangement is necessary and everyone wants assurance that it will be adhered to if he is willingly to do his part. In a large community the degree of mutual confidence in one another's integrity that renders enforcement superfluous is not to be expected. In a well-ordered society the required sanctions are no doubt mild and they may never be applied. Still, the existence of such devices is a normal condition of human life even in this case.

In these remarks I have distinguished between the problems of isolation and assurance.[7] The first sort of problem arises whenever the outcome of the many individuals' decisions made in isolation is worse for

7. This distinction is from A. K. Sen, "Isolation, Assurance and the Social Rate of Discount," *Quarterly Journal of Economics,* vol. 81 (1967).

everyone than some other course of action, even though, taking the conduct of the others as given, each person's decision is perfectly rational. This is simply the general case of the prisoner's dilemma of which Hobbes's state of nature is the classical example.[8] The isolation problem is to identify these situations and to ascertain the binding collective undertaking that would be best from the standpoint of all. The assurance problem is different. Here the aim is to assure the cooperating parties that the common agreement is being carried out. Each person's willingness to contribute is contingent upon the contribution of the others. Therefore to maintain public confidence in the scheme that is superior from everyone's point of view, or better anyway than the situation that would obtain in its absence, some device for administering fines and penalties must be established. It is here that the mere existence of an effective sovereign, or even the general belief in his efficacy, has a crucial role.

A final point about public goods. Since the proportion of social resources devoted to their production is distinct from the question of public ownership of the means of production, there is no necessary connection

8. The prisoner's dilemma (attributed to A. W. Tucker) is an illustration of a two-person noncooperative, nonzero-sum game; noncooperative because agreements are not binding (or enforceable), and nonzero-sum because it is not the case that what one person gains the other loses. Thus imagine two prisoners who are brought before the attorney general and interrogated separately. They both know that if neither confesses, they will receive a short sentence for a lesser offense and spend a year in prison; but that if one confesses and turns state's evidence, he will be released, the other receiving a particularly heavy term of ten years; if both confess each gets five years. In this situation, assuming mutually disinterested motivation, the most reasonable course of action for them—that neither should confess—is unstable. This can be seen from the following gain-and-loss table (with entries representing years in prison):

	Second Prisoner	
First Prisoner	not confess	confess
not confess	1, 1	10, 0
confess	0, 10	5, 5

To protect himself, if not to try to further his own interests, each has a sufficient motive to confess, whatever the other does. Rational decisions from the point of view of each lead to a situation where both prisoners are worse off.

The problem clearly is to find some means of stabilizing the best plan. We may note that if it were shared knowledge between the prisoners that they were either utilitarians, or affirmed the principles of justice (with restricted applications to prisoners), their problem would be solved. Both views in this case support the most sensible arrangement. For a discussion of these matters in connection with the theory of the state, see W. J. Baumol as cited in note 6 above. For an account of the prisoner's dilemma game, see R. D. Luce and Howard Raiffa, *Games and Decisions* (New York, John Wiley and Sons, 1957), ch. V, esp. pp. 94–102. D. P. Gauthier, "Morality and Advantage," *Philosophical Review,* vol. 76 (1967), treats the problem from the standpoint of moral philosophy.

between the two. A private-property economy may allocate a large fraction of national income to these purposes, a socialist society a small one, and vice versa. There are public goods of many kinds, ranging from military equipment to health services. Having agreed politically to allocate and to finance these items, the government may purchase them from the private sector or from publicly owned firms. The particular list of public goods produced and the procedures taken to limit public harms depend upon the society in question. It is a question not of institutional logic but of political sociology, including under this heading the way in which institutions affect the balance of political advantages.

Having considered briefly two aspects of the public sector, I should like to conclude with a few comments about the extent to which economic arrangements may rely upon a system of markets in which prices are freely determined by supply and demand. Several cases need to be distinguished. All regimes will normally use the market to ration out the consumption goods actually produced. Any other procedure is administratively cumbersome, and rationing and other devices will be resorted to only in special cases. But in a free market system the output of commodities is also guided as to kind and quantity by the preferences of households as shown by their purchases on the market. Goods fetching a greater than normal profit will be produced in larger amounts until the excess is reduced. In a socialist regime planners' preferences or collective decisions often have a larger part in determining the direction of production. Both private-property and socialist systems normally allow for the free choice of occupation and of one's place of work. It is only under command systems of either kind that this freedom is overtly interfered with.

Finally, a basic feature is the extent to which the market is used to decide the rate of saving and the direction of investment, as well as the fraction of national wealth devoted to conservation and to the elimination of irremediable injuries to the welfare of future generations. Here there are a number of possibilities. A collective decision may determine the rate of saving while the direction of investment is left largely to individual firms competing for funds. In both a private-property as well as in a socialist society great concern may be expressed for preventing irreversible damages and for husbanding natural resources and preserving the environment. But again either one may do rather badly.

It is evident, then, that there is no essential tie between the use of free markets and private ownership of the instruments of production. The idea

that competitive prices under normal conditions are just or fair goes back at least to medieval times.[9] While the notion that a market economy is in some sense the best scheme has been most carefully investigated by so-called bourgeois economists, this connection is a historical contingency in that, theoretically at least, a socialist regime can avail itself of the advantages of this system.[10] One of these advantages is efficiency. Under certain conditions competitive prices select the goods to be produced and allocate resources to their production in such a manner that there is no way to improve upon either the choice of productive methods by firms, or the distribution of goods that arises from the purchases of households. There exists no rearrangement of the resulting economic configuration that makes one household better off (in view of its preferences) without making another worse off. No further mutually advantageous trades are possible; nor are there any feasible productive processes that will yield more of some desired commodity without requiring a cutback in another. For if this were not so, the situation of some individuals could be made more advantageous without a loss for anyone else. The theory of general equilibrium explains how, given the appropriate conditions, the information supplied by prices leads economic agents to act in ways that sum up to achieve this outcome. Perfect competition is a perfect procedure with respect to efficiency.[11] Of course, the requisite conditions are highly special ones and they are seldom if ever fully satisfied in the real world. Moreover, market failures and imperfections are often serious, and compensating adjustments must be made by the allocation branch (see §43). Monopolistic restrictions, lack of information, external economies and diseconomies, and the like must be recognized and corrected. And the market fails altogether in the case of public goods. But these matters need not concern us here. These idealized arrangements are mentioned in order to clarify the related notion of pure procedural justice. The ideal conception may then be used to appraise existing arrangements and as a framework for identifying the changes that should be undertaken.

A further and more significant advantage of a market system is that, given the requisite background institutions, it is consistent with equal

9. See Mark Blaug, *Economic Theory in Retrospect,* revised edition (Homewood, Ill., Richard D. Irwin, 1968), pp. 31f. See the bibliography, pp. 36f, esp. the articles by R. A. deRoover.

10. For a discussion of this matter, with references to the literature, see Abram Bergson, "Market Socialism Revisited," *Journal of Political Economy,* vol. 75 (1967). See also Jaroslav Vanek, *The General Theory of a Labor Managed Economy* (Ithaca, Cornell University Press, 1970).

11. On the efficiency of competition, see W. J. Baumol, *Economic Theory and Operations Analysis,* 2nd ed. (Englewood Cliffs, N.J., Prentice-Hall, 1965), pp. 355–371; and T. C. Koopmans, *Three Essays on the State of Economic Science* (New York, McGraw-Hill, 1957), the first essay.

liberties and fair equality of opportunity. Citizens have a free choice of careers and occupations. There is no reason at all for the forced and central direction of labor. Indeed, in the absence of some differences in earnings as these arise in a competitive scheme, it is hard to see how, under ordinary circumstances anyway, certain aspects of a command society inconsistent with liberty can be avoided. Moreover, a system of markets decentralizes the exercise of economic power. Whatever the internal nature of firms, whether they are privately or state owned, or whether they are run by entrepreneurs or by managers elected by workers, they take the prices of outputs and inputs as given and draw up their plans accordingly. When markets are truly competitive, firms do not engage in price wars or other contests for market power. In conformity with political decisions reached democratically, the government regulates the economic climate by adjusting certain elements under its control, such as the overall amount of investment, the rate of interest, and the quantity of money, and so on. There is no necessity for comprehensive direct planning. Individual households and firms are free to make their decisions independently, subject to the general conditions of the economy.

In noting the consistency of market arrangements with socialist institutions, it is essential to distinguish between the allocative and the distributive functions of prices. The former is connected with their use to achieve economic efficiency, the latter with their determining the income to be received by individuals in return for what they contribute. It is perfectly consistent for a socialist regime to establish an interest rate to allocate resources among investment projects and to compute rental charges for the use of capital and scarce natural assets such as land and forests. Indeed, this must be done if these means of production are to be employed in the best way. For even if these assets should fall out of the sky without human effort, they are nevertheless productive in the sense that when combined with other factors a greater output results. It does not follow, however, that there need be private persons who as owners of these assets receive the monetary equivalents of these evaluations. Rather these accounting prices are indicators for drawing up an efficient schedule of economic activities. Except in the case of work of all kinds, prices under socialism do not correspond to income paid over to private individuals. Instead, the income imputed to natural and collective assets accrues to the state, and therefore their prices have no distributive function.[12]

12. For the distinction between the allocative and the distributive functions of prices, see J. E.

It is necessary, then, to recognize that market institutions are common to both private-property and socialist regimes, and to distinguish between the allocative and the distributive function of prices. Since under socialism the means of production and natural resources are publicly owned, the distributive function is greatly restricted, whereas a private-property system uses prices in varying degrees for both purposes. Which of these systems and the many intermediate forms most fully answers to the requirements of justice cannot, I think, be determined in advance. There is presumably no general answer to this question, since it depends in large part upon the traditions, institutions, and social forces of each country, and its particular historical circumstances. The theory of justice does not include these matters. But what it can do is to set out in a schematic way the outlines of a just economic system that admits of several variations. The political judgment in any given case will then turn on which variation is most likely to work out best in practice. A conception of justice is a necessary part of any such political assessment, but it is not sufficient.

The ideal scheme sketched in the next several sections makes considerable use of market arrangements. It is only in this way, I believe, that the problem of distribution can be handled as a case of pure procedural justice. Further, we also gain the advantages of efficiency and protect the important liberty of free choice of occupation. At the start I assume that the regime is a property-owning democracy since this case is likely to be better known.[13] But, as I have noted, this is not intended to prejudge the choice of regime in particular cases. Nor, of course, does it imply that actual societies which have private ownership of the means of production are not afflicted with grave injustices. Because there exists an ideal property-owning system that would be just does not imply that historical forms are just, or even tolerable. And, of course, the same is true of socialism.

43. BACKGROUND INSTITUTIONS FOR DISTRIBUTIVE JUSTICE

The main problem of distributive justice is the choice of a social system. The principles of justice apply to the basic structure and regulate how its major institutions are combined into one scheme. Now, as we have seen,

---

Meade, *Efficiency, Equality and the Ownership of Property* (London, George Allen and Unwin, 1964), pp. 11–26.

13. The term "property-owning democracy" is from Meade, ibid., the title of ch. V.

the idea of justice as fairness is to use the notion of pure procedural justice to handle the contingencies of particular situations. The social system is to be designed so that the resulting distribution is just however things turn out. To achieve this end it is necessary to set the social and economic process within the surroundings of suitable political and legal institutions. Without an appropriate scheme of these background institutions the outcome of the distributive process will not be just. Background fairness is lacking. I shall give a brief description of these supporting institutions as they might exist in a properly organized democratic state that allows private ownership of capital and natural resources. These arrangements are familiar, but it may be useful to see how they fit the two principles of justice. Modifications for the case of a socialist regime will be considered briefly later.

First of all, I assume that the basic structure is regulated by a just constitution that secures the liberties of equal citizenship (as described in the preceding chapter). Liberty of conscience and freedom of thought are taken for granted, and the fair value of political liberty is maintained. The political process is conducted, as far as circumstances permit, as a just procedure for choosing between governments and for enacting just legislation. I assume also that there is fair (as opposed to formal) equality of opportunity. This means that in addition to maintaining the usual kinds of social overhead capital, the government tries to insure equal chances of education and culture for persons similarly endowed and motivated either by subsidizing private schools or by establishing a public school system. It also enforces and underwrites equality of opportunity in economic activities and in the free choice of occupation. This is achieved by policing the conduct of firms and private associations and by preventing the establishment of monopolistic restrictions and barriers to the more desirable positions. Finally, the government guarantees a social minimum either by family allowances and special payments for sickness and employment, or more systematically by such devices as a graded income supplement (a so-called negative income tax).

In establishing these background institutions the government may be thought of as divided into four branches.[14] Each branch consists of various agencies, or activities thereof, charged with preserving certain social and economic conditions. These divisions do not overlap with the usual organization of government but are to be understood as different func-

14. For the idea of branches of government, see R. A. Musgrave, *The Theory of Public Finance* (New York, McGraw-Hill, 1959), ch. I.

tions. The allocation branch, for example, is to keep the price system workably competitive and to prevent the formation of unreasonable market power. Such power does not exist as long as markets cannot be made more competitive consistent with the requirements of efficiency and the facts of geography and the preferences of households. The allocation branch is also charged with identifying and correcting, say by suitable taxes and subsidies and by changes in the definition of property rights, the more obvious departures from efficiency caused by the failure of prices to measure accurately social benefits and costs. To this end suitable taxes and subsidies may be used, or the scope and definition of property rights may be revised. The stabilization branch, on the other hand, strives to bring about reasonably full employment in the sense that those who want work can find it and the free choice of occupation and the deployment of finance are supported by strong effective demand. These two branches together are to maintain the efficiency of the market economy generally.

The social minimum is the responsibility of the transfer branch. Later on I shall consider at what level the minimum should be set; but for the moment a few general remarks will suffice. The essential idea is that the workings of this branch take needs into account and assign them an appropriate weight with respect to other claims. A competitive price system gives no consideration to needs and therefore it cannot be the sole device of distribution. There must be a division of labor between the parts of the social system in answering to the common sense precepts of justice. Different institutions meet different claims. Competitive markets properly regulated secure free choice of occupation and lead to an efficient use of resources and allocation of commodities to households. They set a weight on the conventional precepts associated with wages and earnings, whereas the transfer branch guarantees a certain level of well-being and honors the claims of need. Eventually I will discuss these common sense precepts and how they arise within the context of various institutions. The relevant point here is that certain precepts tend to be associated with specific institutions. It is left to the background system as a whole to determine how these precepts are balanced. Since the principles of justice regulate the whole structure, they also regulate the balance of precepts. In general, then, this balance will vary in accordance with the underlying political conception.

It is clear that the justice of distributive shares depends on the background institutions and how they allocate total income, wages and other income plus transfers. There is with reason strong objection to the competitive determination of total income, since this ignores the claims of

need and an appropriate standard of life. From the standpoint of the legislative stage it is rational to insure oneself and one's descendants against these contingencies of the market. Indeed, the difference principle presumably requires this. But once a suitable minimum is provided by transfers, it may be perfectly fair that the rest of total income be settled by the price system, assuming that it is moderately efficient and free from monopolistic restrictions, and unreasonable externalities have been eliminated. Moreover, this way of dealing with the claims of need would appear to be more effective than trying to regulate income by minimum wage standards, and the like. It is better to assign to each branch only such tasks as are compatible with one another. Since the market is not suited to answer the claims of need, these should be met by a separate arrangement. Whether the principles of justice are satisfied, then, turns on whether the total income of the least advantaged (wages plus transfers) is such as to maximize their long-run expectations (consistent with the constraints of equal liberty and fair equality of opportunity).

Finally, there is a distribution branch. Its task is to preserve an approximate justice in distributive shares by means of taxation and the necessary adjustments in the rights of property. Two aspects of this branch may be distinguished. First of all, it imposes a number of inheritance and gift taxes, and sets restrictions on the rights of bequest. The purpose of these levies and regulations is not to raise revenue (release resources to government) but gradually and continually to correct the distribution of wealth and to prevent concentrations of power detrimental to the fair value of political liberty and fair equality of opportunity. For example, the progressive principle might be applied at the beneficiary's end.[15] Doing this would encourage the wide dispersal of property which is a necessary condition, it seems, if the fair value of the equal liberties is to be maintained. The unequal inheritance of wealth is no more inherently unjust than the unequal inheritance of intelligence. It is true that the former is presumably more easily subject to social control; but the essential thing is that as far as possible inequalities founded on either should satisfy the difference principle. Thus inheritance is permissible provided that the resulting inequalities are to the advantage of the least fortunate and compatible with liberty and fair equality of opportunity. As earlier defined, fair equality of opportunity means a certain set of institutions that assures similar chances of education and culture for persons similarly motivated and keeps positions and offices open to all on the basis of qualities and

15. See Meade, *Efficiency, Equality and the Ownership of Property,* pp. 56f.

efforts reasonably related to the relevant duties and tasks. It is these institutions that are put in jeopardy when inequalities of wealth exceed a certain limit; and political liberty likewise tends to lose its value, and representative government to become such in appearance only. The taxes and enactments of the distribution branch are to prevent this limit from being exceeded. Naturally, where this limit lies is a matter of political judgment guided by theory, good sense, and plain hunch, at least within a wide range. On this sort of question the theory of justice has nothing specific to say. Its aim is to formulate the principles that are to regulate the background institutions.

The second part of the distribution branch is a scheme of taxation to raise the revenues that justice requires. Social resources must be released to the government so that it can provide for the public goods and make the transfer payments necessary to satisfy the difference principle. This problem belongs to the distribution branch since the burden of taxation is to be justly shared and it aims at establishing just arrangements. Leaving aside many complications, it is worth noting that a proportional expenditure tax may be part of the best tax scheme.[16] For one thing, it is preferable to an income tax (of any kind) at the level of common sense precepts of justice, since it imposes a levy according to how much a person takes out of the common store of goods and not according to how much he contributes (assuming here that income is fairly earned). Again, a proportional tax on total consumption (for each year say) can contain the usual exemptions for dependents, and so on; and it treats everyone in a uniform way (still assuming that income is fairly earned). It may be better, therefore, to use progressive rates only when they are necessary to preserve the justice of the basic structure with respect to the first principle of justice and fair equality of opportunity, and so to forestall accumulations of property and power likely to undermine the corresponding institutions. Following this rule might help to signal an important distinction in questions of policy. And if proportional taxes should also prove more efficient, say because they interfere less with incentives, this might make the case for them decisive if a feasible scheme could be worked out. As before, these are questions of political judgment and not part of a theory of justice. And in any case we are here considering such a proportional tax as part of an ideal scheme for a well-ordered society in order to illustrate the content of the two principles. It does not follow that, given the injustice of existing institutions, even steeply progressive income taxes are not

16. See Nicholas Kaldor, *An Expenditure Tax* (London, George Allen and Unwin, 1955).

justified when all things are considered. In practice we must usually choose between several unjust, or second best, arrangements; and then we look to nonideal theory to find the least unjust scheme. Sometimes this scheme will include measures and policies that a perfectly just system would reject. Two wrongs can make a right in the sense that the best available arrangement may contain a balance of imperfections, an adjustment of compensating injustices.

The two parts of the distribution branch derive from the two principles of justice. The taxation of inheritance and income at progressive rates (when necessary), and the legal definition of property rights, are to secure the institutions of equal liberty in a property-owning democracy and the fair value of the rights they establish. Proportional expenditure (or income) taxes are to provide revenue for public goods, the transfer branch and the establishment of fair equality of opportunity in education, and the like, so as to carry out the second principle. No mention has been made at any point of the traditional criteria of taxation such as that taxes are to be levied according to benefits received or the ability to pay.[17] The reference to common sense precepts in connection with expenditure taxes is a subordinate consideration. The scope of these criteria is regulated by the principles of justice. Once the problem of distributive shares is recognized as that of designing background institutions, the conventional maxims are seen to have no independent force, however appropriate they may be in certain delimited cases. To suppose otherwise is not to take a sufficiently comprehensive point of view (see §47 below). It is evident also that the design of the distribution branch does not presuppose the utilitarian's standard assumptions about individual utilities. Inheritance and progressive income taxes, for example, are not predicated on the idea that individuals have similar utility functions satisfying the diminishing marginal principle. The aim of the distribution branch is not, of course, to maximize the net balance of satisfaction but to establish just background institutions. Doubts about the shape of utility functions are irrelevant. This problem is one for the utilitarian, not for contract theory.

So far I have assumed that the aim of the branches of government is to establish a democratic regime in which land and capital are widely though not presumably equally held. Society is not so divided that one fairly small sector controls the preponderance of productive resources. When this is achieved and distributive shares satisfy the principles of

17. For a discussion of these tax criteria, see Musgrave, *The Theory of Public Finance,* chs. IV and V.

justice, many socialist criticisms of the market economy are met. But it is clear that, in theory anyway, a liberal socialist regime can also answer to the two principles of justice. We have only to suppose that the means of production are publicly owned and that firms are managed by workers' councils say, or by agents appointed by them. Collective decisions made democratically under the constitution determine the general features of the economy, such as the rate of saving and the proportion of society's production devoted to essential public goods. Given the resulting economic environment, firms regulated by market forces conduct themselves much as before. Although the background institutions will take a different form, especially in the case of the distribution branch, there is no reason in principle why just distributive shares cannot be achieved. The theory of justice does not by itself favor either form of regime. As we have seen, the decision as to which system is best for a given people depends upon their circumstances, institutions, and historical traditions.

Some socialists have objected to all market institutions as inherently degrading, and they have hoped to set up an economy in which men are moved largely by social and altruistic concerns. In regard to the first, the market is not indeed an ideal arrangement, but certainly given the requisite background institutions, the worst aspects of so-called wage slavery are removed. The question then becomes one of the comparison of possible alternatives. It seems improbable that the control of economic activity by the bureaucracy that would be bound to develop in a socially regulated system (whether centrally directed or guided by the agreements reached by industrial associations) would be more just on balance than control exercised by means of prices (assuming as always the necessary framework). To be sure a competitive scheme is impersonal and automatic in the details of its operation; its particular results do not express the conscious decision of individuals. But in many respects this is a virtue of the arrangement; and the use of the market system does not imply a lack of reasonable human autonomy. A democratic society may choose to rely on prices in view of the advantages of doing so, and then to maintain the background institutions which justice requires. This political decision, as well as the regulation of these surrounding arrangements, can be perfectly reasoned and free.

Moreover the theory of justice assumes a definite limit on the strength of social and altruistic motivation. It supposes that individuals and groups put forward competing claims, and while they are willing to act justly, they are not prepared to abandon their interests. There is no need to elaborate further that this presumption does not imply that men are selfish

in the ordinary sense. Rather a society in which all can achieve their complete good, or in which there are no conflicting demands and the wants of all fit together without coercion into a harmonious plan of activity, is a society in a certain sense beyond justice. It has eliminated the occasions when the appeal to the principles of right and justice is necessary.[18] I am not concerned with this ideal case, however desirable it may be. We should note though that even here the theory of justice has an important theoretical role: it defines the conditions under which the spontaneous coherence of the aims and wants of individuals is neither coerced nor contrived but expresses a proper harmony consistent with the ideal good. I cannot pursue these questions further. The main point is that the principles of justice are compatible with quite different types of regime.

A final matter needs to be considered. Let us suppose that the above account of the background institutions is sufficient for our purposes, and that the two principles of justice lead to a definite system of government activities and legal definitions of property together with a schedule of taxes. In this case the total of public expenditures and the necessary sources of revenue is well defined, and the distribution of income and wealth that results is just whatever it is. (See further below §§44, 47.) It does not follow, however, that citizens should not decide to make further public expenditures. If a sufficiently large number of them find the marginal benefits of public goods greater than that of goods available through the market, it is appropriate that ways should be found for government to provide them. Since the distribution of income and wealth is assumed to be just, the guiding principle changes. Let us suppose, then, that there is a fifth branch of government, the exchange branch, which consists of a special representative body taking note of the various social interests and their preferences for public goods. It is authorized by the constitution to consider only such bills as provide for government activities independent from what justice requires, and these are to be enacted only when they satisfy Wicksell's unanimity criterion.[19] This means that no public expenditures are voted upon unless at the same time the means of covering their

18. Some have interpreted Marx's conception of a full communist society as a society beyond justice in this sense. See R. C. Tucker, *The Marxian Revolutionary Idea* (New York, W.W. Norton, 1969), chs. I and II.

19. This criterion was stated by Knut Wicksell in his *Finanztheoretische Untersuchungen* (Jena, 1896). The major part is translated as "A New Principle of Just Taxation" and included in *Classics in the Theory of Public Finance,* ed. R. A. Musgrave and A. T. Peacock (London, Macmillan, 1958), pp. 72–118, esp. pp. 91–93, where the principle is stated. For some difficulties with it, see Hirafumi Shibata, "A Bargaining Model of the Pure Theory of Public Expenditure," *Journal of Political Economy,* vol. 79 (1971), esp. pp. 27f.

costs are agreed upon, if not unanimously, then approximately so. A motion proposing a new public activity is required to contain one or more alternative arrangements for sharing the costs. Wicksell's idea is that if the public good is an efficient use of social resources, there must be some scheme for distributing the extra taxes among different kinds of taxpayers that will gain unanimous approval. If no such proposal exists, the suggested expenditure is wasteful and should not be undertaken. Thus the exchange branch works by the principle of efficiency and institutes, in effect, a special trading body that arranges for public goods and services where the market mechanism breaks down. It must be added, however, that very real difficulties stand in the way of carrying this idea through. Even leaving aside voting strategies and the concealment of preferences, discrepancies in bargaining power, income effects, and the like may prevent an efficient outcome from being reached. Perhaps only a rough and approximate solution is possible. I shall, however, leave aside these problems.

Several comments are called for to prevent misunderstandings. First of all, as Wicksell emphasized, the unanimity criterion assumes the justice of the existing distribution of income and wealth, and of the current definition of the rights of property. Without this important proviso, it would have all the faults of the efficiency principle, since it simply expresses this principle for the case of public expenditures. But when this condition is satisfied, then the unanimity principle is sound. There is no more justification for using the state apparatus to compel some citizens to pay for unwanted benefits that others desire than there is to force them to reimburse others for their private expenses. Thus the benefit criterion now applies whereas it did not before; and those who want further public expenditures of various kinds are to use the exchange branch to see whether the requisite taxes can be agreed to. The size of the exchange budget, as distinct from the national budget, is then determined by the expenditures that are eventually accepted. In theory members of the community can get together to purchase public goods up to the point where their marginal value equals that of private goods.

It should be noted that the exchange branch includes a separate representative body. The reason for this is to emphasize that the basis of this scheme is the benefit principle and not the principles of justice. Since the conception of background institutions is to help us organize our considered judgments of justice, the veil of ignorance applies to the legislative stage. The exchange branch is only a trading arrangement. There are no restrictions upon information (except those required to make the scheme

more efficient), since it depends upon citizens' knowing their relative valuations of public and private goods. We should also observe that in the exchange branch representatives (and citizens through their representatives) are quite properly guided by their interests. Whereas in describing the other branches, we assume the principles of justice to be applied to institutions solely on the basis of general information. We try to work out what rational legislators suitably constrained by the veil of ignorance, and in this sense impartial, would enact to realize the conception of justice. Ideal legislators do not vote their interests. Strictly speaking, then, the idea of the exchange branch is not part of the four-stage sequence. Nevertheless, there is likely to be confusion between government activities and public expenditures required to uphold just background institutions and those that follow from the benefit principle. With the distinction of branches in mind, the conception of justice as fairness becomes, I believe, more plausible. To be sure, it is often hard to distinguish between the two kinds of government activities, and some public goods may appear to fall into both categories. I leave these problems aside here, hoping that the theoretical distinction is clear enough for present purposes.

44. THE PROBLEM OF JUSTICE BETWEEN GENERATIONS

We must now consider the question of justice between generations. There is no need to stress the difficulties that this problem raises. It subjects any ethical theory to severe if not impossible tests. Nevertheless, the account of justice as fairness would be incomplete without some discussion of this important matter. The problem arises in the present context because the question is still open whether the social system as a whole, the competitive economy surrounded by the appropriate family of background institutions, can be made to satisfy the two principles of justice. The answer is bound to depend, to some degree anyway, on the level at which the social minimum is to be set. But this in turn connects up with how far the present generation is bound to respect the claims of its successors.

So far I have said nothing about how generous the social minimum should be. Common sense might be content to say that the right level depends upon the average wealth of the country and that, other things equal, the minimum should be higher when the average increases. Or one might say that the proper level is determined by customary expectations. But these suggestions are unsatisfactory. The first is not precise enough since it does not say how the minimum depends on average wealth and it

overlooks other relevant aspects such as distribution; while the second provides no criterion for telling when customary expectations are themselves reasonable. Once the difference principle is accepted, however, it follows that the minimum is to be set at that point which, taking wages into account, maximizes the expectations of the least advantaged group. By adjusting the amount of transfers (for example, the size of supplementary income payments), it is possible to increase or decrease the prospects of the more disadvantaged, their index of primary goods (as measured by wages plus transfers), so as to achieve the desired result.

Now offhand it might seem that the difference principle requires a very high minimum. One naturally imagines that the greater wealth of those better off is to be scaled down until eventually everyone has nearly the same income. But this is a misconception, although it might hold in special circumstances. The appropriate expectation in applying the difference principle is that of the long-term prospects of the least favored extending over future generations. Each generation must not only preserve the gains of culture and civilization, and maintain intact those just institutions that have been established, but it must also put aside in each period of time a suitable amount of real capital accumulation. This saving may take various forms from net investment in machinery and other means of production to investment in learning and education. Assuming for the moment that a just savings principle is available which tells us how great investment should be, the level of the social minimum is determined. Suppose for simplicity that the minimum is adjusted by transfers paid for by proportional expenditure (or income) taxes. In this case raising the minimum entails increasing the proportion by which consumption (or income) is taxed. Presumably as this fraction becomes larger there comes a point beyond which one of two things happens. Either the appropriate savings cannot be made or the greater taxes interfere so much with economic efficiency that the prospects of the least advantaged in the present generation are no longer improved but begin to decline. In either event the correct minimum has been reached. The difference principle is satisfied and no further increase is called for.

These comments about how to specify the social minimum have led us to the problem of justice between generations. Finding a just savings principle is one aspect of this question.[20] Now I believe that it is not

20. This problem is often discussed by economists in the context of the theory of economic growth. For an exposition see A. K. Sen, "On Optimizing the Rate of Saving," *Economic Journal,* vol. 71 (1961); James Tobin, *National Economic Policy* (New Haven, Yale University Press, 1966), ch. IX; and R. M. Solow, *Growth Theory* (New York, Oxford University Press, 1970), ch. V. In an extensive

possible, at present anyway, to define precise limits on what the rate of savings should be. How the burden of capital accumulation and of raising the standard of civilization and culture is to be shared between generations seems to admit of no definite answer. It does not follow, however, that certain significant ethical constraints cannot be formulated. As I have said, a moral theory characterizes a point of view from which policies are to be assessed; and it may often be clear that a suggested answer is mistaken even if an alternative doctrine is not ready to hand. Thus it seems evident, for example, that the classical principle of utility leads in the wrong direction for questions of justice between generations. For if one takes the size of the population as variable, and postulates a high marginal productivity of capital and a very distant time horizon, maximizing total utility may lead to an excessive rate of accumulation (at least in the near future). Since from a moral point of view there are no grounds for discounting future well-being on the basis of pure time preference, the conclusion is all the more likely that the greater advantages of future generations will be sufficiently large to outweigh most any present sacrifices. This may prove true if only because with more capital and better technology it will be possible to support a sufficiently large population. Thus the utilitarian doctrine may direct us to demand heavy sacrifices of the poorer generations for the sake of greater advantages for later ones that are far better off. But this calculus of advantages, which balances the losses of some against benefits to others, appears even less justified in the case of generations than among contemporaries. Even if we cannot define a precise just savings principle, we should be able to avoid this sort of extreme.

Now the contract doctrine looks at the problem from the standpoint of the original position and requires the parties to adopt an appropriate savings principle. It seems clear that as they stand the two principles of justice must be adjusted to this question. For when the difference principle is applied to the question of saving over generations, it entails either

---

literature, see F. P. Ramsey, "A Mathematical Theory of Saving," *Economic Journal,* vol. 38 (1928), reprinted in Arrow and Scitovsky, *Readings in Welfare Economics;* T. C. Koopmans, "On the Concept of Optimal Economic Growth" (1965) in *Scientific Papers of T. C. Koopmans* (Berlin, Springer Verlag, 1970). Sukamoy Chakravarty, *Capital and Development Planning* (Cambridge, M.I.T. Press, 1969), is a theoretical survey which touches upon the normative questions. If for theoretical purposes one thinks of the ideal society as one whose economy is in a steady state of growth (possibly zero), and which is at the same time just, then the savings problem is to choose a principle for sharing the burdens of getting to that growth path (or to such a path if there is more than one), and of maintaining the justice of the necessary arrangements once this is achieved. In the text, however, I do not pursue this suggestion; my discussion is at a more primitive level.

no saving at all or not enough saving to improve social circumstances sufficiently so that all the equal liberties can be effectively exercised. In following a just savings principle, each generation makes a contribution to those coming later and receives from its predecessors. There is no way for later generations to help the situation of the least fortunate earlier generation. Thus the difference principle does not hold for the question of justice between generations and the problem of saving must be treated in some other manner.

Some have thought the different fortunes of generations to be unjust. Herzen remarks that human development is a kind of chronological unfairness, since those who live later profit from the labor of their predecessors without paying the same price. And Kant thought it disconcerting that earlier generations should carry their burdens only for the sake of the later ones and that only the last should have the good fortune to dwell in the completed building.[21] These feelings while entirely natural are misplaced. For although the relation between generations is a special one, it gives rise to no insuperable difficulty.

It is a natural fact that generations are spread out in time and actual economic benefits flow only in one direction. This situation is unalterable, and so the question of justice does not arise. What is just or unjust is how institutions deal with natural limitations and the way they are set up to take advantage of historical possibilities. Obviously if all generations are to gain (except perhaps the earlier ones), the parties must agree to a savings principle that insures that each generation receives its due from its predecessors and does its fair share for those to come. The only economic exchanges between generations are, so to speak, virtual ones, that is, compensating adjustments that can be made in the original position when a just savings principle is adopted.

Now when the parties consider this problem they do not know to which generation they belong or, what comes to the same thing, the stage of civilization of their society. They have no way of telling whether it is poor or relatively wealthy, largely agricultural or already industrialized, and so on. The veil of ignorance is complete in these respects. But since we take the present time of entry interpretation of the original position (§24), the parties know that they are contemporaries; and so unless we modify our initial assumptions, there is no reason for them to agree to any saving

21. The remark of Alexander Herzen is from Isaiah Berlin's introduction to Franco Venturi, *Roots of Revolution* (New York, Alfred Knopf, 1960), p. xx. For Kant, see "Idea for a Universal History with a Cosmopolitan Purpose," in *Political Writings,* ed. Hans Reiss and trans. H. B. Nisbet (Cambridge, The University Press, 1970), p. 44.

whatever. Earlier generations will have either saved or not; there is nothing the parties can do to affect that. So to achieve a reasonable result, we assume first, that the parties represent family lines, say, who care at least about their more immediate descendants; and second, that the principle adopted must be such that they wish all earlier generations to have followed it (§22). These constraints, together with the veil of ignorance, are to insure that any one generation looks out for all.

In arriving at a just saving principle (or better, limits on such principles), the parties are to ask themselves how much they would be willing to save at each stage of advance on the assumption that all other generations have saved, or will save, in accordance with the same criterion. They are to consider their willingness to save at any given phase of civilization with the understanding that the rates they propose are to regulate the whole span of accumulation. It is essential to note that a savings principle is a rule that assigns an appropriate rate (or range of rates) to each level of advance, that is, a rule that determines a schedule of rates. Presumably different rates are assigned to different stages. When people are poor and saving is difficult, a lower rate of saving should be required; whereas in a wealthier society greater savings may reasonably be expected since the real burden of saving is less. Eventually, once just institutions are firmly established and all the basic liberties effectively realized, the net accumulation asked for falls to zero. At this point a society meets its duty of justice by maintaining just institutions and preserving their material base. The just savings principle applies to what a society is to save as a matter of justice. If its members wish to save for other purposes, that is another matter.

It is impossible to be very specific about the schedule of rates (or the range of rates) that would be acknowledged; the most that we can hope from these intuitive considerations is that certain extremes will be excluded. Thus we may assume that the parties avoid imposing very high rates at the earlier stages of accumulation, for even though they would benefit from this if they come later, they must be able to accept these rates in good faith should their society turn out to be poor. The strains of commitment apply here just as before (§29). On the other hand, they will want all generations to provide some saving (excluding special circumstances), since it is to our advantage if our predecessors have done their share. These observations establish wide limits for the savings rule. To narrow the range somewhat further, we suppose the parties to ask what is reasonable for members of adjacent generations to expect of one another at each level of advance. They try to piece together a just savings sched-

ule by balancing how much they would be willing to save for their more immediate descendants against what they would feel entitled to claim of their more immediate predecessors. Thus imagining themselves to be fathers, say, they are to ascertain how much they should set aside for their sons and grandsons by noting what they would believe themselves entitled to claim of their fathers and grandfathers. When they arrive at the estimate that seems fair from both sides, with due allowance made for the improvement in circumstances, then the fair rate (or range of rates) for that stage is specified. Once this is done for all stages, the just savings principle is defined. Of course, the parties must throughout keep in mind the objective of the accumulation process, namely, a state of society with a material base sufficient to establish effective just institutions within which the basic liberties can all be realized. Assuming that the savings principle answers to these conditions, no generation can find fault with any other when it is followed, no matter how far removed they are in time.

The question of time preference and matters of priority I shall leave aside until the next sections. For the present I wish to point out several features of the contract approach. First of all, while it is evident that a just savings principle cannot literally be adopted democratically, the conception of the original position achieves the same result. Since no one knows to which generation he belongs, the question is viewed from the standpoint of each and a fair accommodation is expressed by the principle adopted. All generations are virtually represented in the original position, since the same principle would always be chosen. An ideally democratic decision will result, one that is fairly adjusted to the claims of each generation and therefore satisfying the precept that what touches all concerns all. Moreover, it is immediately obvious that every generation, except possibly the first, gains when a reasonable rate of saving is maintained. The process of accumulation, once it is begun and carried through, is to the good of all subsequent generations. Each passes on to the next a fair equivalent in real capital as defined by a just savings principle. (It should be kept in mind here that capital is not only factories and machines, and so on, but also the knowledge and culture, as well as the techniques and skills, that make possible just institutions and the fair value of liberty.) This equivalent is in return for what is received from previous generations that enables the later ones to enjoy a better life in a more just society.

It is also characteristic of the contract doctrine to define a just society as the aim of the course of accumulation. This feature derives from the

fact that an ideal conception of a just basic structure is embedded in the principles chosen in the original position. In this respect, justice as fairness contrasts with utilitarian views (§41). The just savings principle can be regarded as an understanding between generations to carry their fair share of the burden of realizing and preserving a just society. The end of the savings process is set up in advance, although only the general outlines can be discerned. Particular circumstances as they arise will in time determine the more detailed aspects. But in any event we are not bound to go on maximizing indefinitely. Indeed, it is for this reason that the savings principle is agreed to after the principles of justice for institutions, even though this principle constrains the difference principle. These principles tell us what to strive for. The savings principle represents an interpretation, arrived at in the original position, of the previously accepted natural duty to uphold and to further just institutions. In this case the ethical problem is that of agreeing on a path over time which treats all generations justly during the whole course of a society's history. What seems fair to persons in the original position defines justice in this instance as in others.

The significance of the last stage of society should not, however, be misinterpreted. While all generations are to do their part in reaching the just state of things beyond which no further net saving is required, this state is not to be thought of as that alone which gives meaning and purpose to the whole process. To the contrary, all generations have their appropriate aims. They are not subordinate to one another any more than individuals are and no generation has stronger claims than any other. The life of a people is conceived as a scheme of cooperation spread out in historical time. It is to be governed by the same conception of justice that regulates the cooperation of contemporaries.

Finally, the last stage at which saving is called for is not one of great abundance. This consideration deserves perhaps some emphasis. Further wealth might not be superfluous for some purposes; and indeed average income may not, in absolute terms, be very high. Justice does not require that early generations save so that later ones are simply more wealthy. Saving is demanded as a condition of bringing about the full realization of just institutions and the equal liberties. If additional accumulation is to be undertaken, it is for other reasons. It is a mistake to believe that a just and good society must wait upon a high material standard of life. What men want is meaningful work in free association with others, these associations regulating their relations to one another within a framework of just basic institutions. To achieve this state of things great wealth is not

necessary. In fact, beyond some point it is more likely to be a positive hindrance, a meaningless distraction at best if not a temptation to indulgence and emptiness. (Of course, the definition of meaningful work is a problem in itself. Though it is not a problem of justice, a few remarks in §79 are addressed to it.)

We now have to combine the just savings principle with the two principles of justice. This is done by supposing that this principle is defined from the standpoint of the least advantaged in each generation. It is the representative men from this group as it extends over time who by virtual adjustments are to specify the rate of accumulation. They undertake in effect to constrain the application of the difference principle. In any generation their expectations are to be maximized subject to the condition of putting aside the savings that would be acknowledged. Thus the complete statement of the difference principle includes the savings principle as a constraint. Whereas the first principle of justice and the principle of fair opportunity are prior to the difference principle within generations, the savings principle limits its scope between them.

Of course, the saving of the less favored need not be done by their taking an active part in the investment process. Rather it normally consists of their approving of the economic and other arrangements necessary for the appropriate accumulation. Saving is achieved by accepting as a political judgment those policies designed to improve the standard of life of later generations of the least advantaged, thereby abstaining from the immediate gains which are available. By supporting these arrangements the required saving can be made, and no representative man in any generation of the most disadvantaged can complain of another for not doing his part.

So much, then, for a brief sketch of some of the main features of the just savings principle. We can now see that persons in different generations have duties and obligations to one another just as contemporaries do. The present generation cannot do as it pleases but is bound by the principles that would be chosen in the original position to define justice between persons at different moments of time. In addition, men have a natural duty to uphold and to further just institutions and for this the improvement of civilization up to a certain level is required. The derivation of these duties and obligations may seem at first a somewhat farfetched application of the contract doctrine. Nevertheless these requirements would be acknowledged in the original position, and so the conception of justice as fairness covers these matters without any change in its basic idea.

45. TIME PREFERENCE

I have assumed that in choosing a principle of savings the persons in the original position have no pure time preference. We need to consider the reasons for this presumption. In the case of an individual the avoidance of pure time preference is a feature of being rational. As Sidgwick maintains, rationality implies an impartial concern for all parts of our life. The mere difference of location in time, of something's being earlier or later, is not in itself a rational ground for having more or less regard for it. Of course, a present or near future advantage may be counted more heavily on account of its greater certainty or probability, and we should take into consideration how our situation and capacity for particular enjoyments will change. But none of these things justifies our preferring a lesser present to a greater future good simply because of its nearer temporal position[22] (§64).

Now Sidgwick thought that the notions of universal good and individual good are in essential respects similar. He held that just as the good of one person is constructed by comparison and integration of the different goods of each moment as they follow one another in time, so the universal good is constructed by the comparison and integration of the good of many different individuals. The relations of the parts to the whole and to each other are analogous in each case, being founded on the aggregative principle of utility.[23] The just savings principle for society must not, then, be affected by pure time preference, since as before the different temporal position of persons and generations does not in itself justify treating them differently.

Since in justice as fairness the principles of justice are not extensions of the principles of rational choice for one person, the argument against time preference must be of another kind. The question is settled by reference to the original position; but once it is seen from this perspective, we reach the same conclusion. There is no reason for the parties to give any weight to mere position in time. They have to choose a rate of saving for each level of civilization. If they make a distinction between earlier and more remote periods because, say, future states of affairs seem less important now, the present state of affairs will seem less important in the future. Although any decision has to be made now, there is no ground for their using today's discount of the future rather than the future's

22. See *The Methods of Ethics,* 7th ed. (London, Macmillan, 1907), p. 381. Time preference is also rejected by Ramsey, "A Mathematical Theory of Saving."

23. *Methods of Ethics,* p. 382. See also §30, note 37.

discount of today. The situation is symmetrical and one choice is as arbitrary as the other.[24] Since the persons in the original position take up the standpoint of each period, being subject to the veil of ignorance, this symmetry is clear to them and they will not consent to a principle that weighs nearer periods more or less heavily. Only in this way can they arrive at a consistent agreement from all points of view, for to acknowledge a principle of time preference is to authorize persons differently situated temporally to assess one another's claims by different weights based solely on this contingency.

As with rational prudence, the rejection of pure time preference is not incompatible with taking uncertainties and changing circumstances into account; nor does it rule out using an interest rate (in either a socialist or a private-property economy) to ration limited funds for investment. The restriction is rather that in first principles of justice we are not allowed to treat generations differently solely on the grounds that they are earlier or later in time. The original position is so defined that it leads to the correct principle in this respect. In the case of the individual, pure time preference is irrational: it means that he is not viewing all moments as equally parts of one life. In the case of society, pure time preference is unjust: it means (in the more common instance when the future is discounted) that the living take advantage of their position in time to favor their own interests.

The contract view agrees, then, with Sidgwick in rejecting time preference as a grounds of social choice. The living may, if they allow themselves to be moved by such considerations, wrong their predecessors and descendants. Now this contention may seem contrary to democratic principles, for it is sometimes said that these require that the wishes of the present generation should determine social policy. Of course, it is assumed that these preferences need to be clarified and ascertained under the appropriate conditions. Collective saving for the future has many aspects of a public good, and the isolation and assurance problems arise in this case.[25] But supposing that these difficulties are overcome and that the informed collective judgment of the present generation is known under the requisite conditions, it may be thought that a democratic view of the state does not countenance the government's intervening for the sake of future generations even when the public judgment is manifestly mistaken.

24. See Sen, "On Optimizing the Rate of Savings," p. 482.

25. See Sen, ibid., p. 479; and S. A. Marglin, "The Social Rate of Discount and the Optimal Rate of Investment," *Quarterly Journal of Economics,* vol. 77 (1963), pp. 100–109.

Whether this contention is correct depends upon how it is interpreted. There can be no objection to it as a description of a democratic constitution. Once the public will is clearly expressed in legislation and social policies, the government cannot override it without ceasing to be democratic. It is not authorized to nullify the views of the electorate as to how much saving is to be undertaken. If a democratic regime is justified, then the government's having this power would normally lead to a greater injustice on balance. We are to decide between constitutional arrangements according to how likely it is that they will yield just and effective legislation. A democrat is one who believes that a democratic constitution best meets this criterion. But his conception of justice includes a provision for the just claims of future generations. Even if as a practical matter in the choice of regimes the electorate should have the final say, this is only because it is more likely to be correct than a government empowered to override its wishes. Since, however, a just constitution even under favorable conditions is a case of imperfect procedural justice, the people may still decide wrongly. By causing irreversible damages say, they may perpetuate grave offenses against other generations which under another form of government might have been prevented. Moreover, the injustice may be perfectly evident and demonstrable as such by the same conception of justice that underlies the democratic regime itself. Several of the principles of this conception may actually be more or less explicit in the constitution and frequently cited by the judiciary and informed opinion in interpreting it.

In these cases, then, there is no reason why a democrat may not oppose the public will by suitable forms of noncompliance, or even as a government official try to circumvent it. Although one believes in the soundness of a democratic constitution and accepts the duty to support it, the duty to comply with particular laws may be overridden in situations where the collective judgment is sufficiently unjust. There is nothing sacrosanct about the public decision concerning the level of savings; and its bias with respect to time preference deserves no special respect. In fact the absence of the injured parties, the future generations, makes it all the more open to question. One does not cease to be a democrat unless one thinks that some other form of government would be better and one's efforts are directed to this end. As long as one does not believe this, but thinks instead that appropriate forms of noncompliance, for example, acts of civil disobedience or conscientious refusal, are both necessary and reasonable ways to correct democratically enacted policies, then one's conduct is consistent with accepting a democratic constitution. In the next

chapter I shall discuss this matter in more detail. For the moment the essential point is that the collective will concerning the provision for the future is subject, as all other social decisions are, to the principles of justice. The peculiar features of this case do not make it an exception.

We should observe that to reject pure time preference as a first principle is compatible with recognizing that a certain discounting of the future may improve otherwise defective criteria. For example, I have already remarked that the utilitarian principle may lead to an extremely high rate of saving which imposes excessive hardships on earlier generations. This consequence can be to some degree corrected by discounting the welfare of those living in the future. Since the well-being of later generations is made to count for less, not so much need be saved as before. It is also possible to vary the accumulation required by adjusting the parameters in the postulated utility function. I cannot discuss these questions here.[26] Unhappily I can only express the opinion that these devices simply mitigate the consequences of mistaken principles. The situation is in some respects similar to that found with the intuitionistic conception which combines the standard of utility with a principle of equality (see §7). There the criterion of equality suitably weighted serves to correct the utility criterion when neither principle taken alone would prove acceptable. Thus in an analogous way, having started with the idea that the appropriate rate of saving is the one which maximizes social utility over time (maximizes some integral), we may obtain a more plausible result if the welfare of future generations is weighted less heavily; and the most suitable discount may depend upon how swiftly population is growing, upon the productivity of capital, and so on. What we are doing is adjusting certain parameters so as to reach a conclusion more in line with our intuitive judgments. We may find that to achieve justice between generations, these modifications in the principle of utility are required. Certainly introducing time preference may be an improvement in such cases; but I believe that its being invoked in this way is an indication that we have started from an incorrect conception. There is a difference between the situation here and the previously mentioned intuitionistic view. Unlike the principle of equality, time preference has no intrinsic ethical appeal. It is introduced in a purely ad hoc way to moderate the consequences of the utility criterion.

26. See Chakravarty, *Capital and Development Planning,* pp. 39f, 47, 63–65, 249f. Solow, *Growth Theory,* pp. 79–87, gives an account of the mathematical problem.

46. FURTHER CASES OF PRIORITY

The problem of just savings may be used to illustrate further cases of the priority of justice. One feature of the contract doctrine is that it places an upper bound on how much a generation can be asked to save for the welfare of later generations. The just savings principle acts as a constraint on the rate of accumulation. Each age is to do its fair share in achieving the conditions necessary for just institutions and the fair value of liberty; but beyond this more cannot be required. Now it may be objected that particularly when the sum of advantages is very great and represents long-term developments, higher rates of saving may be demanded. Some may go further and maintain that inequalities in wealth and authority violating the second principle of justice may be justified if the subsequent economic and social benefits are large enough. To support their view they may point to instances in which we seem to accept such inequalities and rates of accumulation for the sake of the welfare of later generations. Keynes remarks, for example, that the immense accumulations of capital built up before the First World War could never have come about in a society in which wealth was equally divided.[27] Society in the nineteenth century, he says, was arranged so as to place the increased income in the hands of those least likely to consume it. The new rich were not brought up to large expenditures and preferred to the enjoyments of immediate consumption the power which investment gave. It was precisely the inequality of the distribution of wealth which made possible the rapid build-up of capital and the more or less steady improvement in the general standard of living of everyone. It is this fact, in Keynes's opinion, that provided the main justification of the capitalist system. If the rich had spent their new wealth on themselves, such a regime would have been rejected as intolerable. Certainly there are more efficient and just ways of raising the level of well-being and culture than that Keynes describes. It is only in special circumstances, including the frugality of the capitalist class as opposed to the self-indulgence of the aristocracy, that a society should obtain investment funds by endowing the rich with more than they feel they can decently spend on themselves. But the essential point here is that Keynes's justification, whether or not its premises are sound, can be made to turn solely on improving the situation of the working class. Although their circumstances appear harsh, Keynes presumably main-

27. See J. M. Keynes, *The Economic Consequences of the Peace* (London, Macmillan, 1919), pp. 18–22.

tains that while there were many ostensible injustices in the system, there was no real possibility that these could have been removed and the conditions of the less advantaged made better. Under other arrangements the position of the laboring man would have been even worse. We need not consider whether these contentions are true. It suffices to note that, contrary to what one might have thought, Keynes does not say that the hardships of the poor are justified by the greater welfare of later generations. And this accords with the priority of justice over efficiency and a greater sum of advantages. Whenever the constraints of justice in the matter of savings are infringed, it must be shown that circumstances are such that not to trespass upon them would lead to an even greater injury to those on whom the injustice falls. This case is analogous to those already discussed under the heading of the priority of liberty (see §39).

It is clear that the inequalities that Keynes had in mind also violate the principle of fair equality of opportunity. Thus we are led to consider what must be argued to excuse the infringement of this criterion and how to formulate the appropriate priority rule.[28] Many writers hold that fair equality of opportunity would have grave consequences. They believe that some sort of hierarchical social structure and a governing class with pervasive hereditary features are essential for the public good. Political power should be exercised by men experienced in, and educated from childhood to assume, the constitutional traditions of their society, men whose ambitions are moderated by the privileges and amenities of their assured position. Otherwise the stakes become too high and those lacking in culture and conviction contend with one another to control the power of the state for their narrow ends. Thus Burke believed that the great families of the ruling stratum contribute by the wisdom of their political rule to the general welfare from generation to generation.[29] And Hegel thought that restrictions on equality of opportunity such as primogeniture are essential to insure a landed class especially suited to political rule in virtue of its independence from the state, the quest for profit, and the manifold contingencies of civil society.[30] Privileged family and property arrangements prepare those favored by them to take a clearer view of the universal interest for the benefit of the whole society. Of course, one need not favor anything like a rigidly stratified system; one may maintain to

28. In this and the next several paragraphs, I am indebted to Michael Lessnoff. See his essay in *Political Studies,* vol. 19 (1971), pp. 75f. The statement and discussion of the priority rules here and in §39 have benefited from his criticisms.

29. See *Reflections on the Revolution in France* (London, J.M. Dent and Sons, 1910), p. 49; and John Plamenatz, *Man and Society* (London, Longmans, Green, 1963), vol. 1, pp. 346–351.

30. *Philosophy of Right,* §306, trans. T. M. Knox (Oxford, The Clarendon Press, 1942), p. 199.

the contrary that it is essential for the vigor of the governing class that persons of unusual talents should be able to make their way into it and be fully accepted. But this proviso is compatible with denying the principle of fair opportunity.

Now to be consistent with the priority of fair opportunity over the difference principle, it is not enough to argue, as Burke and Hegel appear to, that the whole of society including the least favored benefit from certain restrictions on equality of opportunity. We must also claim that the attempt to eliminate these inequalities would so interfere with the social system and the operations of the economy that in the long run anyway the opportunities of the disadvantaged would be even more limited. The priority of fair opportunity, as in the parallel case of the priority of liberty, means that we must appeal to the chances given to those with the lesser opportunity. We must hold that a wider range of more desirable alternatives is open to them than otherwise would be the case.

I shall not pursue these complications further. We should however note that although the internal life and culture of the family influence, perhaps as much as anything else, a child's motivation and his capacity to gain from education, and so in turn his life prospects, these effects are not necessarily inconsistent with fair equality of opportunity. Even in a well-ordered society that satisfies the two principles of justice, the family may be a barrier to equal chances between individuals. For as I have defined it, the second principle only requires equal life prospects in all sectors of society for those similarly endowed and motivated. If there are variations among families in the same sector in how they shape the child's aspirations, then while fair equality of opportunity may obtain between sectors, equal chances between individuals will not. This possibility raises the question as to how far the notion of equality of opportunity can be carried; but I defer comment on this until later (§77). I shall only remark here that following the difference principle and the priority rules it suggests reduces the urgency to achieve perfect equality of opportunity.

I shall not examine whether there are sound arguments overriding the principle of fair equality of opportunity in favor of a hierarchical class structure. These matters are not part of the theory of justice. The relevant point is that while such contentions may sometimes appear self-serving and hypocritical, they have the right form when they claim (whether correctly or not) that the opportunities of the least favored sectors of the community would be still more limited if these inequalities were removed. One is to hold that they are not unjust, since the conditions for achieving the full realization of the principles of justice do not exist.

Having noted these cases of priority, I now wish to give the final statement of the two principles of justice for institutions. For the sake of completeness, I shall give a full statement including earlier formulations.

FIRST PRINCIPLE

Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.

SECOND PRINCIPLE

Social and economic inequalities are to be arranged so that they are both:

(a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and

(b) attached to offices and positions open to all under conditions of fair equality of opportunity.

FIRST PRIORITY RULE (THE PRIORITY OF LIBERTY)

The principles of justice are to be ranked in lexical order and therefore the basic liberties can be restricted only for the sake of liberty. There are two cases:

(a) a less extensive liberty must strengthen the total system of liberties shared by all;

(b) a less than equal liberty must be acceptable to those with the lesser liberty.

SECOND PRIORITY RULE (THE PRIORITY OF JUSTICE OVER EFFICIENCY AND WELFARE)

The second principle of justice is lexically prior to the principle of efficiency and to that of maximizing the sum of advantages; and fair opportunity is prior to the difference principle. There are two cases:

(a) an inequality of opportunity must enhance the opportunities of those with the lesser opportunity;

(b) an excessive rate of saving must on balance mitigate the burden of those bearing this hardship.

By way of comment, these principles and priority rules are no doubt incomplete. Other modifications will surely have to be made, but I shall not further complicate the statement of the principles. It suffices to observe that when we come to nonideal theory, the lexical ordering of the two principles, and the valuations that this ordering implies, suggest priority rules which seem to be reasonable enough in many cases. By various examples I have tried to illustrate how these rules can be used and to indicate their plausibility. Thus the ranking of the principles of justice in ideal theory reflects back and guides the application of these principles to nonideal situations. It identifies which limitations need to be dealt with first. In the more extreme and tangled instances of nonideal theory this priority of rules will no doubt fail; and indeed, we may be able to find no satisfactory answer at all. But we must try to postpone the day of reckoning as long as possible, and try to arrange society so that it never comes.

47. THE PRECEPTS OF JUSTICE

The sketch of the system of institutions that satisfies the two principles of justice is now complete. Once the just rate of savings is ascertained or the appropriate range of rates specified, we have a criterion for adjusting the level of the social minimum. The sum of transfers and benefits from essential public goods should be arranged so as to enhance the expectations of the least favored consistent with the required savings and the maintenance of equal liberties. When the basic structure takes this form the distribution that results will be just (or at least not unjust) whatever it is. Each receives that total income (earnings plus transfers) to which he is entitled under the public system of rules upon which his legitimate expectations are founded.

Now, as we saw earlier (§14), a central feature of this conception of distributive justice is that it contains a large element of pure procedural justice. No attempt is made to define the just distribution of goods and services on the basis of information about the preferences and claims of particular individuals. This sort of knowledge is regarded as irrelevant from a suitably general point of view; and in any case, it introduces complexities that cannot be handled by principles of tolerable simplicity