

EQUITY

IN THEORY AND PRACTICE

H. Peyton Young

A RUSSELL SAGE FOUNDATION BOOK

PRINCETON UNIVERSITY PRESS PRINCETON, NEW JERSEY

9. <i>The Method of Joseph Hill</i>	52
10. <i>Bias</i>	54
11. <i>Consistency and Priority</i>	57
12. <i>Staying within the Quota</i>	60
13. <i>The Population Paradox</i>	60
14. <i>Summary</i>	62
4. Equity, Equality, Proportionality	64
1. <i>Aristotle's Equity Principle</i>	64
2. <i>Claims Problems</i>	66
3. <i>The Contested Garment Rule</i>	67
4. <i>The Shapley Value</i>	69
5. <i>An Inconsistency in the Shapley Value</i>	71
6. <i>Maimonides' Rule</i>	73
7. <i>Gain vs. Loss</i>	73
8. <i>Varieties of Equality</i>	75
9. <i>Equity, Priority, and Consistency</i>	76
10. <i>Incentive Effects</i>	79
11. <i>Summary</i>	79
5. Cost Sharing	81
1. <i>Sharing Gains from Cooperation</i>	81
2. <i>A Cost-Sharing Problem between Two Towns</i>	81
3. <i>A Cost-Sharing Problem among Three Towns</i>	82
4. <i>The Cooperative Game Model</i>	85
5. <i>The Tennessee Valley Authority</i>	86
6. <i>The Decomposition Principle</i>	88
7. <i>The Shapley Value</i>	90
8. <i>Equitable Core Solutions: The Nucleolus</i>	93
9. <i>Summary</i>	96
6. Progressive Taxation	97
1. <i>Historical Background</i>	97
2. <i>The Progressivity Principle</i>	98
3. <i>The U.S. Federal Income Tax</i>	99
4. <i>Redressing Inequality</i>	100
5. <i>The Benefit Theory</i>	102
6. <i>Ability to Pay and Equal Sacrifice</i>	105
7. <i>The Effect of Progressive Taxation on Work Effort</i>	108
8. <i>Optimal Taxation</i>	109
9. <i>The Effect of Taxation on Risk-Taking</i>	111
10. <i>Summary</i>	114

CONTENTS	ix
7. Fair Bargains	116
1. <i>Bargaining Over Common Property</i>	116
2. <i>The Bargaining Set</i>	116
3. <i>The Coordination Problem</i>	118
4. <i>Classical Bargaining Solutions: Nash and Kalai-Smorodinsky</i>	119
5. <i>Framing Effects</i>	122
6. <i>Equity Criteria Based on Tangible Claims</i>	123
7. <i>Experimental Results on Bargaining</i>	124
8. <i>Empirical Evidence from Sharecropping Practices</i>	127
9. <i>Summary</i>	129
8. Fair Process	130
1. <i>Games of Fair Division</i>	130
2. <i>Auctioning Indivisibles</i>	131
3. <i>Superior and Inferior Modes of Division</i>	134
4. <i>Divide and Choose</i>	135
5. <i>The Divider's Advantage</i>	137
6. <i>Removing the Divider's Advantage by Lottery</i>	138
7. <i>Successively Splitting the Difference: The Raiffa Solution</i>	140
8. <i>Alternating Offers: The Nash Solution</i>	141
9. <i>Bidding to Be Divider: The Egalitarian Solution</i>	143
10. <i>Summary</i>	145
9. Equity, Envy, and Efficiency	146
1. <i>Fair and Efficient Exchange</i>	146
2. <i>Transparent Inequity</i>	146
3. <i>Egalitarianism</i>	147
4. <i>A Difficulty with Egalitarianism</i>	150
5. <i>Competitive Allocations</i>	151
6. <i>The Equity of Competitive Allocation</i>	152
7. <i>The Competitive Standard of Comparison</i>	154
8. <i>Enlarging the Pie</i>	155
9. <i>An Application: Assigning Students to Dormitories</i>	156
10. <i>Restricting the Domain of Exchange</i>	160
11. <i>Summary</i>	161
10. Conclusion	162
Appendix: The Mathematical Theory of Equity	169
A.1. <i>Two Fundamental Principles</i>	169
A.2. <i>Zero-One Allocations</i>	174
A.3. <i>Opinion Aggregation</i>	176
A.4. <i>Integer Allocation</i>	184

Preface

Equity: simply a matter of the length of the judge's ears.

(*Elbert Hubbard*)

WHEN I TEACH this material to students, I warn them that the subject does not exist. Among all nonexistent subjects, in fact, equity occupies a distinguished position because it fails to exist in several different ways. The arguments against existence take three different forms. The first is that equity is merely a word that hypocritical people use to cloak self-interest. It has no intrinsic meaning and therefore fails to exist. The second argument is that, even if equity does exist in some notional sense, it is so hopelessly subjective that it cannot be analyzed scientifically. Thus it fails to exist in an *objective* sense. The third argument is that, even granting that equity might not be entirely subjective, there is no sensible theory about it, and certainly none that is compatible with modern welfare economics. In short, it fails to exist in an *academic* sense.

Set against these arguments is the fact that people who are not acquainted with them insist on using the term "equity" as if it did mean something. In everyday conversation we discuss with seeming abandon the equities and inequities of the tax structure, the health care system, the military draft, the price of telephone service, and how offices are allocated at work. For a term that does not exist this is a pretty good beginning. One is tempted to say that rumors of equity's nonexistence may have been somewhat exaggerated.

I hasten to add that equity, or at least a close relative, is very much alive and occupies a prominent place in moral philosophy. For philosophers, however, the central problem is to define what we mean by a just social order. Equity in this sense is concerned with the proper distribution of resources, rights, duties, opportunities, and obligations in society at large. This grand theme has animated political philosophers since antiquity, from Plato's and Aristotle's conceptions of the ideal state, to the social contract theories of Hobbes, Locke, and Rousseau, to the more modern theories of Rawls, Nozick, and Walzer. I shall refer to these as theories of *social justice*.

This book is not about equity in this sense. Rather, it is about the meaning of equity in concrete situations that we meet every day. Equity is, after all, an everyday concern. Families try to divide up the household chores equitably among their members. Businesses are sensitive to issues of equity in the salaries and perquisites that they offer to their employees. Public agencies worry about equity when they decide who has access to public housing; how much to charge

<i>A.5. Claims and Liabilities</i>	190
<i>A.6. Cooperative Games</i>	199
<i>A.7. Bargaining</i>	205
<i>A.8. Multiple Goods</i>	212
<i>Bibliographical Notes</i>	219
<i>Bibliography</i>	223
<i>Index</i>	233

for basic services such as water, electricity, and public transport; who gets a kidney for transplantation; and who gets into a nursery school (or a nursing home). Equity is a central concern in the most basic political decisions: the distribution of the tax burden, the duty to serve in the army, the apportionment of representation. All of these distributive problems can be and are solved without invoking theories of social justice. I am not saying that these theories are unimportant; they are of the utmost importance. What I am saying is that it is possible to analyze the meaning of equity in the small without resolving what social justice means in the large.

The aim of this book, then, is to examine how societies solve "everyday" distributive problems. Which patient gets the one available kidney for transplantation? Who is released from the army first? How much does the municipality charge for connecting a new house to the water supply? How many seats in Congress does one state receive compared to another? On what theory is one person taxed at a 33 percent rate, another at 15 percent, and a third not at all? What is a fair method for dividing an estate among heirs when they have different preferences for the property it contains? These are the kinds of questions that we shall examine in this book.

Equity is a complex idea that resists simple formulations. It is strongly shaped by cultural values, by precedent, and by the specific types of goods and burdens being distributed. To understand what equity means in a given situation we must therefore look at the contextual details. The book is built around seven illustrative cases: the demobilization of soldiers from the U.S. Army at the end of World War II; the allocation of kidneys among transplant patients; the apportionment of representation among political parties and states; the adjudication of conflicting property claims; the allocation of costs and benefits among participants in joint enterprises; the distribution of the tax burden; and the division of inheritances. This list is by no means exhaustive, nor is the treatment of any one topic complete. Indeed, I am well aware that I have stretched myself quite thinly over a number of different literatures, and probably have not done justice to any one of them. The goal, however, is not to give a definitive treatment of taxation, or political representation, or games of fair division. Rather, my aim is to call attention to common themes that cut across different areas of application, and to show that these pieces fit together into a larger picture. This has forced me to be selective in the treatment of various topics. I apologize to those who feel that the selections should have been made differently.

My interest in fair allocation grew out of a long and fruitful collaboration with my colleague Michel L. Balinski on apportioning seats in the United States House of Representatives. On the face of it the problem is simple. The Constitution stipulates that representation be allocated among the states in proportion to their populations, subject to the requirement that every state receive at least one seat. The apparent solution is to compute the exact proportional share that each state deserves, and round the results. On closer inspection, however, this

“solution” does not work because the rounded numbers typically do not add up to the required total. If we try to circumvent the problem by using some “pseudo-rounding” technique, we quickly run into a host of perplexing paradoxes and conundrums. The practical importance and theoretical challenge of this problem has engaged the attention of both statesmen and mathematicians for over two hundred years.

Many other problems of applied equity follow a similar pattern. What seems simple at first turns out to be riddled with puzzles and contradictions. Inevitably, we must turn to logical analysis to sort them out. The study of equity turns out, therefore, to have close ties with the axiomatic method in mathematics. From simple and intuitively plausible propositions about the meaning of equity, one draws general and sometimes surprising conclusions about the form that an equitable rule must take.

The axiomatic method has two weaknesses however. The first is that, while each axiom seems reasonable by itself, when piled on top of one another they almost inevitably lead to “impossibility” theorems. This confirms the skeptic’s predisposition to believe that the problem had no solution anyway. The proper conclusion, however, is that not all desirable conditions can be satisfied simultaneously. Some choice must be made. A second difficulty with the axiomatic method is that it can easily become disengaged from the problem that it was intended to solve. The invention of axioms and conditions is a fascinating business. The danger is that the exercise can take on a life of its own and lead to results that are mathematically elegant, but that have little or no relation to the realities of the underlying situation. To guard against this tendency I have tried to mix formal definitions and theorems with informal arguments and examples based on real data. The narrative therefore moves back and forth between observation and analysis, between theory and application, in an attempt to bring order to a complex subject without sacrificing too much realism.

The text is intended to be a “primer” and does not presuppose any mathematical background, though a taste for logical argument and some familiarity with economic concepts would certainly help. Technical definitions and arguments are relegated to footnotes and the appendix. The appendix does not cover all mathematical aspects of the subject, but focuses on a particular set of axioms and their consequences that are central to the argument. The text can be read independently of this material. Although there is a definite progression of ideas in the sequencing of the chapters, each can be read on its own for those who prefer to pick and choose among topics. Bibliographical notes are provided at the end for readers who want to go into greater depth.

1

Overview

Fair: equitably, honestly, impartially, justly; according to rule.

(Oxford English Dictionary)

1. The Division of Common Property

Every society has rules for sharing goods and burdens among its members. In primitive hunting tribes, for example, territory is usually held in common, and members of the community divide the kill according to their position in the social structure and their role in the hunt. These sharing rules, which in some cases are elaborately defined,¹ express a notion of *equity* in the division of jointly produced goods. By “equitable” I do not necessarily mean ethical or moral, but that which a given society considers to be *appropriate* to the need, status, and contribution of its various members. Appropriateness is shaped in part by principle and in part by precedent. It expresses what is reasonable and customary in a given distributive situation. To deviate from a rule that is founded on both reason and precedent would violate the stakeholders’ legitimate expectations, and this would be inequitable.

As economies become more complex, however, notions of common property and just shares gradually evolve into a system of private ownership. Resources that had formerly been held in common are divided up among individuals, and a complex legal edifice for protecting property rights and regulating their transfer springs into being. It is a common misconception, however, that this process of dividing common property eventually runs its course, and that in advanced economies virtually all property rights, or at least all important ones, have been privatized. A moment’s reflection will reveal that this is not the case. Property rights are constantly being created, destroyed, revised, and reassigned by soci-

¹ “Among the Birhors [a tribal group in east-central India] after a monkey hunt, once a chief has roasted a little of the meat and offered it to the spirits of the chase, the entrails, tails, and feet of the monkeys are divided equally among the men and women who served as beaters. Each man who provided a net also gets a hind leg. The two flankers each receive a hind leg as dividend, and the chief gets the neck and half of the back meat of each animal killed, in addition to his hunter’s share. If some meat is left after this distribution, it is divided into as many portions as there are eligible persons, plus an extra share for the chief” (Coon, 1971). Coon cites other examples of sharing rules based on contribution and status.

ety, and every time this happens a problem of distribution is encountered. In fact these situations often arise through the operation of the market itself.

One source of new property, for example, is technical innovation. Resources that were once outside the general system of ownership may acquire great value due to technological advances, and the community must then decide how to allot the rights to their use. An example is the spectrum of broadcasting frequencies for radio and satellite communications. Their allocation is governed by Article 33 of the 1965 International Telecommunications Convention, which stipulates that, because “radio frequencies and the geostationary orbit are limited natural resources . . . they must be used efficiently and economically so that countries . . . may have *equitable access* to both . . . according to their needs and the technical facilities at their disposal.” The meaning of equitable access is not spelled out, but has evolved over the years through periodic negotiations among the parties.²

Another example of “new” common property created by technical innovation is the deep ocean bed, which is strewn with nodules rich in nickel, cobalt, manganese, and other valuable minerals. The commercial mining of these nodules first became a realistic prospect in the 1970s, and the allocation of the mining rights was one of the principal topics addressed by the Law of the Sea Conference convened by the United Nations in 1973. The conference participants fell into two groups: industrialized countries such as the United States, West Germany, and Japan, that had the know-how and capital to mine the minerals, and the less developed countries (LDCs) that did not.³ The conference adopted the general principle, which had been enunciated earlier by the General Assembly, that the seabed is part of the common heritage of mankind—a “global commons” in which all countries have a stake. Nodule exploitation was to be undertaken on behalf of the international community. The Conference went on to propose the formation of two new agencies to supervise deep ocean mining: an International Seabed Authority that would license all mining activities, and an international mining entity known as the “Enterprise” to mine the seabed on behalf of the LDCs.

To protect against the possibility that commercial mining interests in the industrialized countries would steal a march on the Enterprise and lock up most of the desirable sites, the Law of the Sea Treaty adopted a variant of an allocation method sometimes used by children to share a piece of cake: one divides and the other chooses. In the context of the Treaty, the “cake” is the ocean bottom. Every time that a mining company applies to the Seabed Authority for permission to mine at a given location, it is required to develop two parallel sites from which the Enterprise can choose one. By this device the

² International Telecommunications Convention, 1973, article 33 [italics added]. See also Stern (1982).

³ The following account is based on Sebenius (1984).

Enterprise is guaranteed right of access to at least one-half of all developed sites, indeed, in their estimation they get the better half.

A second force behind the creation of new property is economic expansion. Traditionally this occurred by the conquest and settlement of new territory, which in feudal societies was distributed by the patronage of princes. In modern society a more complicated problem arises when economic expansion puts pressure on common resources that cannot be physically divided. Increasing air pollution, for example, has led some governments to restrict the quantities of noxious gases that may be released into the atmosphere. By so doing they establish limited rights to the use of the atmosphere and divide these rights among competing claimants. This may be done in a variety of ways. One is to place *uniform limits* on all polluters in a given category, for example, to restrict the amount of sulfur dioxide that each electric power plant can emit in a twenty-four-hour period. A second approach is to allot each polluter a quota of emissions permits, and allow them to trade among themselves. The latter solution is more efficient than the former (at least in theory), but it requires finding an acceptable basis for distributing the permits initially.

A concrete example is a market in sulfur-dioxide emissions permits that was recently established among electric power utilities in the United States. Beginning in 1992, each electric power plant will receive a quota of permits that is proportional to its emissions over a three-year base period. The permits may be traded within certain geographical regions. This allocation is subject, however, to a maximum rate beyond which a plant does not qualify for any more permits. The idea is that past usage defines each plant's relative entitlement up to a point, but a ceiling is imposed that effectively penalizes dirty plants. One could say that this solution is equitable in the sense that it imposes a similar burden of reduction on all plants relative to their status quo position provided that their status quo position is justifiable.

Third, new common property results from the formation of new enterprises. When two firms engage in a joint venture, for example, they must decide how to split the profits from their partnership. Doctors who share an office must divide the cost of equipment, secretarial help, and rent. Similarly, international organizations like NATO and the U.N. who provide various goods and services jointly must decide how to divide the costs among their members.

Finally, many kinds of common property are not privatized even in market economies, but are allocated by society on a contingent basis. The right to vote, the duty to serve in the army, educational opportunities, and health care are rarely left to the workings of the market, but are allocated to specific individuals according to notions of merit or desert, and cannot be transferred to others.

In spite of the fact that distributive issues permeate both economic and political life, they have not been studied very systematically by economists or political scientists. There are exceptions, of course, but I think it is fair to say

that distributive justice has not been a central topic in political economy since the decline of classical utilitarianism in the early part of this century.⁴ In part the neglect is due to the lack of a coherent analytical framework for studying distributive issues that is compatible with the modern ordinalist approach to analyzing economic welfare. In part it is due to the way in which economic theory happens to have developed. The standard model of economic exchange posits that all resources and property rights—including ownership of firms—are fully delineated and vested in individuals. The focus is on the process by which resources are allocated to their most efficient uses through the utility-maximizing choices of individuals. Thus, right at the outset, the issue of how common property should be divided *ex ante* is side-stepped. Somehow property *has been* divided, and one goes on from there to analyze how markets allocate resources efficiently. The division of common property introduces a complicating element into this story that most economists have chosen to ignore.

The fact is, however, that common property is continually being created by the market system itself through economic expansion, the advancement of technology, and the formation of cooperative enterprises. And many of the most important goods and burdens in society are allocated by political institutions rather than by the market. So it is fallacious to assume that the allocation of common property is a once-and-for-all affair; it is an ongoing process.

2. Micro vs. Macro Justice

At this point a caveat is in order. As I said in the preface, we shall not be concerned here with distributive justice in the large, that is, with the question of what constitutes a just social order. Rather, we are interested in distributive justice in the small, that is, how institutions divide specific types of benefits and burdens. Social justice is a crucial issue, of course, because it is ultimately concerned with the legitimacy of different forms of government. For various reasons, however, theories of justice in the large have little to say about what it means in the small. They do not tell us how to solve concrete, everyday distributive problems such as how to adjudicate a property dispute, who should get into medical school, or how much to charge for a subway ride.

The reason for this failure is that issues of local justice tend to be compartmentalized. Society makes no effort to coordinate distributive decisions across different domains, nor would it make sense to try. Kidney agencies do not give precedence to patients who failed to get into a university; a person's tax bill is not reduced because his party received less than its quota of seats in the last election; a soldier does not qualify for extra leave because the government located a waste dump near his home. There is no mechanism comparable to the

⁴ See particularly Baumol (1986), Moulin (1988), and Elster (1992).

invisible hand of the market for coordinating distributive decisions at the micro level into just outcomes at the macro level. To the extent that gross inequities exist because of the cumulative effects of local decisions, they can be rectified at the societal level by redistributive policies, not by trying to coordinate the decisions of local allocative institutions.

The upshot of this compartmentalization is that each type of fair division problem is treated on its own merits based on the norms and precedents that have evolved to deal with that particular class of situations. Property claims are dealt with according to precedents in property law, taxes according to traditions in public finance, kidneys according to standards in the medical profession. The relevant standards differ appreciably from one domain to another and from one culture to another. While generalizations are difficult, however, they are not impossible: similar concepts keep reasserting themselves in diverse applications. This fact allows us to develop a common language, and a set of principles, that apply to a wide variety of distributive problems.

3. The Terms of Discussion

An *allocation problem* arises whenever a bundle of resources, rights, burdens, or costs is temporarily held in common by a group of individuals and must be allotted to them individually. An *allocation* or *distribution* is an assignment of the objects to specific individuals. Allocation is not the same thing as exchange. An allocation is a decision about who gets a good or who bears a burden, and is usually decided by a group or by an institution acting on behalf of the group. Exchange involves many voluntary, decentralized transactions, and can only occur after the goods and burdens have been allocated. Allocation comes first, exchange afterward.

Allocations often involve burdens instead of goods, but we shall usually speak in terms of goods. Moreover, this convention involves no loss of generality, because exemption from a burden is a good. Admittedly, there are situations in which it is natural to focus on the negative rather than the positive side of the ledger, such as taxation, military duty, and layoffs. Furthermore, perceptions of equity may sometimes be affected by this distinction. It undoubtedly makes a difference, for example, whether we frame the job allocation issue as: Who should be fired? or Who should be hired? For purely expository purposes, however, there is no harm in talking about goods rather than burdens.

Goods come in many forms. They may be homogeneous and divisible, like money or water. They may be inhomogeneous and divisible, like land, broadcasting bandwidths, or fishing grounds. They may be homogeneous and indivisible, like seats in a legislature or exemptions from military duty. Or they may be heterogeneous and indivisible, like kidneys for transplantation, places in selective colleges, and jobs.

The supply of the good may be either fixed or variable. The number of seats in the United States House of Representatives, for example, is fixed by statute at 435. The number of kidneys available for transplantation varies from one day to the next. It is important to recognize that the quantity of the good may be affected by the procedure for distributing it. The structure of the income tax code certainly affects the amount that individuals earn and the degree to which they report it, hence it determines the total amount of taxes raised. The perceived fairness of the kidney allocation formula may affect the propensity of donors to give.

More generally, an allocation is usually the result of three different types of decisions.⁵ One decision concerns the total amount of the good to be distributed. Determining the number of seats in the House or the amount to be levied by taxation are examples of such *supply decisions*. The second type of decision concerns the formula or principle by which the good is allotted among the eligible parties. The formula for distributing legislative seats and the personal income tax schedule are examples of such *distributive decisions*. The first and second types of decisions are normally made by institutions. The third type of decision is made by individuals in response to these institutional choices. A taxpayer's choice about how much to earn given the tax rate schedule is an example of a *reactive decision*. The combination of all three levels of decision-making yields an *effective* allocation.

The major focus of this study will be on decisions of the second type, that is, on rules of distribution and the principles invoked to justify such rules. In general an *allocation rule* is a method, process, or formula that allocates any given *supply* of goods among any potential *group* of claimants according to the salient *characteristics* of those claimants. An apportionment rule, for example, determines how many seats each party gets in a legislature as a function of the party vote totals and the total number of seats. A rule of taxation specifies the amount of tax owed by each household as a function of the household's ability to pay and the amount of tax that must be raised. A kidney distribution rule determines who gets the next kidney as a function of each patient's medical condition, how long he or she has been waiting, the suitability of the kidney for the patient, and other factors.

The allocation rules that we see in practice usually exhibit one of three broad conceptions of equity. *Parity* means that the claimants are treated equally, either because they actually are equal or because there is no clear way to distinguish among them. *Proportionality* acknowledges differences among the claimants and divides the good in proportion to these differences. *Priority* asserts that the person with the greatest claim to the good gets it. Parity, proportionality, and

⁵ This taxonomy (though not the terminology) draws on Calabresi and Bobbit (1978) and Elster (1992).

priority also figure prominently in the major theories of distributive justice, as we shall see in the next section.

While these conceptions of equity describe the general *structure* of a rule, however, its *content* derives from specific normative principles. Consider the problem of deciding who has priority to get the next available kidney. On philosophical grounds one could argue that the person who is worst off has highest priority, which suggests giving it to the patient who is in greatest danger or pain. Alternatively, one could argue that the kidney should go where it is likely to do the most good, for example, to the one most likely to survive. Yet a third option is to appeal to a social norm, like first come first served, that derives its legitimacy from precedent. This suggests giving it to the one who has been waiting longest.

These *normative* principles may be contrasted with the ways that distributive decisions are made in practice. Who actually does get the kidney? Who does the company lay off when times are bad? Who gets custody of the child in a divorce? These *empirical rules* of equity, as revealed by the choices that institutions make, are usually more complex and nuanced than any single normative principle. Instead they often represent a balance or compromise between competing principles.

4. Normative Theories of Justice: Aristotle, Bentham, and Rawls

There are three general theories of justice that figure prominently in discussions about equity. The oldest and most prominent is Aristotle's *equity principle*, which states that goods should be divided in proportion to each claimant's contribution. Plausible as this idea may be, it has two substantial limitations. First, we must have some way to measure the contribution of each claimant on a cardinal scale. Sometimes such a measure is natural, for example, the amount of time each worker put into a joint effort. In other situations the measure of contribution is not so clear. In a divorce proceeding, for example, how does one ascertain the relative contribution that husband and wife made to their joint estate or to rearing their children? Second, for proportionality to be workable, the goods must be divisible. When they are not, one could make them divisible by distributing chances at receiving the good, but then proportionality loses some of its plausibility. For example, if *A* has fought twice as long in the army as *B* and only one of them can be discharged, should *A* get *twice* as many chances at being discharged as *B*? Or is it fairer simply to discharge *A* first because he fought longer than *B*?

A second theory of justice is classical utilitarianism, which asserts that goods should be distributed so as to maximize the total welfare of the claimants (the *greatest good for the greatest number*). For this notion to make sense, utility must be understood as a measure of psychic satisfaction or well-being that can be measured on a cardinal scale and added across individuals. This doctrine,

which was the dominant one in nineteenth-century welfare economics, has fallen into disrepute for several reasons. First, no method is provided for comparing levels of satisfaction among different individuals. Modern utility theory defines an individual's utility solely in terms of his preferences for different states of the world: state x has higher utility than state y if and only if x is strictly preferred to y . In this revealed preference approach to utility, the units in which utility is measured are quite arbitrary, so it is meaningless to add and subtract them across individuals. Second, even if we could devise some method for comparing individual utilities, it is not clear that the utilitarian *principle* is ethically sound, since it might require imposing great harm on a few in order to confer a small benefit on the many.

A third approach to social justice that meets these objections to some extent is due to John Rawls (1971). It is impossible to do justice in this short space to Rawls's theory, which is intricately constructed and contains many subtle and important qualifications. However, the central distributive principle may be simply stated: the least well-off group in society should be made as well off as possible. This is known as the *maximin or difference* principle. Contrary to first appearances it is not a welfarist conception of justice, because "well-off" does not refer here to a person's subjective level of satisfaction. Rather, it refers to the *means or instruments* by which satisfaction or happiness can be achieved. Economic income is one such means; others include opportunity, power, and self-respect. Rawls calls these *primary goods*.

If we restrict our attention just to one primary good, such as income, the principle says that income should be distributed so that the person with the least income has as much income as possible. This does not necessarily imply that everyone has equal income, however, because redistributing income from rich to poor may reduce or eliminate the incentive for the rich to become rich, thereby impoverishing everyone. Rather, the principle refers to the effective distribution of income after economic incentives are taken into account. Thus Rawls's principle avoids two of the problems inherent in classical utilitarianism. First, it is based, at least in part, on observable characteristics of individuals (such as income) rather than on interpersonal comparisons of welfare. Second, it avoids the ethical problem of benefiting the many at the expense of the few.

On the other hand, Rawls's theory is not without its drawbacks. In the first place, income is not the only currency in which justice is evaluated: other primary goods (e.g., self-respect) are involved that do not lend themselves readily to objective comparisons. Moreover, even if we could make such comparisons, it is not clear how we should *weight* different primary goods. Second, and most important, it is not clear that the maximin principle satisfies our intuitions about justice. Is it just to impose serious inconveniences on almost everyone in society in order to raise the opportunities, the income, or the self-respect of the least fortunate by a miniscule amount?

5. No Envy

The conceptual difficulties posed by the utilitarian and Rawlsian principles have led some economists to adopt an entirely different approach to distributive justice. Here we shall briefly sketch the central idea, which will be taken up again in chapter 9. A distribution is said to be *envy-free* if no one prefers another's portion to his own. This concept does not require interpersonal comparisons of utility, because each person evaluates every other person's share in terms of his *own* utility function.

Envy-freeness is an admirable—indeed an enviable—idea. It was first proposed in a very strong form by Tinbergen (1953), who argued that an equitable society is one in which no person wants to *change places* with another. “By places we mean their whole situation: especially their occupations and incomes, but also personal conditions such as health, size and health of their families, levels of education, abilities, etc.”⁶ Attractive as this concept may be in theory, it is impossible to achieve if people envy the personal characteristics of others, which is probably the most common object of envy. For example, *A* might envy *B* because *B* is tall. To eliminate *A*'s envy requires that *A* be made wealthier than *B*. On the other hand, *B* might be indifferent between being tall and short. Hence if *A* is compensated, *B* would be envious. So there may be no reasonable way of eliminating envy in Tinbergen's sense.

A more pragmatic formulation was subsequently suggested by Foley (1967). We do not require that society *in general* be envy-free; we only require that no person prefer another's portion *of a particular allocation of goods*. If an estate is being distributed among heirs, for example, the “no envy” criterion says that no heir should prefer another's portion of the property to his own. They might envy each other because of other goods that they own, or because of their different abilities and circumstances of life, but not because someone else received a more desirable portion.

This idea makes sense provided that the parties have equal claims on the goods, and the goods are divisible. An example would be an inheritance in which the heirs are bequeathed equal shares. Of course, if the estate contains only one homogeneous good, such as money, an allocation is envy-free if and only if it is divided equally. Hence there are no novel implications in this case. The principle takes on greater interest when the property consists of different kinds of goods. Suppose, for example, that *A* and *B* are equal heirs to an estate that contains \$200,000 in cash and 300 acres of land (all of which is of similar quality). Suppose further that *A* values the land at \$1000 an acre and *B* values it at \$500 an acre. If *A* gets all of the land and *B* all of the money, each prefers his

⁶ Tinbergen (1953), pp. 55–56. The translation is given in Thomson (1990c), who provides an excellent survey of the literature on envy-freeness and related concepts. See also Thomson and Varian (1985) and Baumol (1986).

own portion to the other's, so the allocation is envy-free. Moreover, it is efficient because they have no desire to trade.

Typically, in fact, there are many allocations that are both envy-free and efficient. In the above example, there is no envy if *A* gets 250 acres and no cash, while *B* gets all the cash plus 50 acres of land. There is also no envy if *A* gets all the land plus \$25,000, while *B* gets \$175,000 but no land. These and all distributions in between are both envy-free and efficient, but it is not clear that all of them are equally fair.

One way to resolve this problem of indeterminacy is to resort to an allocation process that is perceived to be fair by both parties. Consider the traditional method of divide and choose. By the toss of a fair coin one of them is designated to be divider (say *A*). Since *B* gets to select the portion she prefers, she will certainly not envy *A*'s portion. However, *A* can also protect himself from envy by creating two portions that he values equally. Then no matter which portion *B* selects, *A* will not strictly prefer it to the one left over for himself. Note, however, that the divider can manipulate the outcome if he knows the chooser's preferences. For example, *A* could split the estate into the following two portions: one containing \$175,001 and no land, the other containing all the land plus the rest of the money. *B* slightly prefers the first portion to the second, while *A* greatly prefers the second portion to the first. The resulting allocation is envy-free and efficient, but is it fair? This suggests looking for equity principles that discriminate more finely among envy-free allocations, a topic that we explore in chapter 9.

6. Distributive Judgments and Interpersonal Comparisons

The appeal of no envy is that it does not require making interpersonal comparisons of utility, so it has an operational meaning within the framework of modern utility theory. The weakness of the concept is that it only applies when the parties have equal claims on the good, which is often not the case. Indeed, most fair division problems revolve around the question of how *differences* in claims—due to disparities in merit, desert, contribution, need, and so forth—should be taken into account, in which case the no envy principle is more or less irrelevant.

To take a concrete example, consider the problem of allocating one kidney between two patients. Suppose that patient *A* is in her thirties and has no other major health problems, while *B* is in his sixties and is diabetic. On the other hand, *B* has been waiting for two years and cannot survive much longer on dialysis, while *A* has been in the queue for less than two months and can wait longer. Under these circumstances who should receive the transplant? Obviously, whoever does not get the kidney will envy the other in the sense that he or she prefers the other's portion. (Note, however, that *A* may well prefer to be *A*

without a kidney than to be *B* with one.) From society's point of view, envy is not the most relevant issue here. What matters is that medical resources be put to effective use, and that suffering be alleviated. These principles sometimes conflict: the former argues for giving the kidney to *A*, while the latter suggests giving it to *B*. While these choices are not simple, a decision that is objectively based on such values may well be regarded as fair by the losing party, even though in a narrow sense the loser has reason to envy the winner.

The problem with no envy is that it dispenses with interpersonal utility comparisons by not making interpersonal comparisons of any kind. The fact is, however, that distributive decisions almost always involve comparisons and value judgments. It is pointless to assume them away; we need instead to ask *on what basis* they are made. When a doctor or a social agency decides to give a kidney to patient *A* rather than to patient *B*, a value judgment is being made and people are being compared. The decision is not about who wants it most, but about who needs it most or who will benefit most. These judgments are not made by consulting the parties' utility functions, but by looking at objective measures such as expected longevity and time in the queue. Any theory of equity with explanatory power is going to have to come to terms with the fact that we make comparative judgments of this sort all the time. Moreover, there is nothing in *utility theory* that says that such comparisons cannot be made; the theory is simply silent on *how* they are made.

7. Why Classical Formulas Fail

The book is constructed around a stock of examples that illustrate why equity cannot be reduced to simple, all-embracing solutions such as the difference principle, the greatest good principle, or the proportionality principle. To illustrate why these ideas do not get us very far, consider the situation in which a single indivisible object is to be allotted among the members of a group, all of whom lay claim to it to different degrees. For example, which parent gets custody of the child in a divorce? Which soldier gets the privilege of being discharged from the army first? Which patient with renal disease gets the one available kidney? What site should be chosen for a hazardous waste dump? Who inherits the summer house?

One might think that the way out of these difficulties is to transform the indivisible good into a divisible one. In a moment we shall argue that this does not resolve the equity problem, but first let us consider how it could be done. There are three standard methods for transforming an indivisible good into a divisible one: randomization, rotation, and conversion. Under *randomization* each claimant has a probability of getting the good; under *rotation* the claimants take turns at using the good; under *conversion* they exchange the indivisible good for a divisible one and split the proceeds.

The relative suitability of these methods depends on the circumstances. A kidney might be allocated by lottery, for example, but it cannot be shared effectively. By contrast, the heirs to a summer house would probably prefer to share it than take equal chances at getting exclusive use of it, and they might prefer selling it to sharing it if they cannot get along. Rotation is customary in resolving child custody disputes, while lotteries and sale would be considered quite unusual.

Yet another way of handling indivisibilities is to *compensate* those who do not get the good, or, in the case of a burden, to compensate those who do. For example, one might auction off the good among the claimants: the highest bidder gets it and pays appropriate compensation to the losers. Similarly a burden, such as a waste dump, may be auctioned by having each party offer to take on the burden in return for compensation. The one who demands the least compensation accepts the burden and the others divide the cost.

Each of these mechanisms—randomization, rotation, conversion, and compensation—changes the original allocation problem into a new one that involves divisible goods. But none of them goes to the heart of the distributive issue, which is: how much is each claimant entitled to? Does everyone who avoids getting the waste dump pay an equal portion of the cost? Does everyone who has to live with it receive equal compensation? Does patient *A* deserve twice as many chances at getting the kidney as patient *B* because she has been waiting twice as long? Should the mother get five days of custody per week and the father only two? On what theory or principle are these shares determined? This is the nub of the distributive problem, and it is not resolved by introducing divisible goods.

8. The Priority Principle

The evidence suggests, moreover, that indivisible allocations are often handled by confronting the indivisibility directly instead of trying to circumvent it. He who has the greatest claim gets the good; the others do not. We shall call this the *priority principle*. The basis of priority may be quite simple, such as seniority in avoiding job layoffs. It may follow from economic considerations, for example, choose the waste dump site that is least costly. More often priority is based on a mixture of criteria. What is the best site given both cost and safety considerations? Which patient should get the kidney given differences in their chances of survival, time in the queue, and immediate need? (All of these factors enter into the national formula for allocating kidneys in the United States.) Priority in these situations is not one-dimensional, it involves trade-offs among various principles.

This pattern holds true for a wide variety of indivisible goods. There are lists for getting into day care centers, for being admitted to universities, for being

hired, for being laid off, for receiving public housing, for getting into a nursing home. Each such list expresses a concept of who is most deserving, who is next-most deserving, and so forth, given the claimants' circumstances and the good being distributed. Each captures a notion of equity, but it is not equity in the Aristotelian sense of proportionality. It is equity based on *priority*. Priority is an ordinal rather than a cardinal principle because it does not say *how much more* deserving one claimant is compared to another; it simply says that one claimant *is* more deserving than another.

How then is priority determined in a concrete case? In some situations priority is largely a matter of judgment. Deciding which soldier is most deserving of early discharge, which parent wins custody, or who gets the kidney are issues on which reasonable people may differ. In this case fairness reduces to a procedural question of how to strike an *equitable balance* between diverse points of view. This brings us to a classical problem in group decision making, namely, how to design a process that fairly aggregates individual opinions into a collective decision. This project is not without its perils, because collective decisions are inherently more problematic than individual ones due to the "paradox of voting." Nevertheless there do exist equitable ways of aggregating individual opinions into a consensus, as we shall argue in chapter 2.

In other situations, priority is less a matter of judgment and more a matter of logic. An example is the apportionment of representation among political states or parties (chapter 3). Here the equitable ideal is clear—one person, one vote. But in practice the ideal cannot be met due to the lumpy character of the goods. Seats cannot be subdivided. The issue then becomes how to achieve equity *as near as may be*.

9. The Consistency Principle

A fruitful approach to this problem is to begin by asking what solution is most equitable when there are just *two* claimants. This case is usually simple to grasp intuitively. Once a standard of equity has been established for two-claimant situations, we may then solve a many-claimant problem according to the following principle: *allocate the good so that every two claimants divide the amount allotted to them as they would if they were the only two claimants*. This *consistency principle* turns out to be one of the most powerful unifying ideas in the theory of fair allocation.

To illustrate the concept in a concrete case, consider the apportionment of legislative seats among states. Here the *ideal* is one person, one vote. Each state's allotment should be exactly proportional to its population. Unfortunately the ideal cannot be met because the seats come in whole units. To solve this problem, we begin by asking what apportionment comes closest to meeting the ideal when there are just two states. The natural answer is to give each state the

whole number *closest* to its exact share. In other words, we round the shares to the nearest whole number.

For example, suppose there are ten seats to divide between two states with populations 142,000 and 858,000, respectively. The share of the first state is 1.42 and the share of the second is 8.58. The natural solution is to give 1 seat to the first and 9 seats to the second. This satisfies the ideal of one person, one vote as near as may be. When there are more than two states we follow the same principle, that is, we distribute the seats so that every pair of states shares the number of seats allotted to them as near as may be. It is a remarkable fact that such a distribution always exists. Moreover, it is computed by constructing a priority list that determines which state gets the first seat, which state gets the second seat, and so forth until all the seats have been distributed. This is an instance of a very general result, which states roughly that *any allocative rule that satisfies the consistency principle also satisfies the priority principle and vice versa*.

These results are not mere theoretical curiosities. The rule described above, for example, was used to apportion the United States House of Representatives for much of the nineteenth and early twentieth centuries, and it is now one of the standard methods for apportioning legislatures in other countries. Thus we see how a few well-chosen principles can be used to deduce practical solutions to difficult allocative problems. We also see why classical formulas like pure proportionality do not take us as far as we need to go.

10. When Proportionality Fails for Divisible Goods

It might be objected, however, that the above examples are special because the bone of contention is indivisible. Surely when the good is *divisible* and the claimants can be compared by some numerical measure of *entitlement*, classical principles like proportionality work very well. In chapters 4, 5, and 6 we argue, however, that even in this case, proportionality is not necessarily the most appropriate solution.

Suppose, for example, that *A* and *B* are heirs to an estate that contains a field. One document says that *A* gets the whole field, while another document says that *B* gets the eastern half of the field. The documents are inconsistent, but equally valid. How should the inconsistency be resolved? An equitable solution is to give the western half of the field to *A* (since it is undisputed) and to divide the disputed eastern half equally between them. Thus *A* would get three-fourths of the field and *B* one-fourth.

Now consider a slightly different situation: *A* is willed all of the field and *B* is willed half of the field, but the document does not say *which* half. Under the preceding principle *A* would get three-fourths of the field and *B* one-fourth. But the proportional principle would award two-thirds of the field to *A* and one-third to *B*, because *A* was willed twice as much land as *B*. This seems no more

equitable than the first approach, which is a well-known principle in the Talmudic law of contracts (see chapter 4). This does not prove that proportionality is inequitable, but it does show that proportionality is not the *only* equitable solution.

A second difficulty with proportionality (and many other rules) is its failure to take economic incentives into account. Consider, for example, two communities who plan to share the cost of a common waste disposal site. The total development cost is one million dollars. If they fail to agree, town *A* will develop its own site for \$700,000 and town *B* will develop a separate site for \$400,000. Thus there are economies from cooperation. Suppose for the sake of argument that *A* generates three times as much waste as *B*. (*A*'s site does not cost three times as much because of economies of scale.) The Aristotelian solution would be to charge in proportion to their contributions: *A* would pay \$750,000 and *B* would pay \$250,000. *A* is not likely to agree to this proposal, however, because it would be cheaper to go it alone. Thus proportional division is not satisfactory because it fails to take into account the parties' opportunity costs. Equal division fails for the same reason. This issue is taken up in chapter 5, where we show how to define equity in the presence of incentive constraints.

Perhaps the most serious problem with proportionality, however, is that it does not always accord with our *intuitions* about equity. A prominent example is income taxation. Many developed countries tax higher incomes at a higher rate than lower incomes, and have done so almost since the inception of income taxation. This *progressivity principle* is by no means unique to taxation, but is seen in many other kinds of assessments. When hard times force firms to cut wages, for example, it is not unusual for the cuts to be progressive, with higher-paid employees suffering a greater percentage reduction than lower-paid ones. In both cases the rationale is clear: those who are better off can absorb the loss more easily. Of course, taxation is a more complex problem than a one-time assessment, because the tax schedule casts a shadow on the future that affects the propensity of taxpayers to work and invest. Hence distributive goals must be considered in relation to their incentive effects. It turns out, however, that some types of progressive schedules are no worse in this respect than flat-rate taxes; indeed they may even lead to fewer distortions, as we argue in chapter 6.

Each of these cases illustrates why proportionality (as well as the other classical principles) do not always yield satisfactory answers, even when the goods are perfectly divisible. If we want to explain what equity means in practice, we must look for more subtle and nuanced kinds of solutions.

11. Games of Fair Division

In chapters 7 and 8, we turn our attention to situations where the allocative decision is not entrusted to an institution, but is negotiated by the claimants directly. Suppose, for example, that a group of countries are negotiating their

respective shares in a producers' cartel. The purpose is to limit production and drive up the price so each country naturally wants to claim as large a share of the output as possible. The bargaining chip that each holds is the threat to pull out of the agreement and drive down the price by increasing production unilaterally.

At first blush it would seem that equity considerations play little or no role in such bargains. What determines the outcome is the power relation among the claimants: the skill and patience with which they bargain and the credibility of their alternatives. While these factors are important, however, they rarely determine the outcome. Indeed, if everyone simply demands the maximum amount for himself, there is a good chance that the process will end in stalemate. The key to resolving a distributive bargain is not to make self-serving demands, but to make a proposal that the *others* find plausible and justifiable.⁷ This is precisely where equity arguments come in: they *coordinate* the expectations of the bargainers by establishing a plausible basis for the agreement (see chap. 7). Equity principles are the *instruments* that people use to resolve distributive bargains.

Of course there are some situations in which the claimants may find it difficult, if not impossible, to agree in principle. Suppose, for example, that several heirs are bequeathed shares of an estate that consists of various types of divisible goods. Since the shares themselves are not in dispute, each claimant could simply take his share of each of the goods and go home. But this solution is almost certainly inefficient if they value the goods differently. Is there some basis on which they could reach agreement on a distribution that is both fair and efficient? The answer to this question is not immediately apparent. Indeed, since they almost certainly do not know each other's utility functions, they cannot even determine what the efficient outcomes are.

The solution is to design a *procedure* for dividing the goods that the claimants believe to be fair. By a "procedure" we mean a game with prescribed rules and moves that results in a specific division. A classical example is the game in which one player divides the property into two piles and the other gets first choice. How do we know that this game is fair? One answer is that it leads to fair outcomes when played by rational players. For example, the divide and choose game yields a division that is envy-free, so we could argue that it is fair. A second answer is that no advantage accrues from playing one role instead of another. This is analogous to saying that the playing field is *level*: the players can expect to do as well no matter which end of the field they have to play on. In this sense divide and choose is *not* fair because it confers an advantage on the divider. This inequity can be eliminated by changing the game in various ways, as we show in chapter 8. One variation is to allow the role of divider and chooser to alternate between the two players, with the chooser in one round becoming divider in the next round if he does not like the divider's current offer.

⁷ Schelling (1960).

A second approach is to let both parties bid to be divider. Each bid is a fraction between 0 and 1 that represents the bidder's commitment to give up that fraction of all the goods if the other side does not like his proposed division.

These and other modifications of divide and choose make the playing field more or less level. However, they also lead to substantially different outcomes when played by rational, knowledgeable agents. To evaluate which of these processes is fairest, we must have some way to evaluate the fairness of the outcomes they produce. This raises a difficulty, because the equilibrium outcomes of these games are framed in terms of their *utility* to the claimants, which is almost never accessible information. Indeed, the players will not even be able to *reach* these outcomes unless their utility functions are common knowledge, which is almost never the case. Thus, while some games of fair division yield outcomes that are fair in theory, there is no guarantee that these outcomes will result in practice.

12. Equity and Efficiency

This leads us to a natural question. Is there any way to design an allocation procedure that leads to outcomes which are *visibly* fair and efficient, and does not require that the claimants know each other's utility functions? In chapter 9 we suggest an answer to this question. Suppose that the claimants have well-defined shares in the common property (not necessarily equal shares), and different preferences for the goods it contains. A *competitive allocation* is one for which there exists a set of *prices*, such that every claimant likes his portion best among all the portions that he can afford to buy given the value of his share at these prices. Such an allocation can be discovered through a marketlike mechanism that does not require the claimants to know anything about each other's utilities. Moreover, the resulting allocation can be justified on grounds of equity. *It is the only efficient and consistent way of reallocating the property that leaves everyone at least as well off as he was initially.*

This result provides both a theoretical and practical answer to the question of how to allocate divisible goods both fairly and efficiently without making inordinate demands on the players' information. It also sounds a theme that runs throughout the book. In problems of *local* justice, equity and efficiency often complement each other. Principles of equity are the *instruments* by which societies resolve distributive problems when efficiency by itself yields indeterminate results.