1 Microeconomic Foundations

1.1 Fairness: Equal and Unequal Treatment

Justice is blind, and fairness requires anonymous rules of arbitration. "Equals should be treated equally, and unequals unequally, in proportion to relevant similarities and differences" (*Nicomachean Ethics*). Aristotle's celebrated maxim, in its modern rendition, is the first step toward the formal definition of distributive fairness.

Equal treatment of equals is a clear-cut principle, an axiom in the terminology of this book, and the minimal test of fairness throughout the book: if two persons have identical characteristics in all dimensions relevant to the allocation problem at hand, they should receive the same treatment—the same share of goods, of decision power, or of whatever is being distributed. *Unequal treatment of unequals,* on the contrary, is a vague principle open to many interpretations.

Four elementary ideas are at the heart of most nontechnical discussions of distributive justice. They organize neatly our thinking about conflictual interpretations of fairness. The four ideas are *exogenous rights, compensation, reward,* and *fitness;* I discuss them briefly below in that order. A more detailed discussion is the subject of section 2.1.

Equal treatment of equals is the archetypal example of an exogeneous right. Consider the democratic principle "one person, one vote," expressing the basic requirement that voting procedures must not be biased in favor of particular electors, and implemented by the simple device of anonymous ballots (I can't tell who casts which ballot; hence I can't give more weight to the vote of a particular citizen).

One could argue that certain differences among voters are entirely relevant to the way that we should evaluate their votes. At one extreme, we find *conscientious* citizens who collect information about the issues at hand and discuss them with other citizens; at the other extreme, the *whimsical* citizens use their voting right arbitrarily, for spurious reasons or no reason at all, and refuse to make any effort toward gaining some understanding of the issues. Surely the difference between conscientious and whimsical voters affect the normative "value" of their respective ballots, a fact that the "one person, one vote" principle refuses to acknowledge. The exogeneous right to an equal vote is a normative postulate that suffers very few exceptions (insanity, certain criminal records). Other times saw other postulates: medieval religious assemblies typically gave more weight to the vote of their more senior members, and voting rights were commonly linked to wealth throughout the nineteenth century.

When differences in individual characteristics are deemed relevant to fairness, the two ideas of compensation and reward come into play. In order to meet their needs of some essential commodities, such as the regular intake of some vitamins, adequate shelter and clothing—some of us need more resources than others—and the compensation principle justifies this inequality in order to restore equality (or at least, diminish inequality) of the

shares of the essential commodity in question. Those of us who cannot metabolize a key vitamin from their food deserve a free supply in pills, those who lost their home to a natural disaster deserve assistance, those who did not receive adequate basic education deserve a tuition waiver, and so on. The common feature in these instances of compensation is to justify unequal shares of certain resources for the sake of equalizing the shares of a higher-order commodity.

We reward an employee for her contribution to the profit of the firm, an athlete for his contribution to the success of the team, an investor for the risk he took in financing the project. In each case a larger share of the pie is justified by a larger responsibility in making the pie. Symmetrically I must bear a share of joint costs commensurate to my own impact on these costs: businesses sharing a parking lot find it fair to split its maintenance cost in proportion to the traffic that each business generates.

To illustrate the tensions among the three principles above, consider the division of the parent's estate between siblings. Compensation suggests to give more to the poorest sibling, who needs more the extra cash. The devoted child who took care of the parents' business deserves to be rewarded by a bigger share. Finally, strict equality of the shares, no matter what, is a popular and haggle-free division method according to exogeneous rights.[1]

The last commonsense principle of distributive justice, fitness, is the most subtle of the four, and the most fertile ground for microeconomic analysis. Fitness says to give the resources to whomever makes the best use of it: the flute to the flutist, the books to the avid reader, the voting right to the wiser among us, and so on. The difficulty is the interpretation of the "best" usage. If Ann can play popular melodies on the flute that most everybody enjoys, whereas Bob can play esoteric atonal music that a small minority loves intensely, who should get the instrument? The interpretation of the "wise" voters, or the "best" reader is equally fuzzy. In microeconomic language the problem is to make intercomparison of individual welfares, and deal with trade-offs such as the mild pleasure of many when Ann is playing versus the intense aesthetic emotions of a few when Bob does.

The subject of this book is the contribution of modern microeconomic thinking to distributive justice. I submit that this approach is the only conceptual framework to date in which not only fitness (as just described) but also compensation and reward can be analyzed with mathematical rigor.

The main building block is the concept of individual welfare, modeled first with the help of a utility function transforming physical resources into utility, and more generally, some resources devoid of ethical content into a higher-order resource, which is the object

1. Giving more to the more loving child who took care of his parents' needs is also an exogeneous right, inasmuch as the loving care is unrelated to the value of the estate.

of our normative judgment. Thus compensation amounts to the equalization of individual utilities and the maximization of their sum is the interpretation of fitness known as classical utilitarianism: chapters 2 and 3.

A more versatile modeling of welfare is as an ordinal preference relation, describing individual choices but pointedly avoiding the cardinal statement conveyed by the measurement of utility: chapters 4, 6, and 7.

The collective welfare approach remains the most influential application of economic analysis to distributive justice. It offers the first entirely rigorous definition of fitness as Pareto optimality, and operational definitions of fairness by means of collective utility functions—in the cardinal version—and of social welfare orderings—in the ordinal version. It is the subject of chapters 2, 3, and 4, and its methodological premises are reviewed in the next two sections.

If the collective welfare approach yields a systematic interpretation of the fitness and compensation principles, it gives no basis to discuss either reward or exogeneous rights such as private or common ownership of resources. In the last fifty years microeconomic theory has made significant progress toward understanding these two pervasive principles. The two most important discoveries are the formula called the Shapley value (chapters 5 and 6) and the test of no envy (chapters 6 and 7): they offer, respectively, an interpretation of the reward principle in the context of production, and of fair division of resources under common ownership.

In sections 1.4 and 1.5, I introduce the microeconomic background in which these two concepts are born, namely the management of externalities by private agreements (free trade under private ownership) or public contracts (normative principles of equity). Then in section 1.6, I describe the organization of the book and review the contents of its successive chapters. Section 1.7 concludes with an introduction to the literature.

1.2 Collective Welfare: Cardinal

The postulate of rational choice for individual decision-making is the central assumption of microeconomic analysis. It posits that individual choices consist in maximizing a given preference relation (a complete and transitive binary relation over the feasible choices) representing the welfare of the agent in question. Outside economics, this postulate is not universally accepted, and neither is its prerequisite of methodological individualism.

Methodological individualism is the intellectual construction on which the political philosophy known as liberalism is founded. The central postulate is that we can draw a clear line between the self and the world: on the one hand, the human subjects, each one endowed with values, preferences, experiences, beliefs, and so on, and on the other hand, the material

resources of the world, natural and technological. Each individual person is an irreducible atom of the social body, and the public authority can no more alter individual values than the chemist can alter the inner structure of the atoms. Individuals act on the world by consuming or transforming its resources, and by interacting with other individuals (to influence their values and beliefs as well as their actions). The collective authority merely affects the appropriation of the resources of the world (how and by whom are these resources transformed, consumed?) and the interpersonal interactions (who can do what to whom?); it has no tools to influence individual values.

The most sophisticated model of methodological individualism is microeconomic analysis. The physical description of the world is captured by a set of *feasible states of the world* (each state specifies who consumes what resources, who works and for how long, etc.). The values of an individual agent (or citizen) are described by her *welfare,* measuring the degree to which a certain state of the world fulfills her values. And finally, the *rights* of an agent specify which actions are open to her (physically feasible and legally permitted), how these actions influence the state of the world, and in particular, how they interact with other agent's actions. Examples include private ownership (under which all voluntary trades are feasible) and political rights (specifying how each citizen can influence public policy).

Given the feasible states of the world and each citizen endowed with her welfare and her rights, the liberal social order results from the interaction of free wills. Each citizen pursues her own good by her own means; in other words, she chooses her actions (within the limits of her rights) in order to maximize her own welfare. When every citizen acts in this way, the eventual state of the world results from the balance of these individual powers and is called the equilibrium outcome. The only role of the collective authority is to enforce the free exercise of individual rights without influencing in any way the individual choices or the resulting state of the world.

There are two sides to distributive justice in the microeconomic world: Is the distribution of rights fair? Is the outcome of the game fair? Procedural justice addresses the former question, endstate justice addresses the latter.

The collective welfare approach focuses on endstate justice exclusively, by assuming that the decision power lies entirely in the hands of a judge, a benevolent dictator, who must propose a fair and principled compromise taking into account the conflicting interests and unequal characteristics of the participants. This is a convenient stylized description of the public authority, or of the decision-making process of an ad hoc committee. A full-fledged analysis encompassing the procedural and endstate aspects of distributive justice is currently out of our analytical reach, even for a question as basic as the design of a simple voting rule. In this book the discussion of procedural justice within the collective welfare model reduces to a few remarks about the incentives consequences of various collective utility functions in chapter 3, and the brief discussion of strategic voting in chapter 4

(see section 4.3). Procedural justice in fair division receives more attention; see chapter 6, in particular, section 6.4, and chapter 7, sections 7.1 to 7.3.

Thanks to the elimination of individual rights, the welfarist approach to distributive justice provides systematic methods to resolve the conflict between individual welfares. It does not endow the feasible states of the world with any normative content, regards them merely as means toward the satisfaction of individual values. This point of view distinguishes the welfarist methodology from the fair division problems of chapters 5, 6, and 7. There the production and allocation of physical resources play a key role in the definition of fairness.[2]

The concept of individual welfare can be modeled in two ways: as a *cardinal* utility function or an *ordinal* preference relation. These two models partition the welfarist approach in two branches of comparable importance.

Cardinal welfare is measured by a number, called the utility level, and the utility levels of different agents are comparable. The following statement is meaningful: "eating this cookie makes my utility increase (or decrease!) more than your utility would increase if you eat the cookie in question." In this context the interpretation of the sum of individual utilities as an index of collective fitness makes good sense. Imagine that we must allocate scarce medical resources among a given set of critically ill patients (e.g., that blood for transfusion is in short supply and must be rationed). If we can estimate, for each patient, the probability of recovery as a function of the amount of blood he receives, a sensible rule is to distribute the blood so as to maximize the sum of these probabilities, namely the expected total number of recoveries.[3] In other words, we take the probability of recovery as a proxy for utility, follow Bentham's *classical utilitarian* imperative, and so we maximize the sum total of individual utilities.

The more general tool of a *collective utility function* aggregates (by means of a mechanical formula) a profile of individual cardinal utilities into an index of collective utility. This operation is naively anthropomorphic if we try to interpret collective utility as a measure of the "welfare" of the group, or that of its benevolent dictator. However, as a mere technique for selecting a reasonable compromise, it is strikingly effective and easily applicable.

Collective utility functions are well suited to explore the trade-offs between the principles of compensation and fitness. Maximizing the classical utilitarian sum of individual utilities often results in "sacrificing" some individuals: if a majority of the roommates want to watch only sports on the only TV set, the minority who wishes to catch the news will never have its way under the utilitarian rule. Thus this particular collective utility function is orthogonal to the concern for compensation: in the TV example, the latter would require the roommates to tune in the news some of the time.

2. For instance, the Shapley value is computed directly from the cost or production function.

3. This is indeed a common approach to deal with the medical triage problem. See section 2.2.

By contrast, the *egalitarian* utility function is the embodiment of compensation in the welfarist model. This function is computed as the smallest individual utility; hence its maximization leads the benevolent dictator to implement a tiny utility increase for the worst-off member of the group, even if this implies a huge loss for many well-to-do agents. In the TV example the egalitarian judge awards equal time shares to the sport and news channels regardless of the relative support of the two programs, which is surely an unreasonable solution when the minority is small enough.

Chapter 3 describes a rich family of collective utility functions, ranging continuously from the classical utilitarian to the egalitarian one. The *Nash* collective utility function stands out as a remarkable member of this family: it computes the collective utility as the *product* of individual utilities. It strikes an appealing compromise between the two polar methods: in the TV example, it splits viewing time between the two channels in proportion to the number of their supporters.[4]

1.3 Collective Welfare: Ordinal

The core of the cardinal welfarist model is the feasibility of objective interpersonal comparisons of utility. This is not plausible when choices are governed by subjective tastes and values instead of an objective index related to health, nutrition, or any other primary human need. In the TV example, individual tastes for sports versus news are taken as given and radical: those who like sports never show any interest in news (and vice versa); any amount of time spent watching sports, however small, is better than any amount spent on news. Moreover taking "viewing time" as a proxy for utility amounts to ignore the variations in the intensity of feelings among the roommates, from the sports fanatic to the nearly indifferent. Measuring subjective feelings as an outside observer is, for all practical purposes, impossible. If we ask the agents themselves to help us in this assessment (by asking them to fill out a questionnaire), we can expect gross strategic manipulations, which is especially difficult in the TV example.

The objective measurement of utility is not only practically hard to impossible, it is also a logical contradiction of methodological individualism. A fully "objective" notion of welfare is alienated from the "subject" individual; hence it denies the pluralism of individual values, and defeats the purpose of methodological individualism. Paraphrasing Sen, classical utilitarianism views the individuals as petrol tanks among whom we must distribute the gasoline of the world.

4. See example 3.6.

There is a private, inviolable sphere around each of us that cannot be captured by an objective measurement, that cannot even be apprehended by an outside observer, where the self is defined independently of any social construction. In this sphere are produced some components of our personality of great relevance to social interactions—some of our value systems, such as our philosophical and political convictions, our religious beliefs, as well as our body, our psychological disposition to happiness—see next point—and much more. The point is that the public authority should not be allowed to control, or even monitor, these components of our own welfare: even the staunchest welfarist must concede this kernel of inviolable rights. The hope of an objective measurement of welfare is as unrealistic as that of comparing by a common index two different value systems, two outlooks on life. Bentham's sum of pleasures and pains is nonsensical (on stilts!) because my pain and your pain are incommensurable.

The last objection to the interpersonal comparison of welfares is a moral one. It is the strongest of the three.

An important element of personal responsibility is our ability to transform resources into welfare (our good or bad nature, our optimism or pessimism, etc.). Ignoring this important factor, the welfarist allocating resources in order to equalize welfare penalizes those agents who scale down their expectations to their lot.

The story of Tiny Tim, physically challenged and poor, yet fiercely happy is a case in point. Tim will be denied a larger share of resources on account of his good nature, of his ability to generate much welfare from an objectively poor endowment of personal characteristics. The blasé John, on the other hand, is a wealthy spoiled brat with natural talents that he does not enjoy using, and expensive tastes; he gets thoroughly depressed if he cannot drink champagne. Because his depression is real, does John deserve extra cash to buy an adequate supply of Dom Perignon? Similarly, in a political election, paying attention to the intensity of feelings over the various candidates would amount to giving more weight to the opinion of a fanatic than a cool-headed citizen.

Is there a future for welfarism in a world where many—most—factors relevant to welfare escape the eye of, or should not be assessed by, the benevolent dictator? The answer is yes, provided that we replace utilities by *choice*. The *ordinal welfare* of an agent is d*efined* as the set of his choices. If we extract from individual welfares only the choice component (i.e., the underlying ordinal preferences), then we will be unable, by construction, to make any interpersonal comparisons of the levels of welfare; our social decisions will reflect equitable compromises between conflicting preferences/opinions, without any invasion of the private spheres where actual welfare (enjoyment, happiness) is produced.

Under the rational choice postulate, the choices of an agent maximize a certain *preference relation* over the potential outcomes (states of the world). In other words, this relation is a ranking of all outcomes from best to worst (indifferences are permitted) and from any

feasible set of outcomes (the actual choice set); our agent picks the highest ones according to the ranking. An increase in his ordinal welfare means a switch to an outcome higher up on this relation, no more, no less. In particular, unlike a utility level, the notion of ordinal welfare is not logically separable from the physical states of the world under consideration.

All models of fair division in chapters 5, 6, and 7 are built upon ordinal preference relations. The common denominator of collective welfare in these models is the central notion of *Pareto optimality,* also called *efficiency*. An outcome x is Pareto inferior to another outcome y if every concerned agent views y as at least as desirable as x, and at least one of them views y as strictly more desirable than x. In other words, y stands at least as high as x in every agent's ranking (preference relation) and strictly higher for at least one agent. An outcome is *Pareto optimal* if it is Pareto inferior to no feasible outcome.

Pareto optimality is the single most important tool of normative economic analysis. Its desirability is undisputed. In the endstate version of distributive justice, it is the one requirement that cannot be dispensed with. Not so under procedural justice, where the informational context of the "procedure" may push the equilibrium outcome away from the Pareto frontier (the set of efficient outcomes).

Under ordinal preferences the welfarist program of anthropomorphic aggregation is alive and kicking. The tool is the aggregation of a profile of preference relations into a collective preference relation—just as in the cardinal world we aggregate a profile of individual utilities into a collective utility. The aggregation operator is called a *social welfare function,* and this concept is the object of chapter 4.

Because preference relations can only be defined within an explicit set of feasible outcomes, the aggregation method amounts to a decision process very much like voting, where the input (the ballot box) consists of a list of preference relations over the outcomes (candidates), and the output is to elect one particular outcome. The two objects, voting rules and social welfare functions, are so closely related that the two most important aggregation methods are directly adapted from two voting rules introduced more than two centuries ago by Condorcet and Borda.

Borda's proposal is to elicit from each voter a complete ranking of all candidates, and to record a score of 0 for the last ranked candidate, 1 for the next to last candidate, 2 for the next lowest one, and so forth. The total score awarded by all voters determines the winner. Condorcet's proposal elicits the same information but uses it to compute the majority relation: for any two candidates a, b it only records who wins the largest support when only these two are competing. The Condorcet winner is the candidate (if any) who wins all such pairwise contests.

The two rules above offer quite different interpretations of collective welfare. The majority relation compares any two candidates without taking into account their relative rankings with other candidates (a property known as *independence of irrelevant alternatives*). This

ensures that the procedure is immune to strategic misreporting of one's preferences whenever the majority relation has no cycle. However, cyclical majorities do appear for some configurations of individual preferences, and this pattern is unavoidable for any minimally fair aggregation method under independence of irrelevant alternatives (Arrow's theorem).

Borda's scoring rule, on the other hand, is robust to the decentralization of the decision process among subsets of voters. It delivers a social ranking of the candidate that is not plagued by any cycle.

1.4 Externalities and Fair Division

In the welfarist world, the physical resources of the world are fuel for welfare and are pointedly devoid of any ethical content. The allocation of vitamins among medical patients is methodologically identical to that of good wines or subsidies for college education among siblings, or cocaine among drug dealers. Voting to choose the president for millions of citizens is conceptually identical to the choice of the wallpaper among a handful of office mates.

In real life, the physical characteristics of the resources we consume strongly influence our definition of the "fair" production and distribution of these resources. The economic concept of *externalities,* and the related notions of private and public goods, provide important insights into this complicated interaction.

Risky choices are an example of externalities: if I am hurt in an accident because I do not wear my seat belt, my welfare loss is the greatest, but part of the cost of treating me is ultimately borne by my fellow citizens. Therefore my decision to buckle up or not affects more than my own welfare. Polluting activities, the exploitation of exhaustible resources (fish in the sea), create externalities too: the more I fish, the more costly it is for others to capture a certain amount of fish.

A commodity is a *private good* if its consumption by one agent generates no externalities on other agents. Thus a fruit is a private good, whereas a cigarette consumed in a public place is not. A commodity is a (pure) *public good* if it must be consumed identically by all agents in a given community (the economic terminology speaks of consumption without exclusion and without rivalry). Thus the legal system, the police, as well as a radio broadcast (when everybody owns a receiver) are pure public goods. On the other hand, a road system entails partial rivalry (congestion occurs when the number of users is large enough) and possibly exclusion (if only cars with odd plate numbers are allowed to move on a certain day). Fishing in the sea involves a private good as input (fishing effort) and a private good as output (the catch) but partial rivalry—as well as no exclusion—on the *return* of output to input.

A central tenet of economic analysis is that competitive markets work well for the exchange and production of private commodities but do not work well when the consumption or production of these goods entails externalities (e.g., when some of them are public goods). "Working well" means to achieve an efficient (Pareto optimal) utilization of the resources, through the decentralized selfish behavior of the agents, as captured by Adam Smith's invisible hand paradigm.

The microeconomic models developed in chapters 5, 6, and 7 illustrate these two dual propositions of economic wisdom, and show their relevance to the fair division problems. Chapters 5 and 6 focus on the management of a commons, the oldest and simplest story where production externalities hinder efficiency if access to the commons is not regulated. Chapter 7 considers the efficient distribution of pure private goods by the invisible hand. I review the latter first.

Start from a distribution of the property rights over the private goods where every market participant is "small," in the sense that nobody owns a large fraction of the total endowment in the economy; the distribution of property rights is otherwise arbitrary. Then the strategic equilibrium of voluntary, mutually advantageous trades is Pareto optimal and decentralized by a competitive price signal, namely a price for every traded commodity. Each agent receives the price list and requests the net trade (i.e., what quantity of each good he wants to buy or to sell), maximizing his preference ordering given the budget constraint imposed by the prices. The sum of all net trades cancels out; hence it is feasible to meet the demand of every participant in the market.

The price list is the same for every agent, and all competitive net trades are worth zero at these prices. Every agent can afford everyone else's net trade, and the net trade one chooses is the best for her preferences. Therefore she prefers her net trade to everyone else's, and this is the sense in which competitive trade is fair. Whether or not we agree that the initial distribution of property rights is fair, the move from this inefficient initial allocation to the competitive one is fair in the sense just described.

Now consider the problem of dividing a common property, namely a pile of private goods on which all agents have identical rights (think of siblings splitting an estate). Transforming common property into equal shares of private property is unquestionably fair. On the other hand, the allocation of equal shares of all goods to all participants is inefficient, if differences in individual preferences allow for mutually advantageous trades. But the competitive allocation resulting from the initial distribution of equal shares is surely efficient. Moreover it is fair in the sense of the no-envy test: no agent i would prefer to receive agents j''s competitive lot rather than her own competitive lot. This follows, as above, from the fact that everyone chooses optimally from a common budget set (equivalently they can choose the same net trades and the initial lot is the same).

Chapter 7 examines the fair division method just described called the *competitive equilibrium with equal incomes*. The endstate of this solution is just because of no envy. The solution is also the strategic equilibrium of a just procedure, namely the trading game from equal endowments.

We observe in chapter 7 the deep technical link between the competitive equilibrium with equal incomes and the optimum of the Nash collective utility function. We further compare it with another solution of straightforward egalitarian inspiration, called the *egalitarian equivalent* solution.

We turn to the complex issue of fair division under production externalities, where we encounter a fundamental contribution of normative microeconomics originating in cooperative game theory. In the very simple models of the commons in chapters 5 and 6, a given technology is the common property of a set of users, and the problem is to exploit the technology fairly and efficiently. The first difficulty comes from the fact that individual actions (input contributions or output demands) are not aggregated additively by the technology: the marginal return of input or the marginal cost of output is not constant. Therefore we cannot easily separate the impact of a particular user's actions. This issue is at the heart of the reward principle (section 1.1) and is the subject of chapter 5, where efficiency considerations are mostly absent.

The Shapley value is a systematic formula used to divide a joint cost or a jointly produced output. It offers a reasonable definition and computation of the share of cost or surplus for which a user of the commons is deemed responsible. A simple example shows why this question requires a genuinely new interpretation of the reward principle.

This four-story building has one apartment on each of the second, third, and fourth floors; the three apartments are otherwise identical. The manager of the building wishes to split fairly the cost of running an elevator to the three apartments. The cost of an elevator serving only the second floor is \$5,000. That of an elevator serving the second and third floors is \$10,000. An elevator serving all floors would cost \$40,000 because reaching the fourth floor requires structural changes to the building that could be avoided if the elevator shaft ends at the third floor. It has been decided to have the elevator serve all four floors.

The point of this example is that simple cost shares based on equality or proportionality contradict our intuition about individual responsibilities in total cost, as required by the reward principle.

Clearly apartment 4 (on the fourth floor) should be charged more than apartments 2 and 3, but how much more? Charging each apartment in proportion to the service (= number of floors transported) would give shares $1/6$, $2/6$, and $3/6$ respectively to apartments 2, 3, and 4: apartment 2 ends up paying \$6,667 and apartment 3 pays \$13,333, which is grossly unfair because each one of apartments 2 and 3 pays more than the full cost of an elevator stopping at its own floor.

A more reasonable set of shares would be based on the *stand-alone costs*. Here 5K, 10K, and 40K are the respective stand-alone costs of the three apartments. Dividing 40K in proportion to these costs yields:

$$\text{apartment 2:} \quad \frac{5}{5+10+40}\,40{,}000 = \$6{,}636$$

$$\text{apartment 3:} \quad \frac{10}{5+10+40}\,40{,}000 = \$7{,}273$$

$$\text{apartment 4:} \quad \frac{40}{5+10+40}\,40{,}000 = \$29{,}091$$

Apartments 2 and 3 are still subsidizing unfairly apartment 4 because their joint share is \$909 higher than the cost of an elevator stopping at the third floor!

The Shapley value in this example simply splits the cost of each segment of the elevator (from floor 1 to 2, 2 to 3, and 3 to 4) among the users of that segment. Thus the first segment is split three ways, the second is split equally between apartments 3 and 4, the third is charged to apartment 4 alone:

$$\text{apartment 2:} \quad \frac{1}{3}\,5{,}000 = \$1{,}667$$

$$\text{apartment 3:} \quad \frac{1}{3}\,5{,}000 + \frac{1}{2}\,5{,}000 = \$4{,}167$$

$$\text{apartment 4:} \quad \frac{1}{3}\,5{,}000 + \frac{1}{2}\,5{,}000 + 30{,}000 = \$34{,}167$$

In general, the Shapley value assigned to each agent is her expected marginal cost (or surplus) when agents are randomly ordered with uniform probability on all orderings. This formula has strong normative foundations, expressed in a handful of different axiomatizations discovered after Shapley's initial characterization half a century ago. Two of these axiomatizations are discussed in chapter 5.

The applications of the Shapley value are remarkably versatile, as the examples in chapters 5 and 6 demonstrate abundantly. It is the most important contribution of game theory to distributive justice.

In chapter 6 the fairness (reward) and efficiency issues are addressed jointly in a simple model of the commons with variable returns, where all agents may consume at most one indivisible unit of identical output (service), and they only differ with respect to their willingness to pay for the service. Free access to the commons means that each agent decides to "buy" service or not and total cost is equally divided between all buyers. The strategic equilibrium of this natural procedure may, however, be severely inefficient: this difficulty is known as the *tragedy of the commons*.

Because individual preferences (willingness to pay) vary, efficient use of the commons commands to serve the agents with the largest willingness to pay, up to the "marginal" agent whose willingness to pay barely exceeds the marginal cost of production. Other agents are not served, and must be fairly compensated. We compare three competing interpretations of fairness.

The first one is the no-envy test, whereby the "efficient" agents pay the marginal cost of service and the balance (revenue collected at that price minus total cost of production) is divided equally among all agents, efficient and inefficient alike. The corresponding solution is the competitive equilibrium with equal incomes discussed above.

The second solution is defined by the property that any agent's share of surplus (for an efficient agent: willingness to pay for service minus his payment; for an inefficient agent: a positive or negative cash transfer) is the same *as if* everyone was offered the service at a common *virtual price*.

The third solution considers the joint surplus generated by any subset (coalition) of agents *standing alone* to use the technology, and applies the *Shapley value* of this cooperative game.

Each one of these three solutions is subtle in its own way, and proposes a different interpretation of what common property of a technology could mean when there are production externalities (variable returns). We also study the implementation of the three solutions by appropriate procedures.

1.5 Private versus Public Contracts

Libertarian political philosophy offers a simple solution to deal with the externalities just discussed. Assume that the property rights over the existing resources—consumption goods as well as means of production—are clearly defined and entirely distributed among the economic agents. The libertarian postulate (formulated most clearly by Coase) predicts that the interested parties will spontaneously enter private contracts to reallocate the property rights to their mutual advantage.

Say that Ann lives upstream and has the right to pollute the river, thus harming Bob who lives downstream. The opportunity for private recontracting hinges upon the comparison of the cost x to Bob of Ann's polluting activities and the cost y to Ann of refraining from polluting. If x is larger than y, then an offer by Bob to buy Ann's pollution rights at a price between y and x is attractive to both parties, and the postulate says that they will somehow agree on a price—that is, reach an agreement on the division of the surplus $(x - y)$. If, on the other hand, y is larger than x, then it is efficient that Ann continues to pollute, and the status quo will prevail because Bob cannot make an offer that would appeal to both parties. Therefore an efficient utilization of the resources will result regardless of the initial distribution of property rights.

A *private contract* is a mutually advantageous deal: every party is better off when the contract is implemented, therefore signing the contract is the consequence of rational behavior. The contract is explained from no other principle than the pursuit of one's own selfish welfare by all the parties: no social value judgment enters the description of this transaction, and no coercive device is needed to enforce compliance with the terms of the agreement. In other words, the contract is free of value: either party can revoke it if it is advantageous to do so, and the price they agree upon, for instance, is not grounded on any fairness principle: it is simply the equilibrium result of selfish bargaining.

Public contracts are designed by a party of disinterested rational actors, under the veil of ignorance, in the name of compelling arguments of justice—in practice, by a small number of experts, judges, or "founding fathers" who agree on the best organization of a certain institution of concern to all members of society. As explained below, the allocation methods and collective decision rules studied in this book are all examples of public contracts expressed in the microeconomic language and yet broadly applicable. They are miniature social contracts.

Public contracts require a normative justification, must be explained from first principles of rationality and justice, and administered by a central agency, whereas private contracts have no ethical content. Private contracts result from the balance of power between the agents qua players, and from a decentralized, spontaneous process of interaction that requires no social evaluation.

The libertarian credo dismisses public contracts as unwarranted and detrimental to efficiency, as they place constraints on private initiatives to invent beneficial private contracts. In the libertarian world the only interference of the public authority in the lives of private citizens is to enforce their private property rights, which is a "minimal" definition of the role of the state.

Thus the kind of principled discussion of distributive justice that is the object of this book has no place in the libertarian world. The normative/axiomatic approach is built upon a more balanced interpretation of the liberal political philosophy that recognizes the need for public contracts alongside private ones.

The Achilles heel of the libertarian view is to ignore the huge costs, practical and psychological, of direct, value-free bargaining. Exclusive reliance on private contracting to resolve countless externalities and conflicts of opinion—generated by the apparently limitless division of labor in industrialized societies—is plainly utopian.

The transaction cost of reaching an agreement rises more than linearly with the number of parties involved, in any measure of the amount of interpersonal communications implied by collective negotiation. Think of the unanimous consent required in jury decisions: a single stubborn juror can "block" the entire process. The European union will grow soon to more than two dozen members, and must consequently drop its rule of unanimous consent on most

issues. When we choose a rule of universal concern (a bill), we must rely on voting rules to achieve a compromise between conflicting opinions. Passing bills by majority voting is a prima facie liberal institution, yet it is not equivalent, by any stretch of the imagination, to a private contract signed by all citizens: it forces coercively an outcome that certain citizens find detestable.

Thus a contract binding a large community requires coercive intervention of the public authority. The latter is acceptable to all (most) members of the community only if it is justified by a "reason," a general anonymous principle. This is precisely what public contracts are about.

Even transactions on a very small scale are often resolved more effectively by public contracts than by private ones. Consider the division of assets during a divorce. Many couples find it emotionally difficult to reach a reasonable settlement without the help of some guidelines on the division process, such as, provision for child support. These guidelines (whether or not enforceable in a court of law) embody the impersonal principles of fairness of a public contract. The situation is similar in most fair division issues involving a specific type of transaction: sharecropping, profit sharing between an artist and her agent, a publisher and its authors, bankruptcy settlements. A customary division rule is a focal principle of justice that all parties can easily adopt; whether or not this principle is legally enforceable influences how often the parties comply with its recommendation, but the principle remains an instance of a public contract between a very small number of parties.

1.6 Organization and Overview of the Book

In this book I propose a handful of fundamental public contracts, formulated in microeconomic language, as methods for allocating certain resources or rules for reaching a compromise between conflicting opinions. These rules and methods include the equal sacrifice taxation schemes (chapter 2), collective utility functions (chapter 3), voting by majority à la Condorcet or by scoring à la Borda (chapter 4), the Shapley value (chapters 5 and 6), the competitive equilibrium with equal incomes (chapters 6 and 7). They are simple tools for social engineering, and their normative justification is provided by a series of *axioms*.

A full-fledged discussion of the axiomatizations relevant to the rules and methods just listed is beyond the scope of this introductory text. However, the book describes informally some of the main axioms and the corresponding key axiomatic characterization results or impossibility theorems. Specifically, section 3.2 states the central result of cardinal welfarism, namely the characterization of additively separable collective utility functions. Section 4.6 describes Arrow's impossibility theorem, the seminal result of ordinal welfarism, and its key axiom of independence of irrelevant alternatives. The two most important axiomatizations

of the Shapley value are the subject of section 5.5. An axiomatic comparison of the three efficient methods for managing the commons is provided in section 6.6. Section 7.6 compares similarly our two methods for the fair division of heterogeneous commodities, the competitive equilibrium with equal incomes and the egalitarian equivalent solution.

Only two axioms qualify as universal within the confines of this book: they are the basic symmetry property equal treatment of equals, and efficiency/Pareto optimality. All other axioms are used to inform the differences among competing allocation methods. Examples include choices among utilitarian, egalitarian, and Nash collective utility functions in chapter 3, between Condorcet and Borda voting methods in chapter 4, or between the competitive equilibrium with equal incomes and the egalitarian equivalent solution in chapters 6 and 7.

The book is self-contained, yet a familiarity with microeconomic thinking and/or mathematics modeling will make the reader's task easier. It is best suited for advanced undergraduates, in particular, those who have been exposed to intermediate microeconomics, or first-year master or graduate students.

Each chapter, starting with chapter 2, is organized as a sequence of intuitive examples, intertwined with more general discussions in which the mathematical arguments, if any, are elementary. It contains one technically challenging section evoking the relevant axiomatic results (as explained above). Each chapter offers about a dozen detailed exercises, some of them simple numerical examples, others developing some formal properties alluded to earlier in the chapter. It ends with a short introduction to the relevant literature.

The sequence of chapters corresponds to a partly subjective list of the most influential normative models of resource allocation.

In chapter 2 the resources in question are just a sum of money or of any other desirable homogeneous commodity, and the relevant individual characteristics are a single number. Examples of application include rationing an overdemanded commodity, designing a tax schedule, or sharing the benefit from a joint investment. The three basic division methods are simple proportionality of shares to claims, equalization of shares, and equalization of losses (= claims − shares). Depending on the normative interpretation of the resources and of individual characteristics, these methods translate the compensation or the reward principle into an operational formula, exactly as Aristotle had suggested in the first place.

The cardinal welfarist model is the subject of chapter 3. The physical resources generate individual welfare measured by a cardinal index of utility, but the physical allocation process is devoid of normative content. Consequently all relevant information is contained in the utility profiles (specifying the cardinal utility of each concerned agent) achievable by a feasible allocation of the underlying (unspecified) resources. A collective utility function provides an interpretation of the compensation principle compatible with efficiency fitness (Pareto optimality; see section 1.3). Three such functions stand out of the axiomatic

discussion, namely the *classical utilitarian* sum of individual utilities, the *egalitarian* minimum of individual utilities, and the *Nash* product of individual utilities.

Chapter 4 is devoted to the ordinal welfarist model, in which individual welfare is described by a preference relation and collective welfare is computed by an aggregation method called a social welfare function. As explained in section 1.3, the two leading aggregation methods are the majority relation proposed by Condorcet, and the scoring method advocated by Borda. The celebrated impossibility result due to Arrow states that no social welfare function can yield rational collective preferences at every profile of individual preferences (as Borda's method does but Condorcet's does not) and base the collective opinion between two outcomes solely on the profile of individual opinions between these two outcomes (as Condorcet's, but not Borda's, method does). The two general principles at play in chapter 4 are fitness and (equal) exogeneous rights.

The microeconomic problems of fair division discussed in chapters 5, 6, and 7 have been already presented in section 1.4. Recall that the Shapley value, defined in chapter 5, is a mathematical formula cutting the Gordian knot of widespread production externalities, and that it is applicable to a wide array of fair division problems. Its focus is on the reward principle, when a given production process intertwines the inputs of different agents in a complex interaction from which each individual contribution cannot be easily separated (e.g., the impact of the various inputs is not additive). The question is to give every participant in the production effort his or her fair share of the resulting output; alternatively each participant demands a different output share, and we seek to divide fairly the total cost of production.

In chapter 5 we also introduce the stand-alone test, another equity property for the allocation of joint costs or joint surplus. When the test is applied to individual participants, it is never violated by the Shapley value, but in its stronger form, known as the stand-alone core property, it may rule out the Shapley value altogether.

The model of chapter 6 is a simple and familiar example of production externalities often called the problem of the commons. The three solutions defined and compared there are reviewed at the end of section 1.4. They are three different interpretations of the reward principle when a common property technology is used efficiently. Note that chapter 6 stands out in this book for its more demanding technical level, as well as for the originality of the material discussed there. The less technical reader may choose to skip it.

Chapter 7 starts with a brief review of competitive trade under private ownership, and of its relation with the core of the cooperative game of free trade. We show, in particular, that the competitive trade may break down when production exhibits nondecreasing returns to scale, and the same applies to the core stability of free trade.

The second half of chapter 7 deals with the fair and efficient division of a "pie" in the common ownership regime, where a pie means a list of divisible private commodities.

Common ownership, here as in the production model of chapter 6, is no more and no less than the vague principle of equal property rights. An *envy-free* division of the pie is a resting point of the interpersonal comparison of individual shares (as opposed to individual welfares) in the following sense: no one strictly prefers the share received by another participant to her own share. The combination of no envy and of efficiency-fitness (Pareto optimality) leads to the fair division method called competitive equilibrium with equal incomes (section 1.4). To find this solution, we must discover a list of prices, one for each commodity, such that when each participant spends a fair share of the total worth of the pie at those prices, the aggregate demand equals precisely the contents of the pie.

An alternative solution is the egalitarian-equivalent division method, whereby each participant receives a share that he/she views as equivalent to a common fraction of the pie. The numerous examples and the axiomatic discussion in section 7.6 reveal that our two methods (competitive with equal incomes and egalitarian equivalent) are two versatile and plausible normative interpretations of fair division.

The brief chapter 8 provides, for each of chapters 2 to 7, a formal, mathematical definition of the relevant concepts and a precise statement (without proof) of the results alluded to in the course of the successive chapters. Thus chapters 8 serves as a glossary of the technical material underpinning the less formal discussion in the successive examples and exercises.

1.7 Introduction to the Literature

The material covered in this book has deep roots in the economic and political philosophy literature.

The central concept of collective welfare, in its cardinal or ordinal form, is a famously general and far-reaching intellectual construction more than two centuries old. In its cardinal form, it provides the backbone of the political philosophy known as utilitarianism, starting with Bentham and John Stuart Mill; it also delivers practical tools for conflict resolution such as the Nash solution to the bargaining problem. In the ordinal form, the aggregation of individual preferences is an abstract model of democratic elections formalizing the political concept of the "will of the majority."

The systematic discussion of collective welfare, and the contrast between the cardinal and ordinal models (sections 1.2 and 1.3), was the central theme of the "new welfare economics," summarized in the classic Samuelson (1948) book. The seminal axiomatic contribution bear on classical utilitarianism (Harsanyi 1955) and on the aggregation of ordinal preferences (Arrow 1951). The analysis of cardinal collective utilities and ordinal social choice functions was developed mainly throughout the 1960s and 1970s. The organizing principle is the informational content of welfare, and the related axioms of measurement invariance,

discussed in sections 3.2 to 3.4. The classic reference is Sen (1971); more recent surveys include Moulin (1988, chs 1 and 2) and Bossert and Weymark (1996).

The theme of externalities is as old as economic analysis, but the modern concepts discussed in section 1.4 are approximately fifty years old: Shapley invented the "value" in 1953 and the concept of no envy is generally credited to Foley (1967). Further references are given at the end of chapters 5, 6, and 7.

The concepts of private and public contracts sketched in section 1.5 allude to the central debate of contemporary political philosophy, pitting the liberalism against the social contract traditions. The libertarian position is spelled out most clearly by Hayek (1976), Buchanan and Tullock (1962), and Nozick (1974). In my view, this position inspires the extremely influential formal work on repeated games (Aumann 1987) and the spontaneous evolution of cooperation (Axelrod 1984; Binmore 1994). This view is developed in Moulin (2001b).

Rawls's 1971 book, *A Theory of Justice,* inspired a striking revival of the social contract tradition, and in particular of its egalitarian variant. Kolm's 1972 book, *Justice et equité,* offers the first formal presentation of the egalitarian collective utility. The egalitarian position was subsequently articulated both as a philosophical statement, Dworkin (1981), Sen (1985), and Cohen (1995), and as an axiomatic model of economic theory: Roemer (1996, 1999) and Fleurbaey (1996).