3

Efficiency

OVERVIEW

We have thus far focused on individual rationality: what is it to be a rational actor, and how can we formally model such an actor? This chapter begins with that concern, but then turns to an analysis of how rational actors interact—the subject of the remainder of the book. The main concern of this chapter is to explore the relation between rational action and the idea of efficiency: the chapter starts with efficiency of the consumption decisions of one person, and then moves to the idea of an efficient exchange: an interaction between two economically rational agents. The important ideas of *Pareto superiority* and *Pareto optimality* are introduced. The second part of the chapter briefly sketches well-known failures of efficiency, involving various notions of "externalities."

3.1 RATIONALITY AND EFFICIENCY

Everyone knows that economics is about efficiency, and most of us have some pro or con attitude about that. Some of us are all for efficiency while others insist that efficiency is a cold economic value that must not come before equity, concern for the needy, or protection of the environment. Thus it is said we need to "trade off" efficiency against other values.[1] But while most of us know whether we are "for it or agin' it," we are often not sure what "it" is. Just what

is efficiency? Is it simply one value among others—one that economists but not the rest of us find very attractive—or it is somehow a fundamental idea that we cannot do without, which we ought not to sacrifice for other good things?

Efficiency and Rational Individual Choice

Let us reflect on what we know about our rational "economic man." As a fully rational chooser *Homo Economicus* has a well-formed utility function satisfying the requirements of utility theory we examined in Sections 2.2 and 2.3. His preferences are characterized by decreasing marginal utility and downward sloping demand curves (recall here the idea of decreasing rates of marginal substitution of goods in Section 1.3). Suppose, then, that *Homo Economicus* has a preference for pizza. How much pizza is it rational for *Homo Economicus* to consume? Suppose that the cost of pizza is constant at $\mu(c)$: we must remember that by "cost" we mean the total *opportunity costs* of consuming the pizza—the forgone opportunities to satisfy other preferences (see Section 1.3, point 3). It is important to stress that "cost" does not necessarily mean a monetary payment, or something that you don't like (as in "a cost of taking this course is that I have to take tests"). In the economist's sense, the "cost" of getting your first choice of a pizza is that you had to forgo your second choice of a box of chicken wings: when you have to choose between good things, the cost of your decision is that good thing you didn't choose. This can be specified in terms of *Homo Economicus*'s forgone utility—call this $\mu(c)$. It is the utility you would have received from your second choice. Now consider Alf's decision to consume the first slice of pizza. It satisfies his preference for pizza; given his utility function we can represent this by some utility benefit—call it $\mu(b)$. If $\mu(b) \geq \mu(c)$ (that is, if the utility benefits are greater than, or equal to, the utility costs), then it will be rational for him to purchase the slice of pizza. But should he buy only one slice? Well, we know that since the cost of pizza is constant per unit, the cost of the second piece will be $2[\mu(c)]$. But because of decreasing marginal utility, the benefits of the second piece will be less than twice $\mu(b)$: call this $\mu(b) + [\mu(b) - n]$ (where n is a positive number less than $\mu(b)$). The crucial idea here is that whatever the utility benefits of the first piece $[\mu(b)]$, the utility of the second piece will be positive (because more is better than less) but smaller than the utility of the first piece (hence it will be $\mu(b) + [\mu(b) - n]$): that is what is meant by decreasing marginal utility.

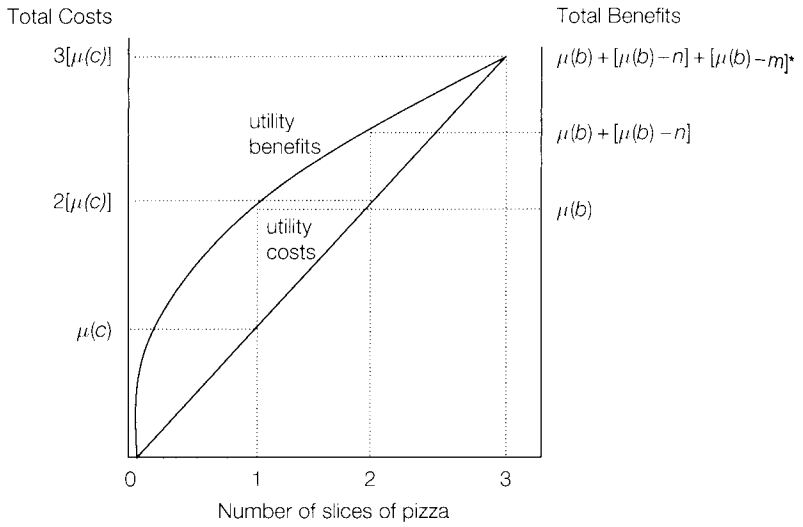

FIGURE 3-1 Efficient Consumption
*where $\mu(b) > m > n$

It will be rational for Alf to consume two slices if the utility benefits of two slices is greater than, or equal to, the utility costs of two slices—$\mu(b) + [\mu(b) - n] \geq 2[\mu(c)]$. Because the costs are constant but the marginal benefits are decreasing, at some point it will be the case that the additional, or as economists say, "marginal" (utility) benefits that Alf gets from some slice of pizza will be less than the marginal (utility) costs he had to incur in order to get that slice; in that case it would be irrational for *Homo Economicus* to consume that additional slice of pizza because the preferences he then would be satisfying are ranked below the preferences he is forgoing. Figure 3-1 is a graphic representation of a specific example of this simple choice problem. On the left vertical axis we measure *Homo Economicus*'s total utility costs while on the right vertical axis we measure his total utility benefits. In Figure 3-1 it is rational for *Homo Economicus* to purchase three slices of pizza (but no more).

Homo Economicus will consume up to the point where *marginal benefits equal marginal costs*. This is one definition of efficiency. *Homo Economicus* chooses in this way not because he "values efficiency" (any more than a rational actor seeks to "maximize utility" [Section 2.4])—this decision is simply required by rationality. To choose any other

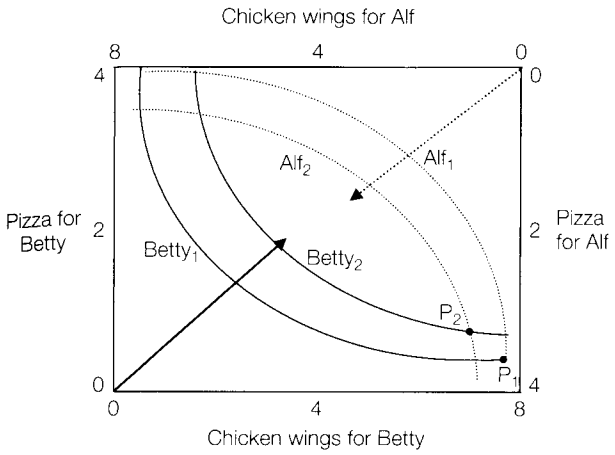

FIGURE 3-2 A Simple Edgeworth Box

way would be to choose a lower- over a higher-ranked preference, and that, of course, is the essence of irrationality. In this sense, then, efficiency is simply rationality.

There is, then, a close relation between efficiency and rational choice: to be rational is just to choose in a way that best satisfies one's preferences—and that means that the marginal gains are at least as great as the marginal costs. We shall examine paradoxical situations in Chapter 4 where rationality and efficiency break apart; in many of these cases we are uncertain just what is the truly rational thing to do.

Efficiency and Rational Exchange

Let us move to a two-person case. We will represent the choice problem in terms of indifference curves, which, it will be recalled, plot one's preferences over bundles of goods given decreasing marginal rates of substitution between them (Section 1.3, point 2). Suppose that Alf has four slices of pizza and Betty has eight Buffalo chicken wings. And suppose that they have pretty much the same preferences over pizza and chicken wings, so that their indifference curves are essentially the same. Figure 3-2 gives their indifference maps; Betty's (solid) indifference curves start in the lower left, Alf's (dotted) indifference curves come down from the upper right. Suppose that in this "Edgeworth Box"[2] Alf and Betty are at point P_1: Alf has four slices of pizza and no wings while Betty has eight chicken wings and no pizza.

It is important to remember that Alf is indifferent between all combinations of pizza and chicken wings on the indifference curve Alf_1; however he prefers all bundles on curve Alf_2 to all bundles on Alf_1. The dotted arrow coming down from the top right corner shows that as Alf moves "southwest" he receives increasingly preferred combinations of pizza and chicken wings. If he moves all the way to the "southwest" corner, he receives all the pizza and all the chicken wings. The solid arrow represents the direction that Betty prefers to move: starting at the lower left corner, she prefers combinations of chicken wings and pizza that are to the "northeast." So Betty is indifferent between all bundles of pizza and chicken wings on the indifference curve $Betty_1$, but she prefers all the bundles on $Betty_2$ to any of the bundles on $Betty_1$. (And, of course, she is indifferent between the bundles on $Betty_2$.)

Now any point in the "eye" formed by the Alf_1-$Betty_1$ indifference curves improves the utility of both Alf and Betty over point P_1: any point in the "eye" moves each to a higher indifference curve. Pick any point in the "eye": you will see that Betty and Alf have both moved from P_1 in their preferred directions. What this shows is that both can be made better off by exchange, *even though no additional goods have been created*. An exchange that moved both Alf and Betty to point P_2 on Figure 3-2 would make both of them better off; both are raised to a higher indifference curve: note that P_2 is on Alf_2 and $Betty_2$. We can say then that point P_2 on Figure 3-2 is *Pareto-superior* to point P_1: at least one person is better off and no one is worse off. In this case, both people are better off.

But although the bargain at P_2 is Pareto-superior to P_1, there are still Pareto-superior bargains that Alf and Betty can make starting at P_2. The gains from trade have not been exhausted: so long as Pareto-superior moves are available, Alf and Betty can keep on trading and at least one will benefit. When are the possible gains from trade exhausted? We can easily see from Figure 3-3 that when they reach a bargain at which their indifference curves are tangent, Pareto-superior moves are exhausted.

Consider, for example, point P_x. At point P_x Alf can only move to a higher indifference curve if Betty moves to a lower indifference curve. But, by definition, that would make her worse off, and a move is only Pareto-superior if no one is made worse off. So, too, starting at point P_x, the only way that Betty can rise to a higher indifference curve is if Alf moves to a lower one—that is, he is worse off. Point P_x is thus a *Pareto-optimal* bargain: any departure from point P_x would

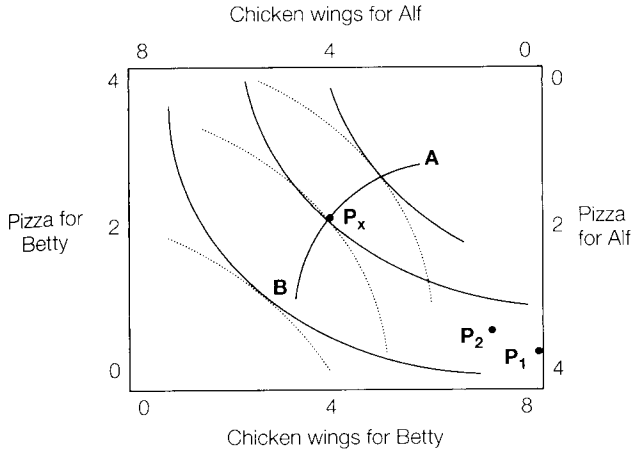

FIGURE 3-3 A Contract Curve

make either Alf or Betty worse off. Point P_x, therefore, is efficient: all the gains from possible exchanges have been exhausted. But point P_x is not unique in this regard: any bargain that occupies a point on which Alf and Betty's indifference curves are tangent is Pareto-optimal. The line A-B (which is called the *contract curve*) represents all such possible efficient bargains—ones that use all the possible gains from trade. Obviously, starting from point P_1, some of the efficient contracts favor Betty while others are better for Alf.[3]

Notice two things. *First*, the assumption of decreasing rates of substitution is crucial in explaining why economically rational people trade with each other. Because they each prefer varied to uniform bundles of goods, if Alf is pizza-rich he will want to exchange with Betty, who is chicken-wing-rich. This is a fundamental point that merits emphasis: *given decreasing rates of marginal substitution (or decreasing marginal utility), everyone can become better off through market exchange without any increase in the total number of goods. Second*, taking part in such trades is simply another example of marginal costs equaling marginal benefits. We have essentially the same story in a different form: here, rather than employing cardinal utility, we are analyzing the problem in terms of ordinal utility (i.e., preferences over bundles of goods). The marginal costs of Alf keeping those last two slices of pizza (his opportunity costs of forgoing acquiring four chicken wings) are greater than the marginal benefits that he gets from those two

slices of pizza. Thus, as I have depicted the problem, it is irrational for Alf and Betty to refrain from trading: if they keep their bundles at P_1 they are satisfying lower- over higher-ranked preferences.

We are now in a position to understand the concept of Pareto, or allocative, efficiency.[4] We can say that distribution D_2 is *Pareto-superior* to (more efficient than) D_1 if and only if no person is on a lower indifference curve in D_2 than that person is in D_1, and at least one person is on a higher indifference curve in D_2 than she is in D_1. If despite the possibility of a move to a Pareto-superior distribution we stay in D_1 there is at least one person who could achieve a higher level of preference satisfaction without lowering anyone else's. Thus in Figure 3-3, the distribution identified by P_x is Pareto-superior to the distribution of P_1. As in all the cases we have discussed thus far (but see Section 4.2) there is something irrational about maintaining Pareto-inferior distributions. In addition to being Pareto-superior to both P_1 and P_2, point P_x is also *Pareto-optimal* just because there is no alternative distribution which is Pareto-superior to it. That is, if Alf and Betty have arrived at P_x, there is no way in which one of them can be raised to a higher indifference curve without the other moving to a lower curve.

Is the Pareto Criterion a Moral Ideal?

Paretian Welfarism The Pareto criterion is often understood not simply as a requirement of *rationality* qua efficiency, but as a standard by which we can judge the *moral* desirability of a distribution or, in general, a social state.[5] To many it seems clear that distribution D_2 is morally better than D_1 if (and only if) some person's welfare is "higher" in D_2 and no one's is lower than it was in D_1. Especially in politics, it is thought, what is good for people—their welfare—must be the (sole) criterion of a good policy. This view has clear roots in utilitarian moral theory, which identified promoting human happiness as the sole goal of morality and politics. Recall the remark quoted in Section 1.1 from Nassau William Senior, a leading political economist of the nineteenth century; economics, he said, could assume that everyone seeks wealth because "wealth and happiness are very seldom opposed." The ultimate aim was clearly human happiness even if the proximate aim of economics was the growth of wealth. Now contemporary *welfare economics* typically understands a person's "welfare" to be measured by her utility function. If the utility of Betty is μ in distribution D_1 and $\mu+n$ in D_2, then it is said

her welfare is higher in D_2 than in D_1, and D_2 is a better distribution than D_1. Here, however, things get complicated. The early utilitarians such as Jeremy Bentham and his followers believed that "utility" (pleasure) was a cardinal measure (it could be measured along a metric) and that, when contemplating a move from D_1 to D_2 we could sensibly add the utility Alf received from the move from D_1 to D_2, to the utility Betty received, and then subtract the loss of utility to Charlie (who, let us say, was better off in D_1). Having done our sums, we could then decide whether, overall, the move from D_1 to D_2 increases overall aggregate utility. But we have seen (Section 2.3) that there is no particularly good reason to add von Neumann–Morgenstern cardinal utilities of different people; unless there is a special case for some additive function, it is simply arbitrary to sum up cardinal utilities. The contemporary welfarist seems to have a problem: how to compare social states without interpersonal comparisons of utilities?

The Pareto criterion seems to offer a way out of this problem: if no one is worse off in D_2 than she was in D_1, and at least one person is better off in D_2 than he was in D_1, then D_2 is Pareto-superior to D_1. And since the welfare economist has identified a person's welfare with her utility, it looks as if we can say that D_2 does better from the perspective of human welfare. Now it is often thought that this cannot be a very useful criterion of "moral betterness": it only yields a judgment that D_2 is better than D_1 if *no one* is worse off in D_2. But how often is it the case that no one is made worse off? On the Pareto test, if in D_2 one million people are made better off than they were in D_1, but one person is worse off, we cannot say that D_2 is Pareto-superior to D_1. Is there ever, we might well wonder, a Pareto-superior move to be made? We are now in a position to see the economist's deep attraction to market transactions. Under certain idealizing conditions (e.g., full information, no third-party effects), each market transaction moves us to a Pareto-superior distribution. When people trade, they prefer what the other person has to what they offer to give up, and so we move to a Pareto-superior distribution. As long as we have not exhausted the possibilities for exchange—as long as there are trades that people want to make—we have not exhausted the possibilities for Pareto-superior moves.

Although market transactions are often moves to Pareto-superior outcomes, it is much harder to see how a collective public policy can meet the Pareto test. It is hard to think of any uniform policy that does not disadvantage someone.[6] To avoid this conclusion (i.e., that

the Pareto criterion must be violated by uniform public policy) some welfare economists and "Paretian" political philosophers have adopted what is known as the Kaldor-Hicks criterion: *the move from distribution D_1 to D_2 is efficient even when some lose by moving from D_1 to D_2 so as long as those who do gain from the move* could *compensate the losers out of their gains.*[7] To grasp what it means to say that a person *could* be compensated for a loss, consider Alf, who, we are supposing, is the sole person who has been made worse off by the move from D_1 to D_2 (to make the case simple, assume that everyone else is better off in D_2). To say that Alf has been made worse off means that he is on a lower indifference curve in D_2 than he was in D_1. Now imagine that after the move to D_2 the gainers transferred enough of their gains to Alf to raise him back to the indifference curve that he occupied in D_1: this would bring about a new distribution D_3, which is indeed Pareto-superior to D_1 because everyone except Alf is at a higher indifference curve in D_3 than they were in D_1, and Alf is now back on the same indifference curve (as he was in D_1). We can say, then, that D_2 is Kaldor-Hicks Pareto-superior to D_1 if there is a distribution D_3 that (1) could be produced by redistributing the gains made by moving from D_1 to D_2 and (2) D_3 is (in the normal sense) Pareto-superior to D_1. Note the Kaldor-Hicks test says that given (1) and (2) D_2 is Pareto-superior to D_1 *even though no actual compensation has been paid*. Distribution D_3 is that in which compensation actually has been made, but Kaldor-Hicks does not say simply that D_3 is Pareto-superior to D_1; it says that D_2 is Pareto-superior to (more efficient than) D_1 even though some people have actually incurred losses by the move from D_1 to D_2. Because D_2 *could* give rise to D_3, and because D_3 *would* be Pareto-superior to D_1, D_2 is Kaldor-Hicks Pareto-superior to D_1.

To many this seems very odd: the Pareto criterion, which was based on the denial that gains for some can outweigh losses for others, is now employed to justify policies that benefit some at the expense of others. The move from D_1 to D_2 makes some people worse off, yet it is justified as a Pareto-superior move! Kaldor-Hicks looks like a backdoor way of getting interpersonal comparisons of utility loss and gains within a Paretian framework.

Even if we put aside the controversial Kaldor-Hicks interpretation of the Pareto criterion, upon reflection the Pareto test is not as uncontroversial as is often thought.[8] Much of the appeal of the Pareto criterion lies in the question "Who could possibly object to an improvement that makes everyone better off?" Figure 3-4 suggests an answer.

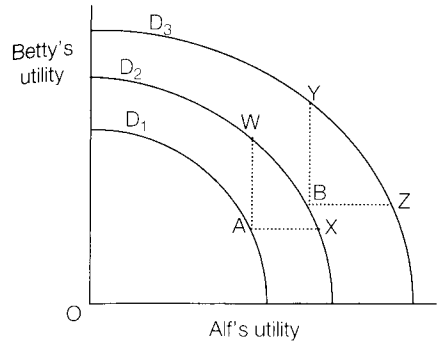

FIGURE 3-4 A Series of Paretian Moves

Suppose we start at point O, and make the Pareto-superior move to point A, which is one possible distribution along D_1—the "Pareto Frontier," the set of possible Pareto improvements from O. Once we are at A, the Pareto improvements to D_2 are limited to those between points W and X; other points on D_2 make either Alf or Betty worse off than in D_1, so are not Pareto improvements. Assume then that at some point distribution D_2 becomes a possibility: Alf and Betty make the Pareto move to point B on D_2. Suppose now that in the future D_3 becomes possible; now the possible Pareto improvements are limited to those falling between Y and Z. We can see that successive applications of the Pareto criterion move distributions along a path that is increasingly beneficial to Alf and of less benefit to Betty. If we had been able to jump to D_3 all at once, everything on it would be a Pareto improvement over O, but once we have made the intermediate moves to A and B, most of D_3 is excluded by the Pareto criterion. Perhaps, then, Betty would have good reason to object to the initial Pareto move to point A. The Pareto principle allows a wide range of moves, and it may matter a lot which of those is actually made, and in what order.

Welfare and Preferences Leaving aside these problems with the Pareto criterion, it also seems doubtful that we should accept the identification of "preference satisfaction" with "welfare." We have seen that preferences need not be about one's own good or self-interest: anytime one ranks an outcome above another, one has a preference (Section 2.1). Recall our last-mango refuser from Section 2.3; her preference is not to take the last mango, though, as Sen notes, she would like that mango and would welcome someone thrusting

the mango on her. In that case it seems doubtful that we should say that her welfare is enhanced by satisfying her preference not to take the last mango, since her "civility" preference instructs her not to do what would be good for her. Those who identify "welfare" with "preference satisfaction" often simply seem driven to stipulating that in this case one's welfare *must* be advanced because one is getting what one "prefers." Here, I think, is a perfect example of the way that the ambiguity between the technical and ordinary senses of "prefer" (Section 2.1) leads to serious confusions.

Preferences regarding others also pose problems for Paretian welfarism. Suppose Alf is a prude who prefers that others do not read a somewhat racy book such as *Lady Chatterley's Lover*. It seems that according to Paretian welfarism Betty's decision to read the book cannot be approved of by the Pareto principle, since she is making Alf "worse off" when she reads the book: her reading the book (which we might normally think she has a right to do) moves him to a lower indifference curve. The way in which the Pareto principle can conflict with an individual's rights to decide what she is going to do has been analyzed by Amartya Sen.[9] Sen conceives of a person having a right as having authority to decide the social preference over at least one pair of alternatives (x,y) such that if a person chooses $x \succ y$ that is the social preference (let us call this social preference xPy); and if the person chooses $y \succ x$ then yPx (i.e., the social preference is y over x). Sen shows that attributing such rights to two persons, and assuming all possible orderings of social states are permissible, the social outcome selected by the rights can conflict with a version of the Pareto principle according to which, if for everyone $x \succ y$, then xPy. That is, if everyone prefers x to y, then the social preference must be x over y. Sen nicely summarizes his argument:

> There is a book (e.g. *Lady Chatterley's Lover*) which may be read by Mr. A ("the prude") or Mr. B ("the lascivious") or by neither. Given other things, these three alternatives define social states, a, b and o respectively. Consider now the following possibility. The prude A most prefers o (no one reading it), then a ("I'll take the hurt on myself"), and lastly b ("Imagine that lascivious lapping it up"). The lascivious [Mr. B] prefers most a ("it will give that lilywhite baby a nice shock"), then b ("it will be fun"), and last o ("what a waste of a good book"). On grounds of individual freedom, since B wants to read the book rather than no one reading it,

> b is socially preferred to o; note that in *either case* A does not read the book here. Similarly, since A does not want to read it, o is socially better than a. But a is Pareto-superior to b, yielding a preference cycle.[10]

So we get bPo (by Mr. B's right), oPa (by Mr. A's right), and aPb (by Pareto, since in both Mr. A's and Mr. B's ordering, $a \succ b$); so we get $bPoPaPb$—an intransitive result. Sen saw this not as a case against rights, but as showing "the unacceptability of the Pareto principle as a universal rule."[11] Sometimes it seems that a commitment to Pareto efficiency can lead us astray. One way to respond to this problem is to exclude some sorts of preferences from consideration: thus we might restrict our welfare Paretianism to self-regarding preferences (preferences over different states of one's own life) and so ignore preferences that other people do rather than not do certain things (or that they not read rather than read certain books). But this certainly does not solve all the problems, for there are problematic sorts of self-regarding preferences. Consider for example the problem of expensive preferences. Suppose I always prefer expensive over cheap goods, but you prefer many cheap goods over many expensive ones. Given these preference structures, Paretian welfarism approves of distributions that raise us both to higher indifference curves—giving me expensive, and you inexpensive, goods. To many this seems unfair.[12]

Fair and Unfair Starting Points Perhaps the most serious problem for Paretian welfarism is that it is insensitive to the distributions from which we begin. Suppose that all of Alf's and Betty's preferences are over quantities of pizza and chicken wings, but presently Alf has all the pizza and all the chicken wings and Betty has none. According to the Pareto principle, this is an efficient distribution. Since Betty has nothing to trade, there is no way to make her better off without making Alf worse off, so we have achieved Pareto optimality. But this hardly seems a moral reason to embrace the distribution. The heart of the Paretian project is to make people better off, and when no one can be made better off without lowering someone else's utility, the Pareto criterion has nothing more to say—we have achieved "optimality." But from the perspective of advancing human welfare it is hard to conceive of situations where Betty has nothing as "optimal." Plausible versions of Paretian welfarism thus seem committed to some idea of a fair starting point, and *then* can hold that Pareto-approved moves made from the initial fair starting point are moral improvements.

3.2 EFFICIENCY, EXTERNALITIES, AND PUBLIC GOODS

Externalities and Property Rights

Let us change our focus from the efficient *allocation*, to the efficient *production*, of a good. Suppose that you produce pizza: how much pizza is it efficient for you to make? The basic idea has already been explained in our analysis of an efficient consumption decision: you should produce up to the point where your marginal benefits equal your marginal costs. If you stop producing pizza while your marginal benefits are still greater than your marginal costs, it looks irrational insofar as production of an extra pizza yields more preference satisfaction than it costs (remember, our concern is opportunity costs). To produce pizza above and beyond the point where your marginal benefits equal your marginal costs means that your last pizza cost you more (in terms of preference satisfaction) than you received in benefits, which again looks to be an irrational decision. So a rational utility maximizer will produce up to the point where her marginal benefits equal her marginal costs.

The problem is that this need not be the efficient level of production for society if there are *positive or negative externalities*. An externality occurs when some person's consumption or production activity has positive or negative impact on the utility of others (where this impact is not included in the producer's or consumer's cost-benefit calculations).[13] If Alf's activity imposes negative externalities (costs) on Betty, then while as a rational economic agent Alf will engage in it up to the point were *his* marginal benefits equal *his* costs, Alf will not take account of the costs to Betty. But if Alf produces just up to the point where his marginal costs and benefits are equal, and there are additional costs to Betty, it looks as if the total social costs (the costs to Alf and Betty) of Alf's last unit of production exceeded the entire social benefits. This would violate the Pareto criterion: Alf has moved them both to a new distribution (with the extra produced unit of pizza) which benefits Alf at a cost to Betty. A similar analysis applies to external benefits: if my production has benefits to you as well as to me, then if I stop production when my marginal benefits equal my marginal costs, from a social point of view (which includes the benefits and costs to everyone) the good has been underproduced: social marginal benefits still exceed social marginal costs, since my decision has not taken into account the benefits you receive.

Externalities are a chief source of "market failure"—the failure of the market to produce efficient results. Only if Betty *fully internalizes* all the costs and benefits of her activity will she stop at just the point where social marginal benefits equal social marginal costs. So too in a trade: only if Alf and Betty fully internalize the costs and benefits of their trade can we say the trade necessarily moves us to a Pareto-superior state. If there are third-party costs (negative externalities), Alf and Betty might make trades where the social marginal costs exceed the social marginal benefits because they do not take account of the costs to Charlie; if third-party benefits exist, Alf and Betty may not trade even though the social marginal benefits exceed the social marginal costs.

Given this, the market would seem to produce efficient outcomes only if we have a scheme of property rights whereby an economic agent *internalizes all the costs and benefits of his activity*: he obtains the full benefits, and pays the full costs, of his activity. Consider the well-known case of the "tragedy of the commons" such as fisheries.[14] Many fisheries around the world are overfished, resulting in a depletion of stocks. Now it would probably be to the benefit of fisherman Alf to reduce his catch this year to secure a good yield next year *if he could be confident of obtaining all the future benefits of his reduced yield this year*. But he cannot: if Betty and Charlie fish anyway, the stocks will still be depleted. Alf will have paid a cost but will not gain the full benefits of his restraint. Conversely, Betty and Charlie do not pay the full costs of their overfishing, since the costs of depletion are transferred to other fishermen such as Alf. Thus the fisheries are overfished, and the marginal social costs exceed the marginal benefits. A scheme of property rights that internalized all costs and benefits would solve the problem.[15] Of course we have this problem just because property rights over fish in the ocean are difficult to institutionalize (fish tend to swim around a lot).

However, the ideal of a system of property rights that internalizes all benefits and costs is unrealizable unless we restrict what counts as an externality. Think about Sen's case of Mr. Prude's and Mr. Lascivious's preferences about reading *Lady Chatterley's Lover*. If Mr. Lascivious exercises his right and reads the book, there is an externality: Mr. Prude is made worse off, since Mr. Prude prefers that no one reads it to Mr. Lascivious reading it. Mr. Lascivious has negatively impacted the utility function of Mr. Prude. If people have preferences over what others do or don't do, then externalities will be everywhere. Suppose Alf prefers that people shop at Target rather than Wal-Mart; if so, every transaction at Wal-Mart involves a negative externality.

One possible solution to this difficulty might be called the *rights-based solution*, according to which Alf's action has a negative externality on Betty if and only if it violates a right of Betty's. Rights, we might say, protect a certain set of preferences: impinging on *those* preferences constitutes cost or harm to an individual. The rights-based solution is suggested by John Stuart Mill, who was especially concerned that people might be held accountable to others for every cost they impose upon them, including "costs" to others resulting from performing actions that their neighbors simply don't like. Mill argued that such "costs" should be ignored, and people should only be said to impose recognized social costs on others when they set back "certain interests, which ... ought to be considered as rights."[16] Thus, says Mill,

> Encroachment on their rights; infliction on them of any loss or damage not justified by his own rights; falsehood or duplicity in dealing with them; unfair or ungenerous use of advantages over them; even selfish abstinence from defending them against injury—these are fit objects of moral reprobation, and, in grave cases, of moral retribution and punishment.[17]

The idea, then, is that we identify a crucial set of interests (or sets of preferences over certain aspects of our life), and hold that if an action or transaction imposes costs on other parties in terms of *these preferences*, the action or transaction has a social cost (i.e., rights have been infringed). That the action involves a social cost does not show that it should be prohibited, since the social benefits may still outweigh the costs.[18]

This results in a moralistic conception of efficiency: we must first know which subsets of a person's preferences are protected by his rights before we can know what constitutes an efficient level of any activity. If, as Mill emphatically argued, no one has a right that others don't read (rather than read) books one finds offensive, Mr. Prude incurs no cost when Mr. Lascivious reads *Lady Chatterley's Lover*—there is no externality because no right was violated. There are, though, real worries about this moralistic view. For one, it does not make sense of a core argument of most liberal political economists, viz., that we should evaluate systems of property rights in terms of their efficiency-promoting characteristics. If we are to say that private property rights promote efficiency we must be able to first identify what an efficient level of production would be, and then show that

private property rights are apt to result in this level. But according to the rights-based solution we must know what our property and other rights are *before* we can identify externalities, and so efficient outcomes. Say that Alf wants to build a tavern on his land and Betty objects. What is the efficient outcome? If Alf has the rights on his side, then the efficient outcome is that he builds it; if Betty has the rights on her side—she has a right not to have her property values lowered by living next to a tavern—then the efficient outcome is that the tavern is not built. If both have rights then we must still somehow weigh up the costs and benefits.

Ronald Coase proposes an analysis that is almost the reverse of the rights-based view: on Coase's view achieving an efficient outcome does not depend on the specific way that the initial rights are assigned.[19] Suppose that we live in a world free of transaction and bargaining costs, and in this world Alf has a factory that produces whatchacallits that generates smoke as a by-product; suppose that Betty has a laundry, and her costs are increased because of the smoke from Alf's factory. Alf's production, then, produces a negative externality. So if Alf produces whatchacallits up to the point where his marginal costs equal his marginal benefits, too many whatchacallits will be produced. Suppose that Alf makes $3,000 per year; Betty presently earns $24,000 from her laundry, but she would make $31,000 if Alf's smoke didn't increase her costs. Alf, then, imposes an externality of $7,000. Assuming no laws against pollution, we can still achieve Pareto efficiency: Betty can pay Alf $3,001 to stop producing whatchacallits; he will be better off and so will she, so the move is Pareto-superior.

It is important that Coase's theorem applies regardless of how the property rights are divided between Alf and Betty: a Pareto outcome can be reached whether Alf has a right to pollute or Betty has a right that he not pollute. Suppose that Alf is now making $10,000 producing whatchacallits while Betty's profits remain the same: $24,000 given Alf's pollution and $31,000 without it. Assume that Betty has the right that Alf not pollute, and so can bring suit against him. Now it is Pareto-efficient for Alf to bribe Betty not to bring suit: he can pay her $7,001 to refrain from bringing suit, and both are better off.[20]

According to Coase, then, *in the absence of transaction and bargaining costs, parties to an activity with externalities will agree to some Pareto-efficient allocation of resources regardless of the initial distribution of property rights.* Coase's theorem calls into question one of the traditional justifications for government regulation. In the absence of a perfect scheme of property rights that internalizes both costs and benefits, it has been widely

argued, government is necessary to regulate the "market failure" that results from externalities. But Coase shows that market transactions can solve the problem of externalities and get us to a Pareto-optimal outcome (though the actual costs involved in negotiation, etc., may preclude this).

Public Goods

Related to the question of externalities is the special case of public goods. Public goods are defined in terms of two characteristics. *First*, they are characterized by *nonrival consumption*. Consider clean air. If it is provided at all, it can be provided to Alf without taking any of it away from Betty. Once the good is there, consumers do not compete for it; everyone can freely use it without diminishing the amount left for others. *Second*, we cannot control the flow of benefits from public goods: they are *nonexcludable*. If a public good is provided, it is provided for all to use. If we clean the air, everyone has clean air. We cannot exclude those who have not paid their share. A pure public good is one that perfectly meets these two conditions. In most cases both conditions are not perfectly met; many goods are thus *quasi-public goods*. Defense, law and order, regulation of air pollution, highways, ports, public works, and elementary education are among the goods usually cited as being quasi-public goods, though some economists have disputed the "publicness" of almost every item on this list. The classic textbook example of a public good was typically a lighthouse: a lighthouse warns all ships away from danger—one ship does not "use up" the light (so there is nonrival consumption) and it is not possible to exclude the light from the ships who did not pay for the lighthouse (so nonexcludability is met).

When we look at game theory in the next chapter, we shall explain more formally why public goods tend to be undersupplied, but the crux of the explanation appears (at least at first sight) clear. Even if everyone prefers having the public good to not having it, each of us will receive it for free if someone else pays for it. After all, the benefits are nonexcludable: if anyone gets the good, everyone does. So we typically have an incentive to *free-ride*: each, hoping the other pays, holds back from paying.

Thus the classic public good argument is for state action to fix the market's failure to generate efficient outcomes. In the interests of efficiency, it is often said, government must require everyone to contribute to the production of such goods. While powerful, the

argument is not quite as simple as it seems. Three points must be kept in mind.

1. For an adequate public goods argument for state action, it must not only be the case that everyone wants the good, but that in everyone's preference ordering {contributing to secure the good & paying my share} is preferred to {not paying my share & not having the good}. If mandatory taxation to supply the good is to move us to a Pareto-superior condition, it is not enough that everyone wants the good; they must prefer having it *and* paying for it to not having it.

2. It is not the case that markets never supply public goods, or never do so efficiently. Suppose Alf's goat wanders into Betty's garden and eats her veggies, and Betty's dog wanders into Alf's property, scaring his goat so that it does not give milk.[21] A fence would be a public good. Assume that each would benefit by unilaterally building the fence (he/she would be better off building the fence alone than not having one) but, of course, each would prefer that the other build the fence. So each has the following ordering: (1) the other builds, (2) we split the cost, (3) I build, (4) neither builds. In such a case, since each would prefer to pay for the entire good rather than do without it, the public good will be provided (and, we shall see in Section 4.3, one person will pay for the entire good). Provision of public goods do not constitute a market failure until we add further conditions, such as that no one individual's utility function is such that it is rational for him to purchase the entire good at the efficient level,[22] there are a large number of people (so each is tempted to free-ride, or we do not know how many people's cooperation is necessary to secure the good), etc. As the number of individuals involved increases, the need for some sort of formal agreement about allocation of contributions becomes necessary.[23] It is, then, not simply public goods per se, but public goods that require collective action of a large number of agents that are most likely not to be adequately supplied by the market.[24]

3. Whether public goods will be voluntarily supplied, and whether universal contribution is efficient, also depends on the relation between securing the goods and individual contribution. We can distinguish three basic types of relations:

   a. *Constant Returns*. If G is the total amount of the good produced, and G_i is the contribution of any individual i,

then $G = G_1 + G_2 + G_3 + \ldots G_n$. Each individual contribution adds to the amount of the public good secured. An example here is picking up after your dog in a public park: each person's contribution helps secure the good of a clean park, where all can walk without fear.

b. *Threshold at the Top.* If G is the total amount of the good produced, and G_i is the contribution of any individual i, then there exists some person k such that $G = \{G_1 + G_2 + G_3 + \ldots G_j\} = \{G_1 + G_2 + G_3 + \ldots G_j + G_k + \ldots G_n\}$. The contribution of individual k and those who follow her add nothing to the amount of public good secured. Suppose that we all support a candidate for office, and she needs 10,000 signatures to get on the ballot. After she has obtained 10,000 valid signatures, additional signatures do her no good (at least from a legal perspective)—they do not help her get on the ballot.

c. *Threshold at the Bottom.* If G is the total amount of the good produced, and G_i is the contribution of any individual i, then there exists some person k such that $\{G_1 + G_2 + G_3 + \ldots G_j\} = 0$, but $\{G_1 + G_2 + G_3 + \ldots G_j + G_k\} > 0$. Until the contribution of individual k is secured, no public good is achieved. Again, think of our candidate: until the 10,000th signature is obtained, no good at all is produced.

We can identify both pure and mixed cases (i.e., cases of public goods that combine these types). We might have, for example, a public good that gives constant returns up to a top threshold, but then no more returns thereafter, or which has a bottom threshold, and constant returns thereafter, and so on. Consider three interesting cases: (i) a simple case of constant returns; (ii) a threshold at the bottom where the k person is also the n, or last person; and (iii) a case of constant returns up to a threshold at the top. Will the good be provided by each person maximizing her own utility, or is some sort of coordination or authority necessary to secure the good?

(i) *A simple case of constant returns.* In the simple case of constant returns each individual contribution helps secure a greater level of the public good. So Alf's action always secures some of the good: he will incur some costs $\mu(c)$, but since he wants the good, he also secures some benefit from his contribution, $\mu(b)$. The

problem is that while Alf's $\mu(c)$ is the total costs of his share of producing the good, his benefits $\mu(b)$ are just a small part of the total social benefits, since everyone gains from his contribution (remember, the good is nonexcludable and nonrival). As a rational economic agent Alf stops contributing when his marginal costs equal the marginal benefits *to him*. But this will not adequately take account of the overall social marginal benefits of his contribution, since all others will benefit from the higher level of the good he will provide. This is a classic case where public goods will be underprovided by uncoordinated individual choices.

(ii) *A threshold at the bottom where the* k *person is also the* n, *or last person.* In the second and very special case, *everyone's contribution is required if the good is to be secured at all.* An example is a crew of a small boat; unless everyone rows, the boat will not make headway against the strong current. No public good is secured unless everyone contributes—the public good of reaching the destination will not be achieved unless everyone does her job. Here in an interesting case where the public good is apt to be achieved by purely voluntary choices based simply on individual utility maximization, since no individual has an incentive to over- or undercontribute.[25]

(iii) *A case of constant returns up to a threshold at the top.* The third case is interesting because here it is *inefficient* to require everyone to contribute all the time. Suppose at Alf's College there is a lawn between two buildings that are located diagonally across from each other.[26] Everyone would prefer a nice lawn between the two buildings to a shoddy one. But everyone also is inconvenienced by having to walk all the way around the quad (where the walkways are). Each person would prefer {having a nice lawn *and* cutting across diagonally}—the shortest route between the two buildings—to {having a nice lawn *and* always using the walkways}. If everyone cuts across, the lawn will be ruined; but if only 10 out of a 100 people do so, there will be no problem. (Hence the threshold at the top; after the 90th person avoids walking across the lawn, no further public good is produced.) Consider three policies: (a) no one crosses, so we have a beautiful lawn but everyone is inconvenienced; (b) 10 people cross, so we have a beautiful lawn and only 90 people are inconvenienced; (c) everyone crosses 10% of the time. The

second and third policies are Pareto improvements on the first. Ten people can be made better off (they can cut across the lawn) without any additional costs to others, or everyone can be made better off 10% of the time: the second and third policies achieve just as much of the public good as the first policy, but at a lower cost. As far as efficiency is concerned, if we are assuming *Homo Economicus* we have no grounds for choosing between the second and third policies (but if we assume that fairness is part of people's utility function, perhaps the third is to be preferred; see Section 2.4). We see here that a public policy based on the pursuit of efficiency does not necessarily seek to eliminate all free-riding (receiving the benefits without paying the costs); it might even (as policy b does) seek to *secure* an optimum level of free-riders (with policy b, 10).

None of this is intended to undermine the basic idea that usually the state should tax everyone to secure public goods. The point is that we should be aware that the necessity and desirability of state action to secure universal contribution is by no means an immediate inference from the mere existence of a public good and the pursuit of efficiency.

SUMMARY

In this chapter I have tried to explain the notion of efficiency and highlight its relation to rationality. This chapter has:

- *Explained why a rational consumer will consume a good up to the point where her marginal benefits equal her marginal costs.* We have also seen that a rational producer should produce up to the point where her marginal benefits equal her marginal costs.
- *Explained why rational consumers will make exchanges that are Pareto improvements.* The Edgeworth Box was explained, and we considered the relation of the contract curve to Pareto-optimal bargains.
- *Explained the idea of Pareto superiority and Pareto optimality.*
- *Considered whether the Pareto principle is suitable as a moral ideal.*
- *Analyzed the notion of an externality, and considered whether the ideal of efficient property rights that internalize all the costs and benefits of*

activity makes sense. The problem, we saw, is that anytime another person negatively impacts my utility, and this is not taken into account in her decision, she imposes an externality on me. But if my preferences are about what she should and should not do, she will impose an externality on me simply by living her life as she sees fit. This problem of adequately defining an externality is a major difficulty for application of the Pareto principle.

- *Sketched the Coase theorem.*
- *Considered why public goods tend to be undersupplied by voluntary action, and described some cases where voluntary action will secure them.*

NOTES

1. See Arthur M. Okun, *Equality and Efficiency: The Big Tradeoff.*
2. Named after Francis Edgeworth (1845–1926) who depicted alternative allocations of resources, and possibilities for contracts, in this way.
3. Of course if we bring in endowment effects (Section 2.5), and Alf and Betty have a preference to keep what they already have, then they may not trade. We can now better see why endowment effects worry economists: they go to the very heart of efficiency.
4. Some distinguish allocative from Pareto efficiency; allocative efficiency is said to obtain when marginal benefits equal marginal costs, and this is distinguished from the Pareto criterion. See John C. Winfrey, *Social Issues: The Ethics and Economics of Taxes and Public Programs,* pp. 26–27. There are other concepts of efficiency employed in economics. *X-efficiency* concerns getting maximum outputs for a given level of inputs; *dynamic efficiency* concerns maximizing growth; and *technological efficiency* concerns the use of the best technology. On the different notions of efficiency see Charles Wolf, Jr., *Markets or Governments: Choosing between Imperfect Alternatives,* pp. 126ff.
5. For a careful and useful analysis of these issues, see Alan Buchanan, *Ethics, Efficiency, and the Market,* especially chaps. 1 and 3.
6. See Daniel Hausman and Michael McPherson, *Economic Analysis and Moral Philosophy,* pp. 93ff.
7. For examples of contemporary welfarists who rely on this test, see Richard A. Epstein, *Skepticism and Freedom;* Louis Kaplow and Steven Shavell, *Fairness versus Welfare.* This test is also fundamental to much social cost-benefit analysis.

8. Russell Hardin makes much of the point in this paragraph. See his *Indeterminacy and Society*, pp. 10–11.
9. Amartya Sen, "The Impossibility of a Paretian Liberal." For an extended, and accessible, discussion, see Sen's, "Liberty, Unanimity, and Rights."
10. Sen, "Liberty, Unanimity, and Rights," p. 218.
11. Ibid., p. 235.
12. See here Hausman and McPherson, *Economic Analysis and Moral Philosophy*, pp. 78–79.
13. See Dennis Mueller, *Public Choice III*, p. 25.
14. See Garret Hardin, "The Tragedy of the Commons."
15. See David Schmidtz, *The Limits of Government: An Essay on the Public Goods Argument*.
16. John Stuart Mill, *On Liberty*, Chapter 4, para. 3.
17. Ibid., Chapter 4, para. 14.
18. Ibid., Chapter 5, para. 3.
19. See Ronald Coase, "The Problem of Social Cost." My explication follows Mueller, *Public Choice III*, pp. 27–30.
20. The Coase theorem can be applied to more realistic cases, where people's activities have variable costs—the more they purchase of a good, the higher the externalities. See Mueller, *Public Choice III*, pp. 27ff.
21. I am following ibid., pp. 16ff.
22. Because the purchase of public goods has a positive externality, they will not generally be supplied at the efficient level.
23. But this is not to say that governmental, coercive action is necessary. See David Schmidtz, *The Limits of Government*.
24. Mueller gives a nice overview of the extent to which individual voluntary choices will secure public goods, and how this results in undersupply. *Public Choice III*, pp. 18ff. The discussion that follows draws on this part of Mueller.
25. See ibid., p. 22. But see the analysis of the assurance problem in Section 4.3.
26. See David Lyons, *The Forms and Limits of Utilitarianism*, pp. 69ff, 162ff. I have discussed this case in more detail in my *Social Philosophy*, pp. 182ff.

4

Game Theory

OVERVIEW

In our examination of utility theory in Chapter 2 we focused on a rational agent as one who has preferences over outcomes and a set of fixed action-options, and who can correlate outcomes with action-options; her preferences over outcomes determined her preferences over action-options. In Chapter 3 we began to consider how rational utility maximizers (who also are characterized by the additional features of *Homo Economicus*) interact, and especially how rational economic agents will engage in efficient transactions. But the analysis of rational interaction in Chapter 3 focused on Pareto-superior moves—roughly, cases in which agents' interests converged. I largely ignored the possibility of conflict. We now move to the theory of games—a general theory of what is rational when interacting with other rational agents, and especially when what is best for Betty may not be best for Alf. In this chapter I employ only the general idea of individuals as utility maximizers: the specific, additional features that are required for *Homo Economicus* are not central to this chapter. (Again we see why it is so important to distinguish *Homo Economicus* from rational utility maximizers in general.)

The chapter begins with the simplest sort of game, a "zero-sum" game in which whatever one person wins the other loses. This is a game of *pure conflict*. We then move on to the famous "Prisoner's Dilemma" and other "variable-sum" games—some will be focused on conflict, but we will also look at games in which rational players will cooperate. It is important to realize that although the theory of