PART I

Origins of the state

CHAPTER 2

The reason for collective choice – allocative efficiency

Had every man sufficient *sagacity* to perceive at all times, the strong interest which binds him to the observance of justice and equity, and *strength of mind* sufficient to persevere in a steady adherence to a general and a distant interest, in opposition to the allurements of present pleasure and advantage, there had never, in that case, been any such thing as government or political society; but each man, following his natural liberty, had lived in entire peace and harmony with all others. (Italics in original)

David Hume

Government is a contrivance of human wisdom to provide for human *wants*. Men have a right that these wants should be provided for by this wisdom. (Italics in original)

Edmund Burke

2.1 Public goods and prisoners' dilemmas

Probably the most important accomplishment of economics is the demonstration that individuals with purely selfish motives can mutually benefit from exchange. If *A* raises cattle and *B* corn, both may improve their welfare by exchanging cattle for corn. With the help of the price system, the process can be extended to accommodate a wide variety of goods and services.

Although often depicted as the perfect example of the beneficial outcome of purely private, individualistic activity in the absence of government, the invisible hand theorem presumes a system of collective choice comparable in sophistication and complexity to the market system it governs. For the choices facing *A* and *B* are not merely to trade or not, as implicitly suggested. *A* can choose to steal *B*'s corn, rather than give up his cattle for it; *B* may do likewise. Unlike trading, which is a positive-sum game benefiting both participants in an exchange, stealing is at best a zero-sum game. What *A* gains, *B* loses. If stealing, and guarding against it, detract from *A* and *B*'s ability to produce corn and cattle, it becomes a negative-sum game. Although with trading each seeks to improve his position and both end up better off, with stealing the selfish pursuits of each leave them both worse off.

The example can be illustrated with strategy Matrix 2.1. To simplify the discussion, let us ignore the trading option and assume that each individual grows only corn. Square 1 gives the allocation when *A* and *B* both refrain from stealing (*A*'s

Matrix 2.1. *Stealing as prisoners' dilemma*

A \ B	*Does not steal*	*Steals*
Does not steal	1 (10, 9)	4 (7, 11)
Steals	2 (12, 6)	3 (8, 8)

allocation precedes *B*'s in each box). Both are better off when they both refrain from stealing, but each is still better off if he alone steals (cells 2 and 4). In Matrix 2.1, stealing is a dominant strategy for both players, so defined because it dominates all other strategy options by promising a higher payoff for the chooser than any other strategy, given any choice of strategy by the other player. In an anarchic environment, the independent choices of both individuals can be expected to lead both to adopt the dominant stealing strategy with the outcome cell 3. The distribution of corn in cell 3 represents a "natural distribution" of goods (so named by Bush [1972]), namely, the distribution that would emerge in an Hobbesian state of nature.

From this "natural" state, both individuals become better off by tacitly or formally agreeing not to steal, provided that the enforcement of such an agreement costs less than they jointly gain from it. The movement from cell 3 to cell 1 is a Pareto move that lifts the individuals out of a Hobbesian state of nature (Bush, 1972; Bush and Mayer, 1974; Buchanan, 1975a; Schotter, 1981). An agreement to make such a move is a form of "constitutional contract" establishing the property rights and behavioral constraints of each individual. The existence of these rights is undoubtedly a necessary precondition for the creation of the "postconstitutional contracts," which make up a system of voluntary exchange (Buchanan, 1975a). Problems of collective choice arise with the departure from Hobbesian anarchy, and are coterminous with the existence of recognizable groups and communities.

A system of property rights and the procedures to enforce them are a Samuelsonian public good in that "each individual's consumption leads to no subtraction from any other individual's consumption of that good."[1] Alternatively, a pure public good can be defined as one that *must* be provided in equal quantities to all members of the community. Familiar examples of pure public goods are national defense and police and fire protection. National defense is the collective provision against external threats; laws and their enforcement safeguard against internal threats; fire departments against fires. Nearly all public goods whose provision requires an expenditure of resources, time, or moral restraint can be depicted with a strategy box analogous to Matrix 2.1. Replace stealing with paying for an army, or a police force,

[1] Samuelson (1954, p. 386). The extent to which individuals can be excluded from the benefits of a public good varies. One man's house cannot be defended from foreign invasion without defending another's, but a house may be allowed to burn down without endangering another. Tullock (1971c) has suggested that voluntary payment schemes for excludable public goods could introduce cases resembling the latter.

or a fire department, and the same strategy choices emerge. Each individual is better off if all contribute to the provision of the public good than if all do not, and each is still better off if only he does not pay for the good.

A pure public good has two salient characteristics: jointness of supply and the impossibility or inefficiency of excluding others from its consumption, once it has been supplied to some members of the community (Musgrave, 1959, pp. 9–12, 86; Head, 1962). Jointness of supply is a property of the production or cost function of the public good. The extreme case of jointness of supply is a good whose production costs are all fixed, and thus whose marginal production costs are zero (e.g., a public monument). For such a good, the addition of more consumers (viewers) does not detract from the benefits enjoyed by others. Even a good with falling average costs, although positive marginal costs, has elements of jointness that raise collective provision issues.

The joint supply characteristic creates the potential gain from a cooperative move from cell 3 to 1. Given jointness of supply, a cooperative consumption decision is necessary to provide the good efficiently. If it took twice as many resources to protect *A* and *B* from one another as it does to protect only one of them, collective action would be unnecessary in the absence of nonexclusion. Each could choose independently whether or not to provide his own protection.

People can be excluded from the benefits from viewing a statue placed within a private gallery if they do not pay to see it. But people cannot be prevented from viewing a statue or monument placed in the central city square. For many public goods, the exclusion of some members of the community from their consumption is impossible or impractical. Failure of the exclusion principle to apply provides an incentive for noncooperative, individualistic behavior, a gain from moving from cell 1 to either cell 2 or cell 4. The impossibility of exclusion raises the likelihood that purely voluntary schemes for providing a public good will break down. Thus, together, the properties of public goods provide the raison d'être for collective choice. Jointness of supply is the carrot, making cooperative-collective decisions beneficial to all; absence of the exclusion principle is the apple tempting individuals into independent, noncooperative behavior.

Although the purest of pure public goods is characterized by both jointness of supply and the impossibility of exclusion, preference revelation problems arise even if only the first of these two properties is present. That is, an alternative definition of a public good is that it *may* be provided in equal quantities to all members of the community at zero marginal cost. The substitution of "may" for "must" in the definition implies that exclusion may be possible. A classic example of a public good fitting this second definition is a bridge. In the absence of crowding, the services of the bridge can be supplied to all members of the community, but they need not be. Exclusion is possible. As long as the marginal cost of someone's crossing the bridge remains zero, however, excluding anyone who would experience a marginal benefit from crossing violates the Pareto principle. Jointness of supply alone can create the need for collective action to achieve Pareto optimality.

Matrix 2.1 depicts the familiar and extensively analyzed prisoners' dilemma. The salient feature of this game is that the row player ranks the four possible outcomes

$2 > 1 > 3 > 4$, while the column player has the ranking $4 > 1 > 3 > 2$.[2] The noncooperative strategy is dominant for both players. It is the best strategy for each player in a single play of the game regardless of the other player's strategy choice. The outcome, square 3, is a Cournot-Nash equilibrium.[3] It has the unfortunate property of being the only outcome of the prisoners' dilemma game that is not Pareto optimal. From each of the other three squares a move must make at least one player worse off, but from 3 a move to 1 makes both better off.

Despite the obvious superiority of the cooperative nonstealing outcome to the joint stealing outcome, the dominance of the stealing strategies ensures that the nonstealing strategies do not constitute an equilibrium pair, at least for a single play of the game. The cooperative solution may emerge, however, as the outcome of a "supergame" of prisoners' dilemma games repeated over and over by the same players. The cooperative solution can arise, even in the absence of direct communication between the players, if each player chooses a supergame strategy that effectively links his choice of the cooperative strategy in a single game to the other player's choice of this strategy. One such supergame strategy is for a player to play the same strategy in the present game as the other player(s) played in the previous game. If both (all) players adopt this strategy, *and* all begin by playing the cooperation strategy, the cooperative outcome emerges in every play of the game. This "tit-for-tat" strategy beat all others proposed by a panel of game theory experts in a computer tournament conducted by Axelrod (1984).

An alternative strategy, which achieves the same outcome, is for each player to play the cooperative strategy as long as the other player(s) does, and then to *punish* the other player(s) for defecting by playing the noncooperative strategy for a series of plays following any defection before returning to the cooperative strategy. Again, if all players begin by playing cooperatively, this outcome continues throughout the game (Taylor, 1987, ch. 3). In both of these cooperative strategies, equilibrium solutions to the prisoners' dilemma supergame, the equilibrium comes about through the *punishment* (or threat thereof) of the noncooperative behavior of any player, in this case by the noncooperation of the other player(s). This idea that noncooperative (antisocial, immoral) behavior must be punished to bring about conformity with group mores is to be found in most, if not all, moral philosophies, and forms a direct linkage between this large literature and the modern theory.[4]

When the number of players in a prisoners' dilemma game is small, it is obviously easier to learn their behavior and predict whether they will respond to cooperative strategy choices in a like manner. It is also easier to detect noncooperative behavior and, if this is possible, single it out for punishment, thereby further encouraging the

[2] An additional assumption that row player's payoff in box 2 and column's in box 4 add up to less than their two payoffs in box 1 is needed to ensure that they do not take turns jointly defecting and cooperating; that is, not stealing from one another for two periods yields higher payoffs than taking turns stealing from one another.

[3] A set of strategies $S = (s_1, s_2, \ldots, s_i, \ldots, s_n)$ constitutes a Nash equilibrium, if for any player i, s_i is his optimal strategy, when all other players $j \neq i$ play their optimal strategies $s_j, s_j \in S$.

[4] For classical discussions of moral behavior and punishment, which are most modern and in line with the prisoners' dilemma discussion, see Hobbes, *Leviathan* (1651, chs. 14, 15, 17, 18), and Hume (1751, pp. 120–7).

cooperative strategies. When numbers are large, it is easy for one or a few players to adopt the noncooperative strategy and either not be detected, since the impact on the rest is small, or not be punished, since they cannot be discovered or it is too costly to the cooperating players to punish them. Thus, voluntary compliance with behavioral sanctions or provision of public goods is more likely in small communities than in large (Coase, 1960; Buchanan, 1965b). Reliance on voluntary compliance in large communities or groups leads to free riding and the under- or nonprovision of the public good (Olson, 1965).

In the large, mobile, heterogeneous community, a formal statement of what behavior is mutually beneficial (e.g., how much each must contribute for a public good) may be needed even for individuals to know what behavior is consistent with the public interest. Given the incentives to free ride, compliance may require the implementation of individualized rewards or sanctions. Olson (1965, pp. 50–1, 132–67) found that individual participation in large, voluntary organizations like labor unions, professional lobbies, and other special interest groups was dependent not on the collective benefits these organizations provided for all of their members, but on the individualized incentives they provided in the form of selective benefits for participation and attendance, or penalties in the form of fines, and other individualized sanctions.

Thus, democracy, with its formal voting procedures for making and enforcing collective choices, is an institution that is needed by communities of only a certain size and impersonality. The family makes an array of collective decisions without ever voting; a tribe votes only occasionally. A metropolis or nation state may have to make a great number of decisions by collective choice processes, although many of them may not correspond to what we have defined here as a democratic process.[5] Similarly, small, stable communities may be able to elicit voluntary compliance with group mores and contributions for the provision of local public goods by the use of informal communication channels and peer group pressure. Larger, more impersonal communities must typically establish formal penalties against asocial behavior (like stealing), levy taxes to provide for public goods, and employ a police force to ensure compliance.

The size of the community, its reliance on formal sanctions and police enforcement, and the breakdown of the prisoners' dilemma may all be dynamically related. Detection of violators of the prisoners' dilemma takes time. An increase in the number of violations can be expected to lead to a further increase in violations but only with a time lag. If, because of an increase in community size or for some other reason, the frequency of violations were to increase, the frequency of violations in later periods could be expected to increase; the frequency of violations in still later periods would increase even further, and with these the need for and reliance on police enforcement of the laws. Buchanan (1975a, pp. 123–9) has described such a process as the erosion of a community's legal (that is, rule-abiding) capital.[6] Today, this form of capital is typically referred to as *social capital*. Putnam (2000) provides

[5] One must also keep in mind that democracy is but one *potential* means for providing public goods. Autocracies and oligarchies also provide public goods to "their" communities. Autocracies are discussed in Chapter 18.

[6] See Buchanan (1965b).

evidence of a dramatic decline in the stock of social capital in the United States over the recent generation.

Taylor (1987, pp. 168–79) relates the breakdown of the cooperative solution to the prisoners' dilemma not to the size of the community, however, but to the level of government intervention itself.[7] Intervention of the state in the provision of a community want or in the enforcement of social mores psychologically "frees" an individual from responsibility for providing for community wants and preserving its mores. State intervention leads to increased asocial behavior requiring more state intervention, and so on. Frey (1997b) makes an analogous argument. State-initiated bribes and sanctions designed to elicit cooperative behavior may "crowd it out" by destroying the intrinsic motivation of individuals to behave morally and as good citizens. These theories might constitute one explanation for the rising government expenditures that have occurred in this century. The increasing mobility and urbanization that have occurred during the century induce less voluntary cooperation by citizens and cause more state intervention. State intervention in turn reduces the internally motivated propensity for citizens to cooperate, necessitating still more state intervention.

This scenario of an unraveling of the social fabric mirrors to a remarkable degree the description by Rawls (1971, pp. 496–504) of the evolution of a just society, in which the moral (just, cooperative) behavior of one individual leads to increasingly moral behavior by others, reinforcing the cooperative behavior of the first and encouraging still more. The dynamic process in these two scenarios is the same, only the direction of change is reversed.

2.2 Coordination games

The prisoners' dilemma is a dilemma because cheating on the cooperative solution to the game is rewarded and, thus, individually rational. All situations in which one person's utility depends on the action of another do not reward "cheating," and thus do not give rise to the kind of collective action problem that characterizes the prisoners' dilemma. One such situation involves a *coordination* game.

Matrix 2.2 depicts one such game. If Row and Column both play strategy A they both receive the positive payoff a. If they coordinate on strategy B, they both receive a positive b, and if they fail to coordinate, they both receive a payoff of zero. Now suppose that each player knows all of the payoffs in Matrix 2.2 and must choose a strategy independently from the other player and in ignorance of the other player's strategy choice. Which strategy should a rational individual choose? Both players know that the other would like to choose the same strategy but, without knowledge of the other player's choice, there is obviously no unequivocal choice that a player can make.

[7] Indeed, "the main point [Taylor] set out to establish" was that "Cooperation can arise in the Prisoners' Dilemma supergame, no matter how many players there are" (1987, p. 104). On the next page he concedes, however, that "it is pretty clear that Cooperation amongst a relatively large number of players is 'less likely' to occur than Cooperation amongst a small number" (p. 105).

Matrix 2.2. *A coordination game*

G \ D	Strategy A	Strategy B
Strategy A	1 (a, a)	4 (0, 0)
Strategy B	2 (0, 0)	3 (b, b)

Suppose, however, that $b > a$. Clearly, both players now have a preference for coordinating on strategy B. Strategy B becomes a form of *Schelling point*, and both can be expected to choose this strategy (Schelling, 1960). But what if $b = a$? Now it would appear that our two players have little choice other than resorting to a coin flip – unless, of course, they were allowed to communicate with one another. With $b = a$, the two players are indifferent between coordinating on strategy A or B. If one of them were to propose that they coordinate on strategy B, the other would have no reason to object, *and he would have no reason to defect once the agreement had been reached.* Coordination games thus have an inherent stability to them that is absent in many other social-dilemma games, like the prisoners' dilemma.

Indeed, because of this inherent stability, Pareto-optimal sets of strategies can be expected to emerge when coordination games are repeated, under far less demanding behavioral assumptions than are needed to sustain Pareto-optimal outcomes in prisoners' dilemma supergames. Assume, for example, that all individuals are ignorant of the payoffs from the different combinations of strategies, the choices that the other player has made in the past, and the current choice of the other player. The only information a player has is what *her own* strategy choices were over a finite number of past plays of the game, and the payoffs she received. Given this limited knowledge she chooses to play the strategy that was most highly rewarded in the recent past.

For example, suppose that she can only recall the outcomes of the last five plays of the game, when she played A three times and B twice. Two of the three times that she played A, she got a; one of the two times that she played B, she was rewarded with b. She opts to increase the frequency with which she plays strategy A. If the other player adopts the same rule of thumb, the two players coordinate over time on strategy A and remain locked in on it so long as the payoff structure does not change.

Recent contributions to evolutionary game theory have modeled individual action as *adaptive learning*, wherein an individual's strategy choice today depends on the payoffs she, or those she can observe, have received in the recent past. These models demonstrate how coordinated strategy choices can emerge in games like that in Matrix 2.2.[8] These results are of great significance because they are based on far more realistic assumptions about the capacities of individuals to engage in rational

[8] See, for example, Sugden (1986); Warneryd (1990); Kandori, Mailath, and Rob (1993); and Young (1993).

Matrix 2.3. *Fence building as a game of chicken*

G \ D	*Contributes to building fence*	*Does not contribute*
Contributes to building fence	1 (3, 3)	4 (2, 3.5)
Does not contribute	2 (3.5, 2)	3 (1, 1)

action and about the ways in which learning takes place. They show how social conventions might evolve to solve coordination problems *without the need for the state.*[9]

Examples of coordination games include various conventions about driving: drive on the right, pass on the left, yield to cars approaching from the right, and so on. If all problems caused by social interaction were as simple as deciding on which side of the road everyone should drive, one might well imagine that it would be possible to do away with the state. But, alas, this is not the case, as our discussion of the prisoners' dilemma has already shown and the game of chicken further illustrates.

2.3 Public goods and chickens

The prisoners' dilemma is the most frequently used characterization of the situations to which public goods give rise. But the technology of public goods provision can be such as to generate other kinds of strategic interactions. Consider the following example.

The properties of two individuals share a common boundary. G owns a goat that occasionally wanders into D's garden and eats the vegetables and flowers. D has a dog that sometimes crosses into G's property, chasing and frightening the goat so that it does not give milk. A fence separating the two properties could stop both from happening.

Matrix 2.3 depicts the situation. With no fence, both D and G experience utility levels on one. The fence costs $1,000 and each would be willing to pay the full cost if necessary to get the benefits of the fence. The utility levels of each (2) are higher with the fence than without it, even when they must pay the full cost alone. This assumption ensures that the utility levels of both individuals are still higher if each must pay only half the cost of the fence (square 1). Last of all, each is, of course, best off if the fence is built and he pays nothing (payoffs of 3.5 to G and D, respectively; in squares 2 and 4). Matrix 2.3 depicts the game of "chicken." It differs from the prisoners' dilemma in that the outcome in which no one contributes (cell 3), which is Pareto inferior to the outcome that both contribute (cell 1), is not an equilibrium. Since each individual is better off even if he must pay for the fence alone, each would be willing to move to square 2 or 4, as the case may be, rather than see the outcome remain at cell 3. Cells 2 and 4 are both equilibria in this game, and they

[9] It is possible that a society would lock in on a strategy A equilibrium, even though $b > a$, so that a limited role for the state in announcing which strategy citizens should coordinate on might still be desirable.

are the only two. The ordering of payoffs in a game of chicken for the row player is cell 2 > 1 > 4 > 3, whereas in a prisoners' dilemma it is 2 > 1 > 3 > 4. The interchange of the last two cells for both players causes the shift in the equilibrium.

In cells 4, 1, and 2, the fence is built. These cells differ only in who pays for the fence and the resulting utility payoffs. In cell 4, G pays the full \$1,000 cost of the fence and experiences a utility level of 2. In cell 1, G pays \$500 and receives a utility level of 3, while in cell 2 G pays nothing for a utility level of 3.5. The lower increment in utility in going from a \$500 fall in income to no change in income, compared with going from a \$1,000 fall in income to a \$500 fall, reflects an assumption of the declining marginal utility of income. If both G and D have declining marginal utilities of income, as assumed in the figures in Matrix 2.3, then the solution that they share the cost of the fence is welfare maximizing as well as equitable. Under alternative assumptions, a stronger, higher fence may be built when the cost is shared, and the result may be an efficiency gain from the cost-sharing solution in cell 1. But the outcome in cell 1 is not an equilibrium. Both D and G will be better off if they can convince the other to pay the full cost of the fence. One way to do this is to precommit oneself not to build the fence, or at least to convince one's neighbor that one has made such a commitment so that the neighbor, say, D, believes that her choice is between cells 2 and 3, and thus naturally chooses cell 2.

The chicken game is often used to depict the interactions of nations (Schelling, 1966, ch. 2). Let D be a superpower, which favors having other countries install democratic institutions, and C a country favoring communist institutions. A civil war rages in small country S between one group seeking to install a communist regime and another group wishing to install a democratic constitution. The situation could easily take on the characteristics of a game of chicken. Each superpower wants to support the group favoring its ideology in S, and wants the other superpower to back down. But if the other superpower, say, C, is supporting its group in S, then D is better off backing off than supporting its group in S and thereby being led into a direct confrontation with the other superpower. Both powers are clearly better off if they both back off than if the confrontation occurs.

Given this game-of-chicken configuration of payoffs, each superpower may try to get the other to back off by precommiting itself to defending democracy (communism) wherever it is threatened around the world. Such a precommitment combined with a reputation for "toughness" could force the other superpower to back down each time a clash between communist and noncommunist forces occurs in a small country.

The danger in this situation, however, is that both superpowers become so committed to their strategy of supporting groups of their ideology, and so committed to preserving their reputations for toughness, that neither side backs down. The confrontation of the superpowers is precipitated by the civil war in S.

As in prisoners' dilemmas, the joint cooperation solution to the chicken game can emerge from a chicken supergame, if each player recognizes the long-run advantages to cooperation and adopts the tit-for-tat supergame strategy or an analogous one (Taylor and Ward, 1982; Ward, 1987). Alternatively, the two superpowers (neighbors) may recognize the dangers inherent in the noncooperative, precommitment strategy and directly approach one another and agree to follow the cooperative strategy. Thus, although the structure of the chicken game differs from that of the prisoners' dilemma, the optimal solutions of the game are similar, requiring some sort of formal or tacit agreement to cooperate. As the number of players increases, the likelihood that a formal agreement is required increases (Taylor and Ward, 1982; Ward, 1987). Thus, for the chicken game, as for the prisoners' dilemma, the need for democratic institutions to achieve the efficient, cooperative solution to the game increases as the number of players rises.

2.4* Voluntary provision of public goods with constant returns to scale

In this section we explore more formally the problems that arise in the voluntary provision of a public good. Consider as the pure public good a levy or dike built of bags of sand. Each member of the community voluntarily supplies as many bags of sand as she chooses. The total number of bags supplied is the summation of the individual contributions of each member. The more bags supplied, the higher and stronger the dike, and the better off are all members of the community. Letting G_i be the contribution to the public good of individual i, then the total quantity of public good supplied is

$$G = G_1 + G_2 + G_3 + \cdots G_n. \tag{2.1}$$

Let each individual's utility function be given as $U_i(X_i, G)$, where X_i is the quantity of private good i consumes.

Now consider the decision of i as to how much of the public good to supply, that is, the optimal G_i, given her budget constraint $Y_i = P_x X_i + P_g G_i$, where Y_i is her income and P_x and P_g are prices of the private and public goods, respectively. In the absence of an institution for coordinating the quantities of public good supplied, each individual must decide independently of the other individuals how much of the public good to supply. In making this decision, it is reasonable to assume that the individual takes the supply of the public good by the rest of the community as fixed. Each i chooses the G_i that maximizes U_i, given the values of G_j chosen by all other individuals j. Individual i's objective function is thus

$$O_i = U_i(X_i, G) + \lambda_i(Y_i - P_x X_i - P_g G_i). \tag{2.2}$$

Maximizing (2.2) with respect to G_i and X_i yields

$$\frac{\partial U_i}{\partial G} - \lambda_i P_g = 0 \tag{2.3}$$

$$\frac{\partial U_i}{\partial X_i} - \lambda_i P_x = 0 \tag{2.4}$$

from which we obtain

$$\frac{\partial U_i/\partial G}{\partial U_i/\partial X_i} = \frac{P_g}{P_x} \tag{2.5}$$

as the condition for utility maximization. Each individual purchases the public good as if it were a private good, taking the purchases of the other members of the community as given. This equilibrium is often referred to as a Cournot or Nash equilibrium, as it resembles the behavioral assumption Cournot made concerning the supply of a homogeneous private good in an oligopolistic market.

Now let us contrast (2.5) with the condition for Pareto optimality. To obtain this, we maximize the following welfare function:

$$W = \gamma_1 U_1 + \gamma_2 U_2 + \cdots + \gamma_n U_n, \tag{2.6}$$

where all $\gamma_i > 0$. Given the positive weights on all individual utilities, any allocation that is not Pareto optimal – that is, from which one person's utility can be increased without lowering anyone else's – cannot be at a maximum for W. Thus, choosing X_i and G_i to maximize W gives us a Pareto-optimal allocation.

Maximizing (2.6) subject to the aggregate budget constraint

$$\sum_{i=1}^{n} Y_i = P_x \sum_{i=1}^{n} X_i + P_g G, \tag{2.7}$$

we obtain the first-order conditions

$$\sum_{i=1}^{n} \gamma_i \frac{\partial U_i}{\partial G} - \lambda P_g = 0 \tag{2.8}$$

and

$$\gamma_i \frac{\partial U_i}{\partial X_i} - \lambda P_x = 0, \qquad i = 1, n, \tag{2.9}$$

where λ is the Lagrangian multiplier on the budget constraint. Using the n equations in (2.9) to eliminate the γ_i in (2.8), we obtain

$$\sum_i \frac{\lambda P_x}{\partial U_i / \partial X_i} \cdot \partial U_i / \partial G = \lambda P_g, \tag{2.10}$$

from which we obtain

$$\sum_i \frac{\partial U_i / \partial G}{\partial U_i / \partial X_i} = \frac{P_g}{P_x}. \tag{2.11}$$

Equation (2.11) is the familiar Samuelsonian (1954) condition for the Pareto-optimal provision of a public good. Independent utility maximization decisions lead each individual to equate her marginal rate of substitution of the public for the private good to their price ratio, as if the public good were a private good (2.5). Pareto optimality, however, requires that the summation of the marginal rates of substitution over all members of the community be equated to this price ratio (2.11).

That the quantity of public good provided under the Cournot-Nash equilibrium (2.5) is less than the Pareto-optimal quantity can be seen by rewriting (2.11) as

$$\frac{\partial U_i/\partial G}{\partial U_i/\partial X_i} = \frac{P_g}{P_x} - \sum_{j \neq i} \frac{\partial U_j/\partial G}{\partial U_j/\partial X_j}. \tag{2.12}$$

If G and X are normal goods in each individual's utility function, then

$$\sum_{j \neq i} \frac{\partial U_j/\partial G}{\partial U_j/\partial X_j} > 0$$

and the marginal rate of substitution of public for private good for individual i defined by (2.12) is less than that defined by (2.5), which implies that a greater quantity of G and a smaller quantity of X_i are being consumed when (2.12) is satisfied than when (2.5) is.

To gain a feel for the quantitative significance of the differences, consider the special case where U_i is a Cobb-Douglas utility function, that is, $U_i = X_i^{\alpha} G^{\beta}$, $0 < \alpha < 1$, and $0 < \beta < 1$. Under this assumption (2.5) becomes

$$\frac{\beta X_i^{\alpha} G^{\beta-1}}{\alpha X_i^{\alpha-1} G^{\beta}} = \frac{P_g}{P_x}, \tag{2.13}$$

from which it follows that

$$G = \frac{P_x}{P_g}\frac{\beta}{\alpha} X_i. \tag{2.14}$$

Substituting from (2.1) and the budget constraint yields

$$\sum_i G_i = \frac{P_x}{P_g}\frac{\beta}{\alpha}\left(\frac{Y_i}{P_x} - \frac{P_g}{P_x} G_i\right), \tag{2.15}$$

from which we obtain

$$\left(1 + \frac{\beta}{\alpha}\right) G_i = -\sum_{j \neq i} G_j + \frac{\beta}{\alpha}\frac{Y_i}{P_g} \tag{2.16}$$

or

$$G_i = -\frac{\alpha}{\alpha + \beta} \sum_{j \neq i} G_j + \frac{\beta}{\alpha + \beta}\frac{Y_i}{P_g}. \tag{2.17}$$

Equation (2.17) implies that individual i voluntarily chooses to supply a smaller amount of the public good, the larger she believes the amount of public good provided by the other citizens to be. With only two individuals in the community, (2.17) defines the familiar reaction curve from duopoly theory. In this situation, it is a negativity-sloped straight line.

If all members of the community have identical incomes, Y, then all will choose the same levels of G_i, and (2.17) can be used to find the contribution in equilibrium of a single individual:

$$G_i = -\frac{\alpha}{\alpha+\beta}(n-1)G_i + \frac{\beta}{\alpha+\beta}\frac{Y}{P_g}, \tag{2.18}$$

from which we obtain

$$G_i = \frac{\beta}{\alpha n+\beta}\frac{Y}{P_g}. \tag{2.19}$$

The amount of the public good provided by the community through independent contributions then becomes

$$G = nG_i = \frac{n\beta}{\alpha n+\beta}\frac{Y}{P_g}. \tag{2.20}$$

These quantities can be compared to the Pareto-optimal quantities. With all individual incomes equal, all individuals contribute the same G_i and have the same X_i left over, so that (2.11) becomes

$$n\frac{\beta X_i^{\alpha} G^{\beta-1}}{\alpha X_i^{\alpha-1} G^{\beta}} = \frac{P_g}{P_x}. \tag{2.21}$$

Using the budget constraint to eliminate the X_i and rearranging yields for the Pareto-optimal contribution of a single individual,

$$G_i = \frac{\beta}{\alpha+\beta}\frac{Y}{P_g} \tag{2.22}$$

and

$$G = nG_i = \frac{n\beta}{\alpha+\beta}\frac{Y}{P_g}. \tag{2.23}$$

Let us call the Pareto-optimal quantity of public good defined by (2.23) G_{PO}, and the quantity under the Cournot-Nash equilibrium (2.20), G_{CN}. Their ratio is then

$$\frac{G_{CN}}{G_{PO}} = \frac{\dfrac{n\beta}{\alpha n+\beta}\dfrac{Y}{P_g}}{\dfrac{n\beta}{\alpha+\beta}\dfrac{Y}{P_g}} = \frac{\alpha+\beta}{\alpha n+\beta}. \tag{2.24}$$

This ratio is less than one, if $n > 1$, and tends toward zero as n becomes increasingly large. Thus, for all communities greater than a solitary individual, voluntary, independent supply of the public good leads to less than the Pareto-optimal quantity being supplied, and the relative gap between the two quantities grows as community size increases.

The extent of underprovision of the public good at a Cournot-Nash equilibrium depends on the nature of the individual utility functions (Cornes and Sandler, 1986, ch. 5). For the Cobb-Douglas utility function, the greater the ratio of β to α, the *smaller* the extent of underprovision. With $\alpha = 0$ – that is, when the marginal utility of the private good is zero – $G_{CN} = G_{PO}$. This equality also holds with right-angled indifference curves, where again the marginal utility of the private good, holding the quantity of the public good fixed, is zero (Cornes and Sandler, 1986, p. 81). But with the familiar, smooth, convex-to-the-origin indifference curves, one can expect an underprovision of a voluntarily provided public good, and an underprovision whose relative size grows with the size of the community. To achieve the Pareto-optimal allocation, some institution for coordinating the contributions of all individuals is needed.

2.5* Voluntary provision of public goods with varying supply technologies

Many public goods might be depicted using the summation technology of the previous section. Public goods of a prisoners' dilemma type – for example, community order, environmental quality – are provided by each individual contributing to the "production" of the public good by not stealing or not polluting. For the typical public good of this kind, the quantity supplied is to some degree additive with respect to each individual's contribution. The more people there are who refrain from stealing, the more secure is the community, and the greater is the welfare of its members.

There are other public goods, however, for which the participation of *all* members is necessary to secure *any* benefits. The crew of a small sailboat, two-man rowboats, and bobsleds are examples. For the rowboat to go in a straight line each rower must pull the oar with equal force. Under- or overcontributions are penalized by the boat's moving in a circle. Only the equal contribution of both rowers is rewarded by the boat's moving forward. With such goods, cells 2, 4, and 3 of Matrix 2.1 collapse into one and cooperative behavior is voluntarily forthcoming.

Goods such as these are produced by what Hirshleifer (1983, 1984) named the "weakest-link" technology. The amount of public good provided is equal to the smallest quantity provided by any member of the community. At the other pole from weakest-link technology one can conceive of a best-shot technology for which the amount of public good provided is equal to the largest quantity provided by any one member of the community. As an example of the best-shot technology, one can think of a community first having each member design a boat (bridge) for crossing a given body of water, and then the best design selected and constructed.

The weakest-link technology is like a fixed coefficient production function for public goods. Individual i's marginal contribution to public good supply, $\partial G/\partial G_i$, is zero, if his contribution exceeds that of any other member of the community ($G_i > G_j$ for some j). But $\partial G/\partial G_i$ equals the community supply function when $G_i < G_j$ for all j. The summation technology assumes an additive and separable production function, whereas the best-shot technology assumes a form of discontinuously increasing returns. The latter seems the least plausible of the three, so we consider only the cases falling in the range between the weakest-link and summation production technologies.

Consider a community of two Australian farmers whose fields are adjacent to another and border on a segment of the bush. Each night the kangaroos come out of the bush and destroy the farmers' crops. The farmers can protect their crops, however, by erecting fences along the border between their property and the bush.

Each farmer is responsible for buying fence for his own segment of the border. The following technologies can be envisaged:

Weakest link: Kangaroos adapt quickly to changes in their environment and discover the lowest point in the fence. The number of kangaroos entering both farmers' fields is determined by the height of the fence at its lowest point.

Unweighted summation: Kangaroos are very dumb and probe the fence at random. The number of kangaroos entering the two fields varies inversely with the average height of the two fences.

Diminishing returns: If one farmer's fence is lower than the other's, some, but not all, kangaroos learn to probe only the lower fence, and the higher fence stops some kangaroos from going over.

Now consider the following general formulation of public good supply: Let G be the number of units of public good provided, defined in this case as the number of kangaroos prevented from entering the fields. Let the units of fence purchased at price P_f be defined so that

$$G = F_1 + wF_2, \qquad 0 \leq F_1 \leq F_2, \qquad 0 \leq w \leq 1, \tag{2.25}$$

where F_i is farmer i's purchase of fence. If $w = 0$, we have the weakest-link case, and $G = F_1$, the smaller of the two contributions. The larger w is, the more 2's contribution beyond 1's contributes to the supply of G, until with $w = 1$, we reach the unweighted summation supply function examined above. To simplify the problem, assume that both farmers have identical utility functions and both G and the private good X are noninferior. Then the farmer with the lower income will always choose to purchase the smaller quantity of fence, so that farmer 1 is the farmer with the smaller income of the two. He maximizes his utility $U_1(X, G)$ by choosing a level of private good consumption X_1 and contribution to the public good F_1 satisfying his budget constraint, $Y_1 = P_x X_1 + P_f F_1$. The solution is again (2.5), with the price of the public good now P_f.

The solution to the utility maximization problem for farmer 2 is, however,

$$\frac{\partial U_2/\partial G}{\partial U_2/\partial X} = \frac{P_f}{wP_x} \tag{2.26}$$

as long as $F_2 > F_1$. In effect, farmer 2 faces a higher relative price for the public good F, since his purchases do not contribute as much on the margin as 1's, owing to the technology defined by (2.25). The smaller w is, the less fence 2 buys (the smaller his optimal contribution to the public good). With small enough w, the solution to (2.26) would require $F_2 < F_1$. But then 2 would be the smaller contributor and his optimal contribution would be defined by (2.5). Since 2 favors a greater contribution than 1, he simply matches 1's contribution if satisfying (2.26) violates $F_2 > F_1$.

To determine the condition for the Pareto-optimal level of G, we choose levels of X_1, X_2, and G to maximize 1's utility, holding 2's utility constant, and satisfying (2.25) and the individual budget constraints; that is, we maximize

$$L = U_1(X_1, G) + \gamma[\bar{U}_2 - U_2(X_2, G)] + \lambda[G - F_1 - wF_2], \tag{2.27}$$

from which it follows that

$$\frac{\partial U_1/\partial G}{\partial U_1/\partial X} + w\frac{\partial U_2/\partial G}{\partial U_2/\partial X} = \frac{P_f}{P_x}. \tag{2.28}$$

Only in the extreme weakest-link case, where $w = 0$, is the condition for Pareto optimality for the community (2.28) satisfied by the two individuals acting independently, for then (2.28) collapses to (2.5), and both farmers purchase the amounts of fence satisfying (2.5).[10] With $w = 1$, on the other hand, we have the unweighted summation supply of public good, and (2.28) becomes (2.11), the Samuelsonian (1954) condition for Pareto optimality, and too little public good is being supplied.

Moreover, the difference between the quantity of public good supplied voluntarily when each farmer acts independently and the Pareto-optimal quantity increases with w. To illustrate this, again let both individuals have identical incomes Y, and identical utility functions $U = X^\alpha G^\beta$. Both then purchase the same quantity of fence F and private good X. From (2.5) and (2.25) we obtain the Cournot-Nash equilibrium quantity of public good supplied through the independent utility-maximizing decisions of the two farmers:

$$G_{CN} = \frac{\beta Y(1+w)}{P_f[\alpha(1+w)+\beta]}. \tag{2.29}$$

In the same way, (2.28) can be used to obtain Pareto-optimal G:

$$G_{PO} = \frac{\beta}{\alpha+\beta}\frac{Y}{P_f}(1+w). \tag{2.30}$$

Dividing (2.29) by (2.30) we obtain the ratio of independently supplied to Pareto-optimal quantities of public good:

$$\frac{G_{CN}}{G_{PO}} = \frac{\alpha+\beta}{\alpha(1+w)+\beta}. \tag{2.31}$$

With $w = 0$, the ratio is one, but it falls as w increases.

With n individuals, (2.28) generalizes to

$$\frac{\partial U_1/\partial G}{\partial U_1/\partial X} + w_2\frac{\partial U_2/\partial G}{\partial U_2/\partial X} + w_3\frac{\partial U_3/\partial G}{\partial U_3/\partial X} + \cdots + w_n\frac{\partial U_n/\partial G}{\partial U_n/\partial X} = \frac{P_f}{P_x} \tag{2.32}$$

and (2.31) generalizes to

$$\frac{G_{CN}}{G_{PO}} = \frac{\alpha+\beta}{\alpha(1+w_2+w_3+\cdots+w_n)+\beta}. \tag{2.33}$$

The gap between the independently provided and Pareto-optimal quantities of public good increases as the number of members of the community increases, and the weights on the additional contributions increase.

[10] This conclusion is contingent on the initial incomes of the two farmers and the implicit constraint that farmer 2 cannot transfer money to 1 or purchase fence for him. With w low enough or Y_2/Y_1 high enough, unconstrained Pareto optimality may require that 2 subsidize 1's purchase of fence. See Hirshleifer (1984).

Experiments by Harrison and Hirshleifer (1986) with two players indicate that individuals will voluntarily provide nearly the Pareto-optimal quantity of public good in weakest-link ($w = 0$) situations, but underprovide in summation and best-shot situations. Experimental results by van de Kragt, Orbell, and Dawes (1983) with small groups also indicate that efficient public good provision is forthcoming in situations resembling the weakest-link technology. Thus, voluntary provision of public goods without coordination or coercion at Pareto-optimal levels is possible when the technology of public good provision conforms to the weakest-link condition. Unfortunately, with large communities it is difficult to think of many public goods for which voluntary provision is feasible, and all w_i for contributions greater than the minimum are zero or close to it. In large communities, therefore, some institutional mechanism for coordinating and coercing individual contributions to the supply of public goods seems likely to be needed.

2.6 Externalities

Public goods are a classic example of the kinds of market failures economists cite as justification for government intervention. Externalities are the second primary category of market failure. An externality occurs when the consumption or production activity of one individual or firm has an *unintended* impact on the utility or production function of another individual or firm. Individual A plants a tree to provide herself shade, but inadvertently blocks her neighbors' view of the valley. The pulp mill discharges waste into the river and inadvertently raises the costs of production for the brewery downstream. These activities may be contrasted with normal market transactions in which A's action, say, buying the tree, has an impact on B, the seller of the tree, but the impact is fully accounted for through the operation of the price system. There is no market for the view of the valley or the quality of water in the river, and thus no price mechanism for coordinating individual actions. Given the existence of externalities, a non-Pareto-optimal allocation of resources often results.

To see the problem more clearly, let us consider a situation in which two individuals each consume private good X, and A consumes externality creating good E. Individual A then purchases X and E so as to maximize her utility subject to the budget constraint, $Y_A = X_A P_x + E_A P_e$; that is, A maximizes

$$L = U_A(X_A, E_A) + \lambda(Y_A - X_A P_x - E_A P_e). \tag{2.34}$$

Maximization of (2.34) with respect to X and E yields the familiar first-order condition for individual utility maximization when there are two private goods:

$$\frac{\partial U_A/\partial E}{\partial U_A/\partial X} = \frac{P_e}{P_x}. \tag{2.35}$$

But E is an activity that produces an externality and thus enters B's utility function also, even though B does not buy or sell E. We can solve for the Pareto-optimal allocation of X and E by maximizing one individual's utility, subject to the constraints that the other individual's utility is held constant, and the combined budget of the two individuals is not exceeded.

$$L_{PO} = U_A(X_A, E_A) + \lambda(\bar{U}_B - U_B(X_B, E_A)) \\ + \gamma(Y_A + Y_B - P_x X_A - P_x X_B - P_e E_A). \quad (2.36)$$

The presence of A's consumption of E, E_A, in B's utility function represents the externality nature of the E activity. Maximizing (2.36) with respect to X_A, X_B, and E_A yields

$$\frac{\partial L_{PO}}{\partial X_A} = \frac{\partial U_A}{\partial X} - \gamma P_x = 0, \quad (2.37)$$

$$\frac{\partial L_{PO}}{\partial X_B} = \lambda\left(-\frac{\partial U_B}{\partial X}\right) - \gamma P_x = 0, \quad (2.38)$$

$$\frac{\partial L_{PO}}{\partial E_A} = \frac{\partial U_A}{\partial E} - \lambda \frac{\partial U_B}{\partial E} - \gamma P_e = 0. \quad (2.39)$$

Using (2.37) and (2.38) to eliminate λ and γ from (2.39), we obtain as the condition for Pareto optimality

$$\frac{\partial U_A/\partial E}{\partial U_A/\partial X} + \frac{\partial U_B/\partial E}{\partial U_B/\partial X} = \frac{P_e}{P_x} \quad (2.40)$$

or

$$\frac{\partial U_A/\partial E}{\partial U_A/\partial X} = \frac{P_e}{P_x} - \frac{\partial U_B/\partial E}{\partial U_B/\partial X}. \quad (2.41)$$

Equation (2.41) gives the condition for Pareto optimality; (2.35), the condition for individual A's optimal allocation of her budget. Equation (2.35) governs the determination of the level of E, since only A decides how much E is purchased. If activity E creates a positive externality,

$$\frac{\partial U_B/\partial E}{\partial U_B/\partial X} > 0,$$

then

$$\frac{\partial U_A/\partial E}{\partial U_A/\partial X}$$

is larger than is required for Pareto optimality. A purchases too little E (and too much X) when E produces a positive external economy. Conversely, when E generates a negative externality,

$$\frac{\partial U_B/\partial E}{\partial U_B/\partial X} < 0,$$

and A buys too much of E.

Although seemingly a separate category of market failure, the Pareto-optimality condition for an externality is identical to that for a pure public good, as a comparison of (2.40) and (2.11) reveals (Buchanan and Stubblebine, 1962). The difference between a pure public good and an externality is that in the case of a public good all members of the community consume the *same* good, whereas for an externality the good (bad) consumed by the second parties may differ from that consumed by the direct purchaser. When A contributes to the purchase of flowers for the town square, she helps finance a public good. When A plants flowers in her backyard, she creates a positive externality for those neighbors who can see and enjoy them. If some of A's neighbors are allergic to pollen from the flowers in her backyard, A's plantings create a negative externality. What is crucial to the issue of Pareto optimality is not that A and B consume precisely the same good, but that A's consumption alters B's utility in a manner not accounted for through the price system. B is not excluded from the side effects of A's consumption, and it is this nonexcludability condition that joins public goods and externalities by one and the same Pareto-optimality condition. It is this nonexcludability condition that necessitates some coordination of A and B's activities to achieve Pareto optimality.

One way to adjust A's consumption of E to bring about Pareto optimality is for the government to levy a tax or offer a subsidy to the E activity. If, for example, E generates a negative externality, a tax on E equal to

$$-\frac{\partial U_B/\partial E}{\partial U_B/\partial X}$$

raises the price of E relative to X by precisely the amount necessary to achieve Pareto optimality. Alternatively, a subsidy to A for each unit of E she consumes, less than the amount implied by (2.35), achieves the same effect. The existence of a government to correct for externalities by levying taxes and offering subsidies is a traditional explanation for government intervention most frequently associated with the name of Pigou (1920).

In most discussions of Pigouvian taxes, the government is assumed to "know" the marginal rates of substitution of the different parties generating and affected by the externalities. Often the government is referred to as an individual, the policymaker, who possesses all of the information relevant to determine the Pareto-optimal allocation of resources and who then announces the optimal taxes and subsidies. But where does the policymaker obtain this information? In some situations – for example, when one factory's activities affect the costs of another – one might think of the government policymaker as gathering engineering data and using these to make a decision. But when individual utilities are affected, the engineer's information-gathering problem is greatly complicated. Much of this book is concerned with describing how democratic institutions reveal information concerning individual preferences on externality-type decisions. The next section discusses a more direct approach to the question.

2.7 The Coase theorem

Ronald Coase, in a classic article published in 1960, challenged the conventional wisdom in economics regarding externalities, taxes, and subsidies. Coase argued that the existence of an external effect associated with a given activity did not inevitably require government intervention in the form of taxes and subsidies. Pareto-optimal resolutions of externality situations could be and often were worked out between the affected parties without the help of the government. Moreover, the nature of the outcome was independent of the assignment of property rights, that is, in the case of a negative externality associated with E whether the law granted the purchaser of E the right to purchase E in unlimited quantities, or the law granted B the right to be protected from any adverse effects from A's consumption of E.

Although Coase develops his argument by example, and neither states nor proves any theorems, the main results of the paper are commonly referred to as the Coase theorem. The theorem can be expressed as follows:

The Coase theorem: *In the absence of transaction and bargaining costs, affected parties to an externality will agree on an allocation of resources that is both Pareto optimal and independent of any prior assignment of property rights.*

Pigou was wrong; government intervention is not needed to resolve externality issues.

Consider first a discrete case of the theorem. Let A be a factory producing widgets with a by-product of smoke. Let C be a laundry whose costs are raised by A's emissions of smoke. Given that A is in business, C's profits are \$24,000, but if A were to cease production altogether, C's profits would rise to \$31,000. A's profits are \$3,000. Assuming A's factors of production can be costlessly redeployed, society is better off if A ceases production. C then earns a net surplus over costs of \$31,000, while the combined surplus when both A and C operate is only \$27,000.

But suppose that there are no laws prohibiting smoke emissions. A is then free to produce, and the socially inferior outcome would appear to ensue. It would, however, pay C to bribe the owners of A to cease production by promising to pay them \$3,000 per annum. Alternatively, C could acquire A and close it down. If i is the cost of capital, and the market expects A to earn \$3,000 profits per year in perpetuity, then the market value of A is $\$3{,}000/i$. The present discounted value to C of shutting down A is $\$7{,}000/i$, however. The owners of C realize an increase in wealth of $\$4{,}000/i$ by acquiring and closing A.

To see that the socially efficient outcome arises regardless of the assignment of property rights, assume that A's annual profit is \$10,000 and the figures for C are as before. Now the efficient solution requires that A continue to operate. Suppose, however, that the property rights lie with C. Strict air pollution laws exist and C can file a complaint against A and force it to cease production. However, the profits of A are now such that A can offer C a bribe of $\$7{,}000 + \alpha$, $0 \le \alpha \le \$3{,}000$, not to file a complaint. The owners of both firms are as well or better off under this alternative than they are if A closes, and the socially efficient outcome can again be expected to occur.

Note that under the conditions of the first example, where A's profits were only \$3,000, it would not pay A to bribe C to allow it to continue to produce, and the socially efficient outcome would again occur.

When the externality-producing activity has a variable effect on the second party as the level of the activity changes, the Coase theorem still holds. If A's marginal rate of substitution of E for X (MRS^A_{EX}) falls as E increases, then $MRS^A_{EX} - P_e/P_x$ is negative sloped, as in Figure 2.1. The point where $MRS^A_{EX} - P_e/P_x$ crosses the

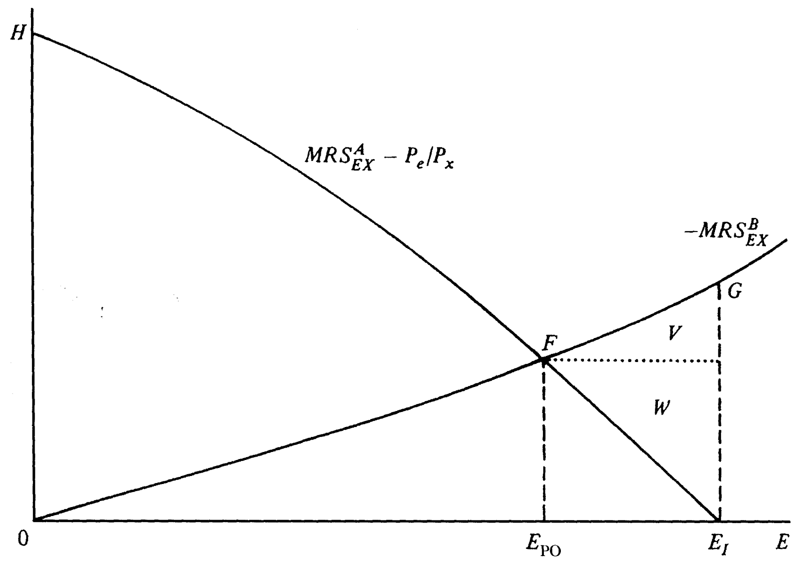

Figure 2.1. Pareto-optimal quantity of a good with external effects.

horizontal axis, E_1, is the level of E that A chooses when she acts independently of B. It is the level of E satisfying (2.35).

If E creates a negative externality on B, then $-MRS^B_{EX}$ is positive. In Figure 2.1, $-MRS^B_{EX}$ is drawn under the reasonable assumption that B is willing to give up an increasing amount of X to prevent A from consuming another unit of E, the higher E is. E_{PO} is the Pareto-optimal level of E, the level satisfying (2.41).

The area $E_{PO}FGE_1$ measures the utility loss to B from A's consumption of E_1 instead of E_{PO}. $E_{PO}FE_1$ measures A's utility gain from these extra units of E. Both B and A are made better off if A accepts a bribe of Z from B to consume E_{PO} rather than E_1, where $E_{PO}FE_1 < Z < E_{PO}FGE_1$. In particular, if B were to offer A a bribe of $E_{PO}F$ for each unit of E she refrained from consuming, A would choose to consume exactly E_{PO} units of E, and A would be better off by the area W and B by the area V as against the independent action outcome at E_1.

With the property rights reversed, B could forbid A from consuming E and force the outcome at 0. But then A would be foregoing $OHFE_{PO}$ benefits, while B gains only OFE_{PO}, as opposed to the Pareto-optimal allocation E_{PO}. Self-interest would lead A to propose and B to accept a bribe Z', to allow A to consume E_{PO}, where $OFE_{PO} < Z' < OHFE_{PO}$.[11]

Coase demonstrated his theorem with four examples drawn from actual cases. Several experiments have been run in which student subjects are given payoff tables that resemble those one would observe in an externality situation. Pareto-optimal

[11] For the quantity of E bought to be the same, whether A receives or pays the bribe, no income effects must be present. When they exist, precise solutions require the use of compensated demand functions (Buchanan and Stubblebine, 1962).

I also abstract from the difficulty of people moving close to a negative externality to receive a bribe, as discussed by Baumol (1972).

outcomes are observed in well over 90 percent of the experiments.[12] The Coase theorem offers a logical and empirically relevant alternative to government action in externality situations. But does it hold up as the number of parties involved in the externality increases? We now address this question.

2.8 Coase and the core

The examples presented by Coase and those discussed above involve but two parties. Does the theorem hold when more than two parties are involved? Hoffman and Spitzer (1986) present experimental results in which Pareto-optimal allocations are achieved in Coasian bargains among as many as 38 parties. But Aivazian and Callen (1981) present an example in which the theorem breaks down with only 3 parties. Let us consider their example.

They deal with a factory, A, producing smoke and a laundry, C, as in our previous example. Representing company profits using the characteristic function notation of game theory, we can restate the previous example as having the following attributes: $V(A) = \$3{,}000$, $V(C) = \$24{,}000$, and $V(A, C) = \$31{,}000$, where $V(A, C)$ is a coalition between A and C, that is, a merger of A and C that results in A's ceasing production.

Now assume the existence of a second factory, B, producing smoke. Let the characteristic functions for this problem be defined as follows:

$$V(A) = \$3{,}000 \qquad V(B) = \$8{,}000 \qquad V(C) = \$24{,}000$$
$$V(A, B) = \$15{,}000 \qquad V(A, C) = \$31{,}000 \qquad V(B, C) = \$36{,}000$$
$$V(A, B, C) = \$40{,}000$$

The Pareto-optimal outcome is the grand coalition $V(A, B, C)$; that is, A and B cease production. If the property right lies with C, the Pareto outcome occurs, C forbids A and B to produce, and neither a coalition between A and B ($V[A, B] = \$15{,}000$) nor the two firms independently ($3,000 + $8,000) can offer C a large-enough bribe to offset its $16,000 gain from going from $V(C)$ to $V(A, B, C)$.

Suppose, however, that A and B have the right to emit smoke. C offers A and B $3,000 and $8,000, respectively, to cease production. Such a proposal can be blocked by A offering to form a coalition with B and share $V(A, B) = \$15{,}000$ with allocations, say, of $X_A = \$6{,}500$, $X_B = \$8{,}500$. But C in turn can block a coalition between A and B by proposing a coalition between itself and B, with, say, $X_B = \$9{,}000$ and $X_C = \$27{,}000$. But this allocation can also be blocked.

To prove generally that the grand coalition is unstable, we show that it is not within the *core*. Basically, a grand coalition is within the core if no subset of the coalition can form, including an individual acting independently, and provide its members higher payoffs than they can obtain in the grand coalition. If (X_A, X_B, X_C) is an

[12] See Hoffman and Spitzer (1982, 1986); Harrison and McKee (1985); and Coursey, Hoffman, and Spitzer (1987).

allocation in the core, then it must satisfy conditions (2.42), (2.43), and (2.44):

$$X_A + X_B + X_C = V(A, B, C) \tag{2.42}$$

$$X_A \geq V(A), X_B \geq V(B), X_C \geq V(C) \tag{2.43}$$

$$X_A + X_B \geq V(A, B), X_A + X_C \geq V(A, C), X_B + X_C \geq V(B, C). \tag{2.44}$$

Condition (2.44) implies that

$$X_A + X_B + X_C \geq \frac{1}{2}[V(A, B) + V(A, C) + V(B, C)], \tag{2.45}$$

which from (2.42) implies that

$$V(A, B, C) \geq \frac{1}{2}[V(A, B) + V(A, C) + V(B, C)]. \tag{2.46}$$

But the numbers of the example contradict (2.46):

$$\$40{,}000 < \frac{1}{2}(\$15{,}000 + \$31{,}000 + \$36{,}000) = \$41{,}000.$$

The grand coalition is not in the core.

The primary issue in the present example is the externality of smoke caused by factories A and B imposed upon the laundry C. That there are gains from internalizing this externality is represented by the assumptions that

$$V(A, C) > V(A) + V(C) \tag{2.47}$$

$$V(B, C) > V(B) + V(C) \tag{2.48}$$

$$V(A, B, C) > V(A) + V(B, C) \tag{2.49}$$

$$V(A, B, C) > V(B) + V(A, C). \tag{2.50}$$

In their example, Aivazian and Callen also make the assumption that an externality exists between the two smoking factories; that is, there are gains to their forming a coalition independent of the laundry C:

$$V(A, B) > V(A) + V(B). \tag{2.51}$$

Now this is clearly a separate externality from that involving C and either or both of the two factories. Aivazian and Callen (p. 177) assume the existence of an economy of scale between A and B. But the existence of this second externality is crucial to the proof that no core exists. Combining (2.49) and (2.50) we obtain

$$V(A, B, C) > \frac{1}{2}[V(A) + V(B) + V(B, C) + V(A, C)]. \tag{2.52}$$

If now $V(A, B) \leq V(A) + V(B)$ – that is, there are no economies to forming the A, B coalition – then

$$V(A, B, C) > \frac{1}{2}[V(A, B) + V(B, C) + V(A, C)] \tag{2.53}$$

and condition (2.46) is satisfied. The grand coalition is now in the core. Aivazian and Callen's demonstration that no core exists when property rights are assigned to the factories comes about not simply because a third player has been added to the game, but because a second externality has also been added, namely, the gain from combining A and B. Moreover, the absence of the core hinges on the requirement that both externalities be eliminated simultaneously with the help of but one liability rule.

To what extent does this example weaken Coase's theorem? As long as we are concerned with eliminating the inefficiency caused by a single externality, I do not think that the example has much relevance. Suppose, for example, that the property rights are with A and B, but that the law allows C to close them if it pays just compensation. C offers the owners of A and B \$3,000 and \$8,000 per annum in perpetuity if they cease to operate. They refuse, demanding \$15,000. If the matter were to go to court, should the court consider an argument for awarding \$15,000 on the grounds that A and B could earn that much if they continued to operate *and if they decided to merge*? I doubt that any court would entertain such an argument. Nevertheless, by including the value of the coalition between A and B in the examination of the existence of the core, we have given legitimacy to a threat by A and B to merge and eliminate one externality as a hindrance to the formation of a coalition among C, A, and B to eliminate another. Conceptually, it seems preferable to assume that either A and B definitely will merge, absent agreement with C, or they will not. If they will, negotiation is between C and the coalition A, B and the Coase theorem holds, since $V(A, B, C) > V(C) + V(A, B)$. If A and B will not merge, (2.52) is the relevant condition for determining the existence of the core, and the theorem again holds.[13]

2.9 A generalization of the Coase theorem

The Coase theorem breaks down in Aivazian and Callen's example, because no stable coalition can form among the three actors. If firm C approaches A and proposes that they form a coalition that would increase both firms' profits, B steps forward and makes A a better offer. But this coalition is also vulnerable to a counteroffer from C. This form of *cycling* from one possible outcome to another will pop up throughout the book. It arises because each actor can *unilaterally* break any "agreement" and accept a better offer.

Bernholz (1997a, 1998) has proposed to rescue the Coase theorem, therefore, by restricting an individual's freedom to break a contract once made. Specifically, Bernholz requires that all *external* contracts and all *internal* contracts be *binding*, meaning that a contract, once made, can only be broken if *all* parties agree to break it. An example of an external contract would be an agreement between firms A and

[13] The combined market values of A and B must lie between \$11,000/i, the value the market places on the firms if it assigns a zero probability to their merging (\$3,000/i + \$8,000/i), and \$15,000/i, the value of a merged firm. Thus, the option of C buying A and B and forming the grand coalition through merger must exist if ownership claims to A and B are for sale. Thus, in the spirit of the Coase theorem, individual actions and the market for firms can optimally eliminate the externality without government intervention.

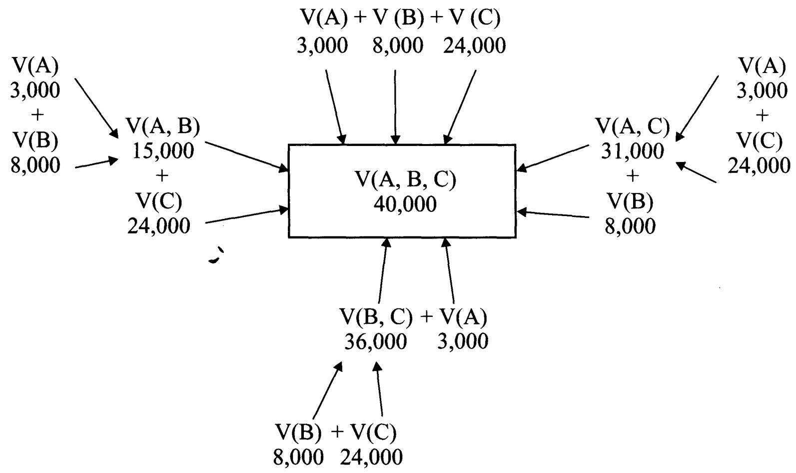

Figure 2.2. Alternative paths to the grand coalition.

C to merge and form a new firm. Once this contract has been signed, the requirement that all internal contracts be binding implies that *A* could accept an offer to merge with *B* only if *C* agreed. Since *C* is worse off playing the game alone, *C* would only agree to let *A* break and join with *B* if *A* and *B* offered *C* a compensating bribe. But the gain to *A* and *B* from forming a coalition is not sufficient to compensate *C* for its loss if *C* breaks with *A*, and thus *C* will never agree to allow *A* to merge with *B*. Once *A* and *C* have agreed to merge, the only new agreement possible is one to form the grand coalition, and it will be forthcoming, since it can lead to an improvement in the positions of all parties. Thus, when all internal and external contracts are binding, one of the four sequences of moves depicted in Figure 2.2 must take place. Either the three firms form the grand coalition immediately, or a pair of them merge, and then this pair goes on to merge with the remaining third company.

Given the presence of well-defined property rights and the absence of transaction costs, Bernholz (1997a, 1998) proves that the existence of binding internal and external contracts suffices to ensure that the Pareto frontier is reached. Starting from a state of anarchy, rational self-interested individuals could and would join a series of contracts that would carry them out to the Pareto frontier. No cycling problems of the type posed by Aivazian and Callen would arise, nor of the types discussed later in this book.[14] In a world of zero transaction costs, the state's only role would be to define the initial set of property rights and enforce all contracts to ensure that they are, indeed, binding. Coase's initial insight – that two rational individuals would, in the absence of transaction costs, contract to resolve a conflict over an externality in a way that achieves Pareto optimality – can be generalized to *all* individuals contracting to resolve *all* collective action problems optimally. (Bernholz's theorem does not, of course, overturn the demonstration that no core exists in the example of three factories, as well as in a much broader set of examples.

[14] Bernholz makes some additional assumptions, but the key assumptions for the proof are those of zero transaction costs and binding contracts.

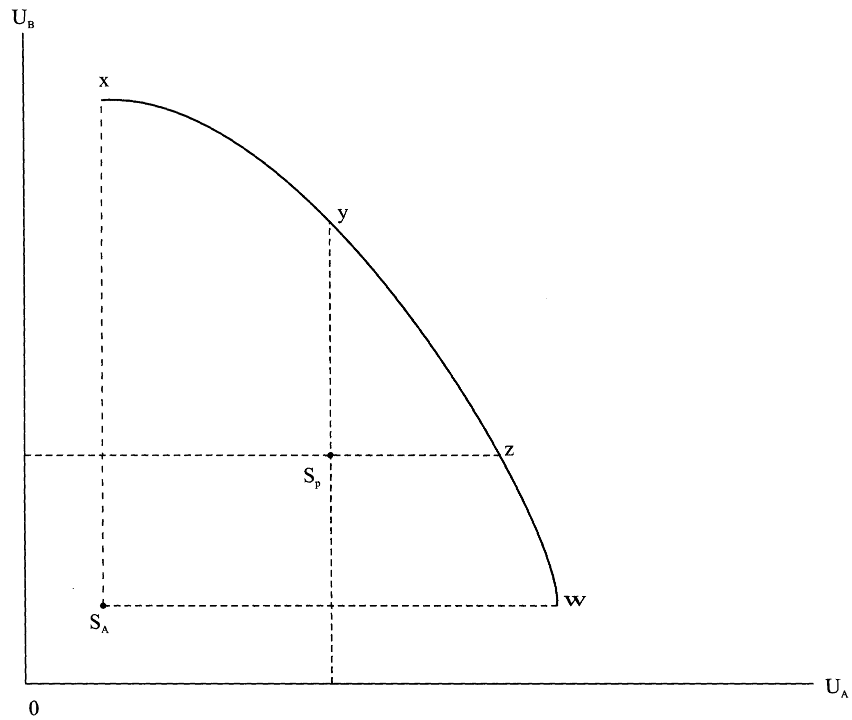

Figure 2.3. Utility possibilities in presence of an externality.

Thus, the possibility cannot be ruled out that a Pareto-optimal set of contracts is never achieved. Just as Buridan's ass stood paralyzed unable to choose between two equidistant stacks of hay, individuals faced with several contractual options, each of which would improve their welfare, may be unable to choose any one, and thus fail to join any. Although a logical possibility, for individuals who are more rational than Buridan's ass, one expects that they would eventually join one advantageous contract and then move on to others as they march toward the Pareto frontier.)

2.10 Does the Coase theorem hold without predefined property rights?

In our statement of the Coase theorem, the Pareto-optimal allocation is reached *independent of any initial assignment of property rights*. What happens, however, if there is no initial assignment of property rights? Does the Coase theorem still hold?

To see what is involved, consider Figure 2.3. A undertakes activity E which creates an externality that harms B, as discussed in the example involving Figure 2.1. The initial assignment of property rights favors A. S_P represents the levels of utility that A and B experience when A purchases E without regard for B (E_I in Figure 2.1). The minimum bribe that A will accept to achieve the Pareto-optimal outcome equals the triangular area under her demand schedule between E_I and E_{PO}.

If B pays only this minimum bribe his utility increases by the equivalent of $W + V$ in Figure 2.1, and the outcome shifts from S_P to y. If, on the other hand, all of the gains from reducing the level of E go to A, the outcome shifts from S_P to z. The curve connecting points y and z represents all of the combinations of utility that A and B can attain by reducing A's consumption of E to its Pareto-optimal level. The Coase theorem states that in the absence of transaction costs some point between y and z is attained.

What happens, however, if there are no assigned property rights? Presumably A will want to consume E_I. B will want to prevent A from consuming any E. To do so, B might buy a gun or hire a thug to intimidate A. Violence might ensue. Without assigned property rights A and B are thrust back into anarchy and additional resources might be wasted in the struggle to determine how much E, if any, A will be able to consume. The status quo under anarchy shifts back from S_P to S_A.

But if there are zero transaction costs, A and B will not stay at S_A; they will agree to move costlessly to some point on $y - z$. If by zero transaction costs we mean zero *bargaining costs*, then rational self-interested individuals will never expend resources to resolve conflicts, since these conflicts can always be resolved at no cost to both parties' advantage. A and B move instantaneously from S_A to $y - z$.

Such an interpretation of the zero transaction costs assumption both trivializes it and converts the Coase theorem into a tautology, which merely states that rational people will never pass up opportunities to make themselves better off at zero cost.[15]

At the same time, however, the argument helps illustrate just how important are the assumptions that we make about transaction costs, and it gives additional insight as to why property rights are valuable. The range of utility combinations that make both A and B better off is much greater when they are bargaining from point S_A than from point S_P. Thus, the stakes involved in the bargaining are much greater at S_A than at S_P. In the real world, where bargains are not costlessly consummated, it might be easier for A and B to strike a bargain if they start from point S_P, since the stakes are much smaller there. This in turn explains why individuals might choose from a state of anarchy like that represented by S_A to define property rights. Such rights may reduce future transaction and bargaining costs.[16]

2.11 Externalities with large numbers of individuals

The Coase theorem implies that when transaction costs are zero, all collective choices that promise a Pareto improvement are made. No public good with benefits greater than costs goes unprovided; no Pareto-relevant external effect is left unaltered; no firm that would make a profit fails to get started, no matter how large the number of participants needed to bring about the optimal collective choice.

In the next section, we shall indicate why the zero transaction costs assumption becomes increasingly implausible as the number of participants in a collective action increases. Now, however, we consider an argument that the Coase theorem is

[15] See Mueller (1991) and Usher (1998).
[16] See, again, Mueller (1991). We return to the issue of why rights might be defined in Chapters 26 and 27.

"undermined" by increasing numbers of participants, *even* when transaction costs remain zero.[17]

We have already demonstrated this proposition in Sections 2.4 and 2.5 for the case of voluntary individual contributions to a public good taking the contributions of all other individuals as given. Except in the case of the most extreme weakest-link technology, the quantity of the public good provided as a percentage of the Pareto-optimal amount becomes vanishingly small as the number of contributors increases.

Consider now a slightly different example involving a discrete public good that would seem to make the attainment of Pareto optimality through voluntary action more likely.[18] A dike that will forever protect a community from flooding can be built at a cost of C. Each of the N members of the community has identical tastes and income and would experience a utility gain of V if the dike were built. Obviously, the dike should be built if $NV > C$. But a collective decision must be made to provide this public good. A meeting is called to which all N members of the community are invited. Each person is free to attend or not. Those attending can decide whether to provide the public good and share its costs amongst themselves, or not. Absent an institution like the state that can *compel* contributions, however, those who do not attend the meeting cannot be forced to contribute to the public good's costs.

Given the zero transaction (bargaining) costs assumption, we can assume that the n individuals who show up at the meeting choose to build the dike, if $nV > C$, and, let us say, they decide to share its costs equally. Knowing this, each individual must decide whether to attend the meeting. With all individuals identical, it is reasonable to confine our attention to symmetric strategy choices. There are only two *pure* strategy choices – to participate or to abstain – and thus only two possible, symmetric Nash equilibria in pure strategies – one where all participate and one where all abstain. Let M be the minimum number of participants that suffices for the dike to be built, $(M-1)V < C < MV$. Then participation is a symmetric, Nash equilibrium if and only if $M = N$. With $M < N$ and all other persons participating, an individual is better off abstaining and free-riding on the provision of the public good by the rest of the community. The case $M = N$ corresponds to the extreme form of weakest-link technology described in Section 2.4, and again produces the Pareto-optimal quantity of the public good with voluntary participation.

Abstention is a symmetric Nash equilibrium for any M above one. If two or more individuals must participate for the dike to be built, and all other $(N-1)$ individuals

[17] We follow the development of the argument by Dixit and Olson (2000). See also, however, Palfrey and Rosenthal (1984).

[18] Voluntary contributions should be more likely with discrete public goods, because *no* public good is provided at all unless the total amount contributed exceeds the lump sum cost of the public good – referred to in the experimental literature as the "provision point." Although the existence of a provision point by itself does not seem to mitigate free-rider behavior in public goods experiments (Isaac, Schmidtz, and Walker, 1989; Asch, Gigliotti, and Polito, 1993), Isaac, Schmidtz, and Walker (1989) and Bagnoli and McKee (1991) do find significantly higher voluntary contributions in experiments that include both provision points and a give-back option. In these experiments an individual only "loses his contribution" if the provision point is reached and the public good is provided. This combination of a provision point and a give-back option characterizes the following example, and thus we would expect from these experiments that the participants at the meetings would decide whether the public good is provided would contribute the required amount.

are abstaining, there is no reason for the Nth individual not to abstain also. With even modestly large Ns, the number of situations in which $M \geq 2$ is likely to be far greater than the number satisfying $M = N$. Thus, if pure strategy equilibria were to emerge, they would most likely involve all members of the community abstaining.

Recognizing this, our sophisticated resident might choose to adopt a *mixed* strategy, that is, to participate with probability P, $0 < P < 1$, and to abstain with probability $(1 - P)$. This way, if all persons choose the same P, there is at least a positive probability that the public good is provided. Of course, there must then also be a positive probability that the public good is *not* provided, and this alone undermines the Coase theorem to a degree.

Consider now the decision of Tip, a typical member of the community. If Tip participates, and the public good is provided, his net benefits are $(V - C/n)$ with n participants. His expected benefits if he participates are then the probability that the public good is provided, that is, the probability that $n \geq M \times (V - C/n)$.

$$\sum_{n=M}^{N} \frac{(N-1)!}{(n-1)!((N-1)-(n-1))!} P^{n-1}(1-P)^{(N-1)-(n-1)} \left[V - \frac{C}{n} \right]. \tag{2.54}$$

The expected benefit from abstention is V times the probability that the public good is provided even when he abstains:

$$\sum_{n=M}^{N-1} \frac{(N-1)!}{n!(N-1-n)!} P^{n}(1-P)^{N-1-n}\ V. \tag{2.55}$$

Whenever $n > M$, the public good would have been provided without Tip's participation, and he loses C/n. He experiences a *net* gain by participating only when his participation raises n to equality with M, an invent whose probability falls as N increases, holding M/N constant. Dixit and Olson (2000) calculate P, and the cumulative probability that enough people participate so that the public good is provided, π, for various values of C, M, and N, holding V fixed at 1.0. A few of their calculations are reproduced in Table 2.1.

When one person's participation is decisive, $C/M < V < C/(M+1)$. The size of the gain from this person's participation $(V - C/M)$ is then the crucial number to induce participation. Thus, seemingly small changes in C can have big effects on P and π. With $M = 10$ and $N = 20$, the probability of an individual's participating falls from 0.091 to 0.011 as C goes from 9.1 to 9.9. But even in the case where $P = 0.091$, the probability that 10 or more people choose to participate is a mere 0.0000032. Even this probability looks large compared to the other entries in the table. Only for very small communities are the probabilities of participation, and that the public good is provided, reasonably high. (If $V = 1.0$, $C = 1.5$, $M = 2$, and $N = 6$, then $P = 0.176$ and $\pi = 0.285$.)

What would happen if someone called a meeting to provide a pure public good and no one came? Obviously, the public good would not be provided. But equally as obvious – if there are zero transaction costs – it would pay to call another meeting. Surely, if the public good failed to be provided at the first meeting, individuals would

Table 2.1. *Optimal participation probabilities, P, and public good provision probabilities, π, when participation is voluntary*

	$V = 1.0$					
	$C = 9.1$		$C = 9.5$		$C = 9.9$	
N	P	π	P	π	P	π
			$M = 10$			
20	.091	$.32 \times 10^{-5}$	.053	$.18 \times 10^{-7}$	.011	$.40 \times 10^{-14}$
30	.048	$.76 \times 10^{-6}$	.027	$.37 \times 10^{-8}$	.005	$.66 \times 10^{-15}$
40	.032	$.43 \times 10^{-6}$	.018	$.20 \times 10^{-8}$	.004	$.33 \times 10^{-15}$
80	.014	$.20 \times 10^{-6}$	.008	$.87 \times 10^{-9}$	.002	$.14 \times 10^{-15}$
160	.007	$.15 \times 10^{-6}$	.004	$.61 \times 10^{-9}$	.001	$.94 \times 10^{-16}$
			$M = 50$			
60	.084	$.60 \times 10^{-43}$	.049	$.97 \times 10^{-55}$	.010	$.11 \times 10^{-88}$
100	.018	$.27 \times 10^{-58}$	.010	$.10 \times 10^{-70}$	.002	$.26 \times 10^{-105}$
150	.009	$.74 \times 10^{-62}$	.005	$.23 \times 10^{-74}$	.001[a]	$.48 \times 10^{-109}$
200	.006	$.30 \times 10^{-63}$	.003[a]	$.88 \times 10^{-76}$	.001[a]	$.17 \times 10^{-110}$
250	.005	$.56 \times 10^{-64}$	.003[a]	$.16 \times 10^{-76}$	.001[a]	$.29 \times 10^{-111}$

[a] These numbers differ from the identical numbers in this column when written to four decimal places.
Source: Dixit and Olson (2000, Tables 1 and 3).

reevaluate their decisions to abstain, and show up at the second meeting, or the third, or the fourth. Alas, quite to the contrary. If more meetings were held, a rational, self-interested individual would be encouraged to *lower* his P and take a chance that enough people to provide the good show up at a meeting before he does.[19]

To ensure that the public good is provided in a reasonable amount of time, it is necessary to both call a meeting *and announce* that the public good will be provided only in the event that all N members of the community participate. The "threat" of not providing the public good if $M \leq n < N$ is credible, so long as there are no costs to calling another meeting, since in a meeting where $n < N$, all participants gain by adjourning and waiting until $n = N$. Knowing that the public good will only be provided when everyone attends the meeting, each person might as well attend the first meeting called. The Coase theorem is reconfirmed under the proviso that some agent (the state?) both calls a meeting of all community members and announces that the community will only reach a positive decision if all members participate.

We are thus forced to qualify the implications of the generalized Coase theorem discussed in Section 2.9. The requirement of binding external and internal contracts may not suffice to ensure that all Pareto-preferred contracts actually are written. When a nonexcludable public good is involved, it may be necessary to require that

[19] Of course, $\pi > 0$ if $P > 0$. Thus, as long as P does not go to zero, the possibility remains that the public good is provided, even if π becomes infinitesimal. If the zero transaction costs assumption is interpreted as implying that an infinitely large number of meetings could be called in an infinitely short period of time, then the Coase theorem is reconfirmed.

all members of the community participate in the writing of the binding contract to provide it.[20]

2.12 Externalities with large numbers of individuals – a second time

Several years ago residents of the community of Shangrila unanimously voted to tax themselves to pay for a dike that would protect them from floods. At that time they formed the Preservation of Shangrila from Floods Club (PSFC). The PSFC meets once a year to decide on the taxes needed to maintain the dike.

As Shangrila has grown and prospered, a second problem has arisen. The number of autos has grown so large that the air of Shangrila has become polluted. Jane, a jogger who owns a bike but no car, surmises that there are many like herself who would be willing to tax themselves to offer all automobile drivers a bribe to reduce the pollution from their cars. She decides to form a club – the Preservation of Shangrila from Pollution Club (PSPC). Consider now the task confronting Jane. She must first approach all of those who, like herself, desire cleaner air, and ask them to attend a meeting to form the PSPC. If they have read the previous section, some may choose not to attend this meeting in the hope that the meeting will agree to offer motorists the bribe and succeed in reducing pollution without their having to contribute anything. But even if all potential contributors attend, the meeting faces the task of deciding how much to collect from each participant and how much to offer as bribes. Should the PSPC form and overcome this obstacle, it still faces the formidable task of contacting all motorists and getting them to agree to undertake the measures necessary to improve air quality in exchange for the bribes. The zero transaction costs assumption is clearly untenable. The transaction costs of organizing these two groups of individuals are mind-boggling.

In desolation Jane is about to abandon her idea, when she remembers that she is already a member of a club that includes all of the relevant parties – the PSFC. She can make a tax/bribe proposal at the next meeting of the PSFC. If a Pareto-optimal reduction in pollution is possible, there must exist a combination of taxes and subsidies that will win the unanimous support of all citizens of Shangrila. Having resolved this issue, the meeting might go on to consider other issues, like protecting the community from fires and theft, lighting the streets, and so on.

We have discovered another possible reason for the state's existence: to economize on the transaction costs of making collective decisions. Although a separate, voluntary, contractual agreement might be relied upon to correct every market failure in a world of zero transaction costs, in the real world the costs of forming each separate club and writing each contract would be enormous. Once a club that includes all members of the community has been formed to resolve one market

[20] Dixit and Olson show, however, that this result is not robust to the introduction of a modest transaction cost in the form of a cost of attending the meeting. Given such a cost, each individual has an incentive to abstain to avoid it. Should enough persons attend a meeting to provide the public good ($n \geq M$), they now have the incentive to do so, even if $n < N$, so as to avoid incurring the cost of attending the meeting another time.

failure, considerable savings can be made in the costs of bringing together the different groups involved, if this same club is used to resolve other market failures. Thus, the state can be defined as a kind of involuntary membership club that exists to economize on transaction costs when resolving the many market failures that a community faces.[21]

2.13 Experimental results in the voluntary provision of public goods

The assumption of rational, self-interested behavior leads to the following two predictions:

1. In a two-person prisoners' dilemma game that is played only once, both players select the noncooperative strategy.
2. If a two-person prisoners' dilemma game is repeated indefinitely, both players *may* at some point begin to select the cooperative strategy at each new play of the game.

Neither of these predictions has been well supported in laboratory experiments in which subjects, typically university students, play prisoners' dilemma games or, what amounts to the same thing, decide how much to contribute voluntarily to the provision of a public good. Roughly half the participants in one-shot, two-person prisoners' dilemma games cooperate; voluntary contributions to pure public goods average roughly half of the cooperative strategy contribution in one-shot games and in the first round of repeated games. Contributions *fall* if the game is repeated with the same players, reaching the level consistent with the optimal noncooperative strategy after a half dozen or so plays of the game. Both sets of findings contradict the assumption that the subjects in these experiments would behave as rational egoists.[22]

Somewhat more reassuring for the prisoners' dilemma supergame predictions are results from oligopoly experiments, which show first a decline in cooperation as in prisoners' dilemma experiments, and then a continual increase in cooperation until the perfect-collusion/cooperative outcome is reestablished. This cooperative solution does not reemerge, however, until the oligopoly game has been replicated some 35 or more times (Alger, 1987; Benson and Faminow, 1988).

A behavioral assumption that is consistent with the results in these various experiments is that the subjects are *adaptive* egoists. Their current behaviors reflect their past conditioning. Most people since their childhoods have been rewarded for cooperating in prisoners' dilemma situations (being honest, helpful, generous)

[21] When one uses the government to correct for more than one externality and to determine public goods levels *simultaneously*, one confronts head on the problem raised by Aivazian and Callen (1981). One might then anticipate that the absence of a core – that is, the absence of an equilibrium – will be a problem with respect to government decisions on public goods and externalities. This anticipation is correct. See, also, Aivazian and Callen (2000).

[22] The number of experiments of this type is immense. The findings have been surveyed by Davis and Holt (1993, ch. 6), Roth (1995, pp. 26–35), Ledyard (1995), Ostrom and Walker (1997), and Hoffman (1997).

and punished for not cooperating. When they first view the payoffs in a typical voluntary-contribution-public-good experiment, they recognize this as a situation in which cooperation is expected and in the past has been rewarded. Their conditioned reaction is to cooperate, at least to a degree. Such cooperative behavior can quickly be extinguished, however, by the noncooperative or half-cooperative behavior of the other player(s). Indeed, the tit-for-tat strategy, which has fared so well in computer-simulated prisoners' dilemma games, is nothing more than a strategy for conditioning cooperation through the play of the game by rewarding past cooperation and punishing noncooperation.[23]

Evidence of the importance of prior conditioning for determining an individual's behavior in game situations has recently been provided by Glaeser, Laibson, Scheinkman, and Soutter (GLSS, 2000). Their experiments involved individuals' propensity to *trust* other individuals rather than to contribute to a public good; but if background variables are important in one context they are likely to be important in the other. GLSS found that people who disagreed with the statement "you can't trust strangers anymore" were more trusting in the experiments in which they later participated. Both whites and nonwhites tended to be more trusting of members of their own race than of members of a different race. This behavior seems likely to have been conditioned by the individuals' past experiences with strangers and members of other social groups.[24]

There are two reasons to expect the amount of cooperation in a prisoners' dilemma game, or contributions in a voluntary-contribution-public-good game to fall as the number of players increases: (1) the marginal gain from contributing falls as the number of players increases, and (2) it becomes more difficult to identify and punish defectors. The first explanation is the basis for the increasing inefficiency outcome of the voluntary contribution examples discussed in Sections 2.4, 2.5, and 2.11. This prediction has been well supported in the experimental literature. Although individuals do not free-ride to the degree predicted by the rational actor model, they do respond to marginal incentives and contribute more when there are greater marginal gains from doing so.[25]

In a two-person prisoners' dilemma game, defection by the other player can be easily detected and punished. With three or more players, it may be difficult to determine which other player defects, and it is certainly impossible to punish a player who has defected without also punishing all others. This important difference between

[23] Ahn, Ostrom, Schmidt, Shupp, and Walker (2001) and Clark and Sefton (2001) provide experimental evidence of this sort of conditioning of players in repeated game situations.

[24] We shall discuss the potential explanatory power of the adaptive egoism postulate at greater length in Chapter 14, when we attempt to explain another paradox for the rational actor model – why people vote.

[25] See Ledyard (1995, pp. 149–51). An exception to this finding is reported by Isaac, Walker, and Williams (1994), who find that increasing the marginal reward from a contribution while holding the number of players constant has either no effect, or perversely *reduces* the level of contributions, when the number of players is held constant. They do find, however, that when the marginal reward is reduced and the number of players is simultaneously increased, contributions fall. Fisher, Isaac, Schatzenberg, and Walker (1995) find that differences in marginal rewards from contributions within a group are associated with significant differences in contributions with higher marginal incentives associated with higher contributions. See also the discussion in Ostrom and Walker (1997, pp. 49–69).

two-person and n-person ($n > 2$) prisoners' dilemmas may explain why cooperation, in the form of perfect collusion, is often observed in duopoly games, where Cournot and other noncooperative equilibria dominate in all oligopoly games with three or more players (Holt, 1995, pp. 406–9). Although this conclusion is not without controversy, the results from voluntary-contribution-public-good experiments seem to imply that a player's contribution either remains constant or *increases* as the number of players increases, when the marginal gain from an individual contribution is held constant (Ledyard, 1995, pp. 151–8; Ostrom and Walker, 1997, pp. 49–69).

None of these experimental findings offers unqualified support for the predictions of the rational actor models regarding human behavior in prisoners' dilemma-type situations. These findings should not be viewed as undermining the explanation for the existence of the state that rests on prisoners' dilemma/market failure/free-rider behavior, however. In an experimental setting, cooperators and defectors can only be rewarded and punished through the play of the game, or perhaps if communication is allowed, through the verbal rewards and reprimands of the other players. In the real world, a much richer set of rewards and punishments is available, from the slap on the hand or a pat on the head given to a child, to chopping off a hand or a head, in the case of an adult. In real-world settings, individuals do not need to *discover* what their behavior should be and what the other "players" are likely to do, as often is the case in experiments; they are usually told directly. In many real-world settings, communication among the players is possible, and, in this regard, the consistent finding in experiments that cooperation increases when communication is allowed is reassuring.[26]

Thus, if anything, the results from the many prisoners' dilemma and voluntary-contribution-public-good experiments underline the need for an institution like the state that announces what behavior is expected of all individuals in these situations, and helps ensure that this behavior is forthcoming.

Bibliographical notes

Several studies have chosen the state of anarchy as a starting point and shown how property rights, or private protection agencies or the state might emerge as institutional solutions to the social dilemma presented by anarchy. See Skaperdas (1992), Usher (1992), and Sutter (1995).

The best, short introduction to the prisoners' dilemma game is probably by Luce and Raiffa (1957, pp. 94–113). Rapoport and Chammah (1965) have a book on the subject. Taylor (1987, pp. 60–108) presents in a collective choice context an exhaustive discussion of the possibilities of the cooperative solution emerging as an equilibrium in a prisoners' dilemma supergame. Hardin (1982, 1997) also discusses the prisoners' dilemma in a public choice context. Axelrod (1984) explores in depth

[26] See Davis and Holt (1993, pp. 334–8) and Ledyard (1995). Of particular interest in this regard are experiments by Gächter and Fehr (1997), who find that even a minimal opportunity to discuss contributions before and after the experiments serves as a sufficient degree of social sanctioning to induce students to contribute significantly more to the provision of a public good.

the tit-for-tat solution to the prisoners' dilemma supergame and its relevance to the achievement of cooperative outcomes in real-world situations.

Other works that link the prisoners' dilemma to public goods include Runciman and Sen (1965), Hardin (1971, 1982, 1997), Riker and Ordeshook (1973, pp. 296–300), and Taylor (1987, ch. 1). In his excellent survey of the public choice field, Inman (1987, pp. 649–72) discusses several additional explanations of a prisoner's dilemma and why government intervention may improve allocative efficiency.

The experimental literature on prisoners' dilemmas and voluntary contributions to public goods is surveyed by Davis and Holt (1993), Roth (1995), Ledyard (1995), Ostrom and Walker (1997), and Hoffman (1997).

Hamlin (1986) reviews the normative issues surrounding a rational choice theory of the state, placing heavy emphasis on prisoners' dilemma-type rationales for collective action.

Some interesting examples of real-world situations that take on the characteristics of the chicken game, as well as an analysis of solutions to the game, are given by Taylor and Ward (1982).

Classic discussions of externalities include the essays by Meade (1952) and Scitovsky (1954), as well as Buchanan and Stubblebine's (1962) paper and Baumol's book (1967b). Mishan (1971) surveys the literature and Ng (1980, ch. 7) has an interesting discussion of both externalities and the Coase theorem. Cornes and Sandler (1986) provide an integrated analysis of externalities and of both pure and quasi-public goods.

The core is discussed and defined by Luce and Raiffa (1957, pp. 192–6).

Dahlman (1979) links transaction costs and government intervention to the Coase theorem. Frohlich and Oppenheimer (1970) show that more than individual rationality and self-interest (e.g., transaction costs) are needed to conclude that the extent of free-riding increases with group size.