they have no need for intercommittee logrolls and would therefore only bear the cost of passing legislation that did not reflect their interests.

This line of argument is far from showing that intercommittee logrolls do not or cannot occur. But it is sufficient to raise some doubts about the simplest version of the committee government model, in which distinctive panels reciprocally defer to one another's proposals on the floor. Only about a quarter of House committees have been persistently unrepresentative of chamber preferences in the postwar era. Is this number enough to sustain a system of committee-based reciprocity as the dominant feature of legislation?

PART TWO

A Theory of Party Organization

In the previous part we scrutinized several of the key building blocks of the committee government model, concentrating on the notions that committees are autonomous and distinctive. Autonomy can mean many things; we focused on the extent to which the committee personnel process can reasonably be described as autonomous, concluding that it cannot. Distinctiveness also can mean many things; we focused on the geographical and ideological representativeness of House committees, concluding that most committees are representative most of the time.

In this part we shift gears from an examination of empirical details to a broad theoretical question: Why and how might a group of legally equal and often contentious legislators nonetheless create and maintain parties? The answer that we give to this fundamental question is similar in essential respects to the "theory of the firm" developed in the industrial organization literature over the past two decades. But one need not be familiar with this literature to follow the argument. The basic ideas—which are also available in the Hobbesian theory of the state, the theory of political entrepreneurship, and elsewhere—boil down to this: parties are invented, structured, and restructured in order to solve a variety of collective dilemmas that legislators face.[1] These "collective dilem-

1. This statement may suggest that we emphasize the degree to which organizational design and reform are intentional. We certainly do emphasize this intentionality more than, say, Hayek (1960) does. But we do not mean to imply that those who attempt to design institutions are infallible or to preclude the kind of evolutionary perspective articulated in Alchian (1950).

mas"—situations in which the rational but unorganized action of group members may lead to an outcome that all consider worse than outcomes attainable by organized action—are inherent in the drive to be reelected in a mass electorate and in the process of passing legislation by majority rule.

The primary method of "solving" the collective dilemmas that legislators face, we argue, is the creation of leadership posts that are both attractive and elective. In chapter 4 we survey theories of organizational leadership, explaining how it is that the institution of leadership positions can ameliorate the dilemmas facing groups of workers, citizens, legislators, and so forth. In chapter 5 we adapt the ideas sketched out in chapter 4 to the specific case of elected legislators.

FOUR

Institutions as Solutions to Collective Dilemmas

Starting with this and the next chapter, we begin to articulate a view of parties as legislative cartels. This metaphor seems apt to us in part because both cartels and parties—indeed, organizations in general—face a variety of collective dilemmas that must be solved if the organization is to operate effectively. This chapter accordingly deals with the general topic of organizational design and structure. The next chapter then focuses more specifically on legislative parties.

Social scientists from a variety of disciplines study institutions such as legislatures, business firms, public and private bureaucracies, armies, and trade associations. This chapter reviews what we consider to be the most satisfying and comprehensive theory of institutional origins and design: what we shall refer to as the neo-institutional or neocontractarian theory. This theory, exposited fully in no single source, appears in remarkably similar form in a variety of fields. It will be familiar to normative political theorists as a generalized version of the Hobbesian theory of the state, to positive political theorists as a variant on the idea of a "political entrepreneur," and to industrial organization economists as an elaboration on the Alchian-Demsetz theory of the firm. Our purpose here is to underscore the similarity of these various theories—all of which seek to explain institutional features in terms of the choices made by rational individuals facing collective dilemmas—and to examine the answers given to two key questions: How do institutions

"solve" collective dilemmas? What are the costs entailed by institutional solutions?[1]

In the first section of this chapter we define the notion of a collective dilemma more precisely by looking at two examples. In the rest of the chapter we concentrate specifically on the prisoner's dilemma, by far the most studied of collective dilemmas, and on a particular "institutional" solution to it: central authority.[2] In sections 2 and 3 we describe the basics of central authority and why it is resorted to in theory and practice. In sections 4 and 5 we consider some of the costs of central authority and the possibility of doing without it.

1. COLLECTIVE DILEMMAS

A collective dilemma is a situation in which rational behavior on the part of individuals can lead to unanimously dispreferred outcomes.[3] More formally, collective dilemmas are situations that can be modeled by games possessing Pareto-inefficient Nash equilibria.[4] Two well-known game situations can illustrate what we mean.

1.1 STANDARDIZATION

Consider first some problems of standardization. Two railroads would both benefit if they used the same gauge of track, but each prefers the gauge that fits its own trains. Two grain merchants would both benefit if they used the same unit of weight to measure the grain, but each prefers the unit with which he has experience. Two politicians would both benefit if they joined forces to promote the same bill, but each

1. We do not ignore criticisms of the neo-institutional view, but it is not our primary task to deal with them here.

2. Other, more decentralized, institutional solutions, such as a system of property rights, are not dealt with here.

3. We thus agree with the usage in Taylor 1987, 19; and Bates 1987.

4. The usual games included under the rubric of "collective dilemmas" are the prisoner's dilemma, chicken, battle of the sexes, assurance games, and pure coordination games. The formal definition given in the text yields—in light of the folk theorem, discussed in the text below—a great many collective dilemmas when iterated games are considered. In light of the multiplicity of (inefficient) Nash equilibria in the theoretical world, it would seem sensible to adopt some stronger equilibrium concept. Unfortunately, the standard refinement of Nash equilibrium—subgame perfection—also falls prey to a folk theorem. At present there is no well-worked-out refinement of the Nash concept that avoids the multiple equilibrium problem in iterated games. Consequently, we do not pursue any refinements here.

prefers her own version of the bill.[5] In each of these situations, the two agents—call them A and B—have two basic strategies: they can either stick with their current standard or switch to the other player's standard. In the case of the politicians, it may also be reasonable to offer a compromise of some sort, but we will not consider that possibility here.

In some situations of standard setting, the benefits of coordinating are high enough so that both players would prefer to switch to the other's standard rather than fail to coordinate on a standard at all. These situations pose the following problem of interaction: if the other player is going to insist on his or her standard, then it is wise to switch to that standard; but if the other player will soon give in, then one should insist on one's own standard.

The standard-setting problem poses a collective dilemma because both players may rationally insist on their own standard, resulting in an outcome (no shared standard) that both consider inferior to some other that could have been attained (both adopting A's standard or both adopting B's standard). Plausible real-world examples of this inefficient outcome include the multiplicity of different railroad gauges in the nineteenth century and the difference between the U.S. and metric measuring systems.

If we turn to the more formal definition of a collective dilemma—a situation that can be modeled by a game possessing Pareto-inefficient Nash equilibria—then some care must be taken in the present example to choose the right game. A *game* is a formal representation of a strategic choice or a class of strategic choices. It consists of a set of players (in this case, A and B); a specification of the options or strategies of the players (A can either stick with his standard or switch to B's, and similarly for B); a specification of what outcome results from each possible set of options chosen by the players (e.g., if A sticks and B switches, then coordination is achieved on A's standard and B pays some costs in switching); and a specification of how players rank the various possible outcomes (A most prefers that they coordinate on his standard, next that they coordinate on B's standard, next that neither switch, and last that both switch). A matrix representation of the standardization game—in which the value of agreeing on a standard is (arbitrarily) taken to be 5 and the cost of switching to be 2—is given in figure 4.[6]

5. Obviously, the example of the politicians begins to stretch the usual meaning of *standardization* into something one would more usually describe as *coalition*. The point of including it is to emphasize the abstract similarity of these problems, which, in game theory, are both discussed under the general heading of coordination games.

6. If one assumes that "both switching" means that both actually switch to the other's

		Player B: Stick	Player B: Switch
Player A	Stick	0, 0	5, 3
	Switch	3, 5	-2, -2

FIGURE 4. A Standardization Game

Given a game, a *Nash equilibrium* for that game is a set of strategies, one for each player, such that each player does best playing his equilibrium strategy (all other players continuing to play their equilibrium strategies). In the example above, the pair of strategies (Switch, Stick)—where A's strategy is listed first—is a Nash equilibrium because A prefers the equilibrium outcome (coordination on B's standard, with some costs of switching) to the outcome she could get by changing her strategy to Stick (failure to coordinate on either strategy, with no costs of switching). This equilibrium outcome is also *Pareto efficient*, meaning that no other outcome exists that both players prefer to it. A little thought will reveal that the only other equilibrium is (Stick, Switch) and that this, too, is efficient.

It might appear that as the game we have described has no inefficient equilibria, there is by our definition no collective dilemma. However, another game that models the same situation a bit more accurately *does* possess inefficient equilibria. This game consists simply of repeated plays of the game sketched above. Intuitively, such a game introduces the element of time into the situation facing A and B; if they fail to coordinate in the first period, they have another chance. This situation opens up the possibility that each will try to outlast or outbluff the other. Farrell (1987) has shown that this game does have inefficient equilibria, the reason being simply that each player tries to get the other to switch before giving in himself. Thus, in equilibrium, a few rounds of bluffing are expected in which neither party gives in and both forgo the benefits of coordination. This problem might be exacerbated if each player invested more and more heavily in his or her particular standard as time went by, for there might come a time when standardization would no longer be worth it to either (the costs of switching having become too high).

standard—and pay the associated costs—then it seems reasonable that this option would be ranked last, as in figure 4. If one takes switching to be "giving in at the negotiating table," then simultaneous switching would presumably give rise to further negotiation. One particularly simple form of further negotiation would be to flip a coin. In this case the entries in the Switch, Switch cell in figure 4 should be 4, 4 rather than −2, −2. This change would make the game one of chicken. As it stands, it is a variant on the battle of the sexes.

		Player B: Bribe	Player B: Do Not
Player A	Bribe	-1, -1	1, -2
	Do Not	-2, 1	-0, 0

FIGURE 5. A Prisoner's Dilemma

1.2 THE PRISONER'S DILEMMA

In standardization or coordination games the problem is one of strategic uncertainty: will the players settle on one of the multiple Nash equilibria in the game (and if so, which one), or will they fail to coordinate entirely? Another collective dilemma is the famed prisoner's dilemma, in which the problem is one of a "nasty" incentive structure. A two-person version of this game (figure 5) models the situation facing candidates for elective office before effective bribery laws were enacted (e.g., in England before 1883). Each candidate can either bribe some voters or not. Bribery, of course, is costly but—depending on what the other candidate does—it can secure an electoral advantage. Specifically, if both (or neither) bribe, then neither gains an advantage; but if one bribes and the other does not, then the briber is advantaged. If one assumes that the advantage gained by bribing exceeds the cost, then the situation facing the candidates is a prisoner's dilemma. In figure 5 we assume that the utility of the electoral advantage is 2 for both candidates, with a symmetric utility loss of 2 for the disadvantaged candidate; and that the cost of bribery in utiles is 1 (this cost includes the monetary expenditure on bribery; the fines to be paid if caught, discounted by the probability of being caught; and any moral repugnance the candidate may feel). Both candidates have a dominant strategy to bribe; that is, their best strategy is to bribe regardless of what they think their opponent will do. Thus, the strategy pair (Bribe, Bribe) is a Nash

equilibrium. The collective dilemma comes in that this equilibrium is inefficient: both candidates could be made better off if neither bribed. Unfortunately, neither can trust the other not to bribe, and so both incur the costs of bribery without realizing any benefits.[7]

Another prisoner's dilemma (actually a close cousin) can be illustrated by the difficulties facing teams of laborers. In prerevolutionary China large gangs of men would tug fair-sized boats up the Yangtze. The problem was that each man was tempted to slack off a bit. After all, if enough others were pulling, the boat would still progress; if too few others were pulling, it did not matter how hard one pulled anyway.[8] This situation is a collective dilemma because it is a Nash equilibrium for no one to pull at all (if no one else pulls, then the efforts of just one person are futile); yet this equilibrium is inefficient because everyone prefers the outcome in which everyone both pulls and gets paid.

In the remainder of this chapter we discuss some solutions to the prisoner's dilemma, focusing in particular on central authority. In many cases, of course, action can be taken to *create* prisoner's dilemmas. Examples include the district attorney's separation of the suspects in the original prisoner's dilemma, antitrust laws, and open shop laws. We are not concerned with institutions or rules that create dilemmas, however, only with those that solve them.

2. CENTRAL AUTHORITY: THE BASICS

This section concerns what we call "central authority." The gist of this notion can be suggested by recalling the case of the Chinese riverboat pullers discussed above. Cheung (1983, 8) has noted that the problem of loafing was so severe that workers "actually agreed to the hiring of [someone] to whip them," thereby ensuring that everyone both pulled and got paid. This simple idea underpins a wide array of institutional

7. One plausible explanation of the timing of the Corrupt Practices Act of 1883 in the United Kingdom is that the aristocratic politics of the prereform era, in which the prisoner's dilemma could fairly often be kept in bounds by negotiation between candidates who knew one another personally, gave way to an increasingly competitive and open system in which the prisoner's dilemma became less tractable to solution by repeat play.

8. If just the right number of others are pulling, then whether or not you pull can make the difference between the boat progressing and stalling. For this reason, shirking is not a dominant strategy. Analysts generally reserve the term *prisoner's dilemma* for situations in which noncooperation is a dominant strategy—e.g., Schelling 1978; Taylor 1987. Kavka 1987 calls this type of case a "quasi-prisoner's dilemma."

theories, including Frohlich and Oppenheimer's (1978) theory of political entrepreneurship (see also Olson 1965; Salisbury 1969; Frohlich, Oppenheimer, and Young 1971), the Alchian and Demsetz (1972) theory of the firm, and Hobbes's theory of the state (see Gauthier 1969; Kavka 1987).

2.1 POLITICAL ENTREPRENEURS

The theory of political entrepreneurship suggests that *n*-prisoner's dilemmas can be solved by "political entrepreneurs." Political entrepreneurs have three essential features: (1) they bear the direct costs of *monitoring* the community faced with the collective dilemma; (2) they possess *selective incentives* (individually targetable punishments and rewards) with which to reward those whom they find cooperating or punish those whom they find "defecting"; (3) they are *paid,* in various ways, for the valuable service they provide.

A recent example of political entrepreneurship is Dan Rostenkowski's handling of what became the 1986 Tax Reform Act. The Democratic leadership was anxious to avoid the impression that the Democratically controlled House of Representatives had killed tax reform. Rostenkowski, as chair of the tax-writing Ways and Means Committee, was faced with the formidable task of ensuring that the members of his committee did not cave in to special and constituency interests clamoring for the preservation of tax loopholes (the collective dilemma arose because each committee member wished to preserve loopholes benefiting his constituency, but if enough loopholes were preserved, the Democrats could be accused of gutting reform). The selective incentives Rostenkowski had available to reward cooperative members included a variety of legislative favors at his disposal as chair, the most obvious of which were the so-called transition rules—special dispensations publicly justified on the grounds that they allowed a smooth transition from the old to the new tax rules, and politically justified on the grounds that they enabled key supporters to deliver benefits to important constituents. Rostenkowski made liberal use of the transition rules and ultimately was successful in reporting out a bill that made substantial reforms in the tax system (at the same time clearly benefiting traditional Democratic constituencies more than Republican constituencies). The chair was paid for his troubles with a share of the transition-rule largesse, continuance in office, and (perhaps) an increased chance that he would someday become Speaker.

The moral of this story is that a long-standing feature of the institutional structure of the House—to wit, the office of chairman of the Ways and Means Committee—facilitated the successful handling of a potential dilemma for the Democratic party. The basic reason (on which we expand greatly in the next chapter) is that the position is both powerful and essentially elective so that its occupant has both the wherewithal and the incentive to ameliorate collective dilemmas.

2.2 ECONOMIC ENTREPRENEURS

The above view of political institutions has a direct analog in the theory of the firm advanced by Alchian and Demsetz (1972). Business entrepreneurs, in their view, have three distinguishing features. First, they are specialists in monitoring, whose function it is to prevent shirking by workers engaged in team production. Second, they alone have the right to hire and fire individual workers and to negotiate pay. Third, they have a residual claim to all profits produced by the enterprise. Each of these features is directly analogous to the features of political entrepreneurship identified above. First, Alchian and Demsetz define team production in such a way as to make the problem of shirking essentially an *n*-prisoner's dilemma, very similar to the problem facing the Chinese riverboat pullers. Thus, monitoring serves the same purpose in the Alchian-Demsetz firm as in the Frohlich-Oppenheimer political organization. Second, the right to hire, fire, and negotiate pay gives the business entrepreneur some particularly potent selective incentives with which to reward cooperative and punish noncooperative behavior. Third, the business entrepeneur's residual claim is compensation for services rendered. Alchian and Demsetz emphasize the inadequacy of paying a flat salary to the monitor, who then has no economic incentive other than the fear of dismissal to perform his or her duties. Having a claim to all profits in excess of the sum needed to pay workers' wages, however, motivates the entrepreneur to promote efficient collective action in order to maximize output and thus profits.[9]

9. There are several other major versions of the theory of the firm. A good recent review can be found in Tirole 1988. Here it suffices to note that collective dilemmas lurk at the heart of the other major theories as well. Consider, for example, Williamson's (1975) notion of specific investment—an investment of time or money that will have a high range of payoffs if the investor can trade with a specific party and a low range of payoffs if the investor is forced to trade with others. If the specificity of the investment is known to the prospective trading partner, then once the investment is made, the partner may be able to "hold up" the investor for essentially its full value. This possibility, if anticipated, removes the investor's incentive to invest. Williamson argues, among other things, that problems of investment and asset specificity—which are wrinkles on the prisoner's dilemma—are easier to solve within firms than between them. This is one reason why firms exist.

2.3 THE HOBBESIAN STATE

A final example of central authority as a solution to a prisoner's dilemma is the Hobbesian state. The "war of all against all" can be viewed as the collectively irrational outcome of an *n*-prisoner's dilemma (Gauthier 1969; Kavka 1987; Taylor 1987). Hobbes's suggested solution is the institution of an absolute sovereign—an individual or assembly with unlimited authority to act for all members of the polity and with accompanying unlimited lawmaking and enforcement. The monarch monitors and punishes unlawful and aggressive behavior (Kavka 1987, secs. 4.4, 6.1). He is *able* to do so because of the vast power of his office, and he is *motivated* to do so by the fees collected in his courts, by the taxes collected by his officials, and by other devices that provide monarchs with a personal interest in promoting the peace and prosperity of their kingdoms (Hirschman 1977).

2.4 CENTRAL AGENTS

All of the institutional theories surveyed here involve a central agent—whether political entrepreneur, businessperson, or monarch—with three common features: (1) the agent bears the direct costs of *monitoring* the population faced with the collective dilemma; (2) the agent possesses, by virtue of his or her institutional position, *selective incentives* with which to punish noncooperative and reward cooperative behavior; and (3) the agent is motivated to bear the costs of monitoring and to expend scarce resources on selective incentives in punishing and rewarding those whom he or she monitors, either by receiving a substantial share of the collective output, or by receiving a claim to the residual of collective output above some preassigned level, or by some other compensation scheme designed to align the personal interests of the agent with the level of collective output. The essential purpose of establishing a central authority is to create an institutional position whose occupant has a personal incentive to ensure that the collective dilemma is overcome. In Olson's terms, one can think of central authority as an institutional means of transforming latent groups into privileged ones (Olson 1965).[10]

10. One criticism of theories that point to central authority as a solution to collective

It might be noted that central agents are not always confined in the literature to the role of supervisors, as they have been in the discussion above. In some recent versions of the theory of the firm, for example, corporate management is viewed as an arbiter of intrafirm disputes. The gist of this view is that (1) many important transactions carried out within corporations are difficult to fully specify in advance; (2) this vagueness may lead to costly disputes when unforeseen contingencies arise; and (3) the CEO of the corporation thus has an interest in providing cheap, knowledgeable, and rapid "justice" when disputes arise. Part of the reason the corporation exists, then, is that the "legal system" provided within the firm by management is more flexible, cheap, and fair than the state-provided legal system to which the divisions of the corporation would have to appeal were all their transactions conducted in the open market.

This notion of central agent as provider of justice is, of course, a familiar one in the history of the state. The economic value of establishing a reliable system of property and justice, even in a local area, is evident throughout history. From a contractarian perspective, it is one of the clearest reasons to have a state.

The central agent as provider of justice is also visible within parties. Mayhew's *Party Loyalty among Congressmen* (1966) provides a number of examples of the leadership helping to hold together Democratic logrolls on the floor. We pursue this idea at greater length below.

3. WHY CENTRAL AUTHORITY IS SOMETIMES NECESSARY

This section discusses why purely voluntary agreements cannot always be relied on to solve organizational dilemmas. We focus on economic organization, contrasting the fortunes of workers who organize into an Alchian-Demsetz firm with those of workers who remain unorganized

dilemmas is that they presuppose the solution of a prior collective action problem—that is, the creation and support of the central authority (Taylor 1987, 22). This is a valid point, but the collective action problem entailed in creating a position of authority is often more tractable than the original problem. In the case of the riverboat pullers, for example, the workers need only agree that someone be given a whip and a share of the pay. Those who refuse to contribute toward the purchase of a whip—should this be necessary—are simply excluded from the group. The whipper's share is just as secure as that of any ordinary puller. More generally, Hardin (1991) points out that the problem of creating a state when none exists is a coordination problem (of the battle of the sexes kind) rather than a prisoner's dilemma.

(leaving all their transactions to "the market").[11] The focus on economic rather than political organization is chosen for a variety of reasons: political scientists generally are less familiar with this literature and may profit from exposure to it; the theory of economic organization is more fully and formally developed than is the corresponding political theory; the principles of economic and political organization are fundamentally similar despite the fairly obvious initial differences (e.g., economic organizations produce private goods almost exclusively, whereas political organizations usually produce a mixture of both private and public goods). Political organization will, of course, come in for the bulk of our attention in the remaining sections and chapters.

But for now we shall concentrate on the organization of production. Consider a group of n workers producing for sale in the marketplace. Each worker i chooses an action, a_i, from a set of available actions, $A_i = [0,\infty)$. We shall interpret the action $a_i = 0$ as "exert no effort" or "do nothing" and adopt the convention that action a_i requires more effort than action b_i if and only if $a_i > b_i$. Effort is assumed to be costly, so that the ith worker bears a cost—$v_i(a_i)$—for taking action a_i, where v_i is strictly increasing and such that $v_i(0) = 0$. Given a vector of actions, $a = (a_1, \ldots, a_n)$, one for each of the n workers, a total output, $y(a)$, is determined. For simplicity, we shall assume that the price of the output y is \$1, so that the total revenue produced is simply $y(a)$. Can a group of n unorganized workers agree on a method of sharing this revenue such that all workers are properly motivated to work?

This question has been posed, in a precise fashion, by Holmström (1982), who makes three assumptions. (1) *Unobservability:* The particular action taken by each worker is unobservable, so that the share each receives can depend only on total revenue. Holmström denotes i's share as $s_i(y)$. (2) *Budget balancing:* Regardless of the level of total revenue, the shares of the n workers add up to one ($\sum s_i(y) = 1$ for all y). (3) *Concavity of production:* The function y is strictly increasing, concave and differentiable with $y(0, \ldots, 0) = 0$ (no effort, no output). Under these conditions Holmström shows that no n-tuple of actions exists that is both a Nash equilibrium and Pareto efficient. Put another way, he shows that any Nash equilibrium must be Pareto inefficient, so that any group facing the three conditions of unobservability, budget balancing,

11. The reader will notice similarities between the theory of the firm summarized in the following paragraphs and the Hobbesian theory of the state, which contrasts the fortunes of individuals who organize into a state with those of individuals who remain unorganized (leaving all their transactions to "anarchy").

and concavity of production must be mired in what we have called a collective dilemma.

The reason for this can be seen in the context of a simple example. Suppose that $s_i(y) = y/n$ for all i (that is, each worker gets one nth of the total revenue). Each will choose an action a_i in order to maximize the difference between his share of total revenue ($y(a)/n$) and the cost of his action ($v_i(a_i)$). Denoting the partial derivative of y with respect to a_i by y_i', this implies that $y_i'/n = v_i'$. That is, the worker continues increasing his level of effort until the marginal cost of this effort equals one nth of his marginal contribution to total output. But Pareto efficiency requires that each worker equate marginal cost to his full marginal contribution to total output.

It is important to note some limitations on Holmström's result. Gary Miller (1987, 28) interprets it as showing that "with any budget-balancing incentive scheme . . . there will be a tension between individual self-interest and group efficiency—exactly the tension described by the prisoner's dilemma." But schemes satisfying the budget-balancing and unobservability assumptions do exist that support efficient Nash equilibria.

As an example, suppose that $n = 3$, $s_i(y) = y/3$ for all i, $v_i(a_i) = a_i$ for all i, and $y(a) = 0$ unless a_1, a_2, and a_3 are all at least 1, in which case $y(a) = 99$. Note that the specification of y violates the concavity assumption. It says that all three workers must exert a particular minimum level of effort or no salable output is produced at all. All the other major conditions of Holmström's model are satisfied. In this example, however, there exists a Nash equilibrium that is Pareto efficient. Pareto efficiency requires that all three workers choose action $a_i = 1$, yielding a payoff to each of $99/3 - 1 = 32$. But no worker has an incentive unilaterally to depart from this triple of actions. On the one hand, if i lowers her level of effort below 1, her share of output drops 33 (to zero) while she saves at most \$1 in costs. On the other hand, if i raises her level of effort, no more output is produced but additional costs are incurred.

The message of this example is that unorganized workers with unusually complementary skills may achieve efficient equilibria via simple share-of-output agreements. What is required is that the value of total output be quite low until all workers perform their tasks at an acceptable level. In other and more evocative terms, each worker's contribution must be like a link in a chain, not like a drop in the bucket.[12] An

12. This terminology was suggested to us by our colleague Samuel Popkin.

economic example approximating such a production function might be coauthorship when neither coauthor has the other's expertise. A political example sometimes occurs in voting: against a solid minority of 49, a majority of 50 can produce "victory" (and, with it, spoils) only if all members of the majority do their part and vote.

Despite this caveat regarding extreme complementarities in production, however, Holmström's result does show that for a wide class of situations, when actions are unobservable and budgets are balanced, inefficient equilibria are unavoidable. This suggests that a group of workers in an industry where workers cannot monitor one another are inevitably faced with a collective dilemma.

Holmström suggests that the way around this problem is to relax the balanced-budget assumption. In other words, let the workers share the output in some budget-balancing way if output attains the efficient level; otherwise, give all the output away to some third party or destroy it. This scheme does, in principle, allow for efficient equilibria: supposing that everyone is currently working at efficient levels, each is faced with a choice between shirking (which saves some effort but costs the entire output) and not shirking (which requires effort but is remunerated with some share of the total output). As long as each worker's share of the total output exceeds the total cost of his effort, a condition that is satisfied *ex hypothesi,* each worker will work.[13] Holmström's technique works for basically the same reason that efficient equilibria can be attained with extreme production complementarities. Indeed, Holmström can be interpreted as using sharing rules, together with detailed knowledge of the production function y, to *create* the same interdependencies among workers that were posited as a feature of production technology alone in the above example. However, just as such production technologies are the exception rather than the rule, so, too, does it appear that Holmströmian employment contracts are exceptional.

The *rule* in employment contracts is based on a violation of the unobservability assumption. Business entrepreneurs expend resources to monitor the actions of employees and base their pay chiefly on their observed actions rather than on total output. This corresponds to the Alchian-Demsetz model of the firm, or what we have referred to generally as central authority. Central authority can ameliorate the collec-

13. When all workers work at the efficient level, the value of total output must exceed the total social cost of effort for there to be a collective dilemma in the first place. Thus, there always will exist sharing rules that give each worker sufficient remuneration to cover his or her effort costs.

tive dilemma facing workers (in the sense of effecting a Pareto improvement) because workers can be effectively motivated to work by the system of monitoring and sanctions that the central agent implements. Monitoring of employee actions need not be perfect in order to achieve a Pareto improvement. Holmström has shown that partial information about the actions of workers is always valuable in the sense that were such information available at no cost, and workers' pay based in part on it, then all workers could in principle be made better off. The reason for this is simply that the workers' incentives actually to work are greatly improved when their pay is based to some extent directly on their level of work. Output thus can go up rather dramatically when pay is based on direct information about effort (how much it increases depending on the quality of the information), and in principle everyone can share in the profits from increased production. In practice, of course, information is costly. Because of its value in stimulating effort, however, it may be worth a substantial price. It may even be worth the cost of hiring an $(n+1)$st worker whose only job is to monitor and sanction the original workers, as in the example of the Chinese riverboat pullers.

The possibility that monitoring can so improve incentives to work that Pareto improvements result, even after the cost of monitoring is taken into account, is one of the central insights of the Alchian-Demsetz model of the firm. Nonetheless, it should be noted that full or "first-best" efficiency is never attained in Alchian-Demsetz firms. The first-best solution is for every worker to perform the efficient action, a_i^*, and for no resources to be expended on monitoring. The Alchian-Demsetz firm mitigates the incentive problem but does so at the cost of expending real resources in monitoring—an otherwise useless endeavor. A lower bound on the amount by which such firms fall short of first-best efficiency is simply the amount of resources devoted to monitoring (it is a lower bound because workers may shirk even with monitoring—albeit less than without it).

If infinitely high penalties can be imposed on workers caught shirking, then the amount of actual monitoring can be reduced to near zero while still providing workers with sufficient incentives to work at the efficient level. Such a scheme could approximate first-best efficiency as closely as desired (Mirrlees 1976). But, although a "Pascal's wager" solution might work if managers could rent fire and brimstone, bankruptcy laws and other legal devices seem to prevent the infliction of some punishments utilized in hell.

To summarize the discussion so far, it it difficult to achieve first-best

efficiency in group production of private goods. If there are extreme complementarities in production, or if these complementarities are mimicked by the employment contract as suggested by Holmström, then full efficiency can be attained in equilibrium by a simple share-of-output contract with no need for the organization brought by a central agent. But the typical real-world case involves neither extreme complementarities nor extreme contracts, and in this case workers are insufficiently motivated to work if they merely receive a share of total output. This insufficiency of motivation prompts the development of firms in which certain agents monitor and sanction the actions of others. This monitoring cum sanctions gives workers an incentive to work, and output can increase enough to cover the costs of monitoring. Nonetheless, monitoring is costly and would be avoided in a first-best world.[14]

4. MULTIPERIOD CONSIDERATIONS

The model of collective production considered in the last section was for the short term, focusing on the more or less immediate rewards and punishments available to motivate behavior. But the possibility of voluntary or anarchistic cooperation in long-term interactions recently has been prominently argued in the literature. Taylor (1976; 1987) and Axelrod (1981; 1984) have shown that a simple tit-for-tat strategy in two-person iterated prisoner's dilemmas can support cooperation in equilibrium without any apparent institutional structure.[15] The gist of this result is that current noncooperation can be deterred by the threat of future retaliation in kind. If one takes this "shadow of the future" argument seriously enough, the question arises as to why central authority is ever necessary.

14. The results regarding group production of *public* goods are even less encouraging as regards Pareto efficiency. Because by definition everyone consumes the entire quantity of a public good available, whether or not he or she has contributed to its production, the incentive to contribute cannot be manipulated by adjusting the share of output that an agent receives. If extreme complementarities in production exist (everyone must contribute or nothing is produced), the efficient level of output may be achieved. Otherwise, production of the public good would have to rely on the type of selective incentives that Olson (1965) identifies and that Frohlich and Oppenheimer (1978) expand on; that is, some central agent would have to monitor the contributions of individuals and mete out rewards and punishments accordingly. The analog to Holmström's technique would require that the public good be destroyed if not produced in the efficient amount. This is hard to imagine in concrete instances. Would a group failing to clean up a local pond to the extent agreed on then set about the costly task of restoring all the pollutants they had extracted?

15. The equilibrium they identify is not perfect. See, for example, Sugden 1986 for a robust perfection.

Part of the answer has to do with two of the assumptions that make the "shadow of the future" formidable—that both parties can observe whether or not the other has cooperated and that both expect the interaction to last a long time. Both of these assumptions rely for their approximate fulfillment in the real world on the existence of appropriate institutions.

Consider first the problem of unobservability. If the players in an iterated prisoner's dilemma can neither observe whether others have cooperated in each stage of the game nor infer this from what they can observe, then policing noncooperation by in-kind retaliation is obviously problematic. This is essentially the difficulty facing many arms control agreements. An agreement not to develop certain kinds of weapons may be concluded, but typically neither side can easily observe compliance. Moreover, neither side can observe a payoff in increased security from which compliance might be inferred. Hence, elaborate verification procedures are resorted to in an attempt to provide sufficient observability so that both sides feel they have a credible deterrent threat. The role of U.N. monitors in verifying the winding down of the Iran-Iraq war provides a recent and more explicitly institutional example of the same point.[16]

A second prerequisite for decentralized, purely voluntary cooperation is that both players believe their interaction will last long enough so that the possibility of *future* gains can deter *present* noncooperation. On the one hand, this belief can be endangered in a number of ways, giving rise to a collapse of cooperation. On the other hand, it can be shored up institutionally. For example, Kreps (1990) illustrates how business firms—artificial persons with indefinitely long lives—can replace natural persons for the purposes of many transactions. If an individual has a reputation for dealing honestly with customers whom he might cheat, there is a possibility that his customers will become nervous when they believe he is near retirement or death. If a firm has a reputation for dealing honestly with customers whom it might cheat, there is less reason to become nervous when the current owner nears retirement or death, for a firm's reputation for honest dealing is a valuable asset that contributes to the sale price of the firm. Thus, the owner nearing retirement recognizes that any cheating of customers in the twilight of his career may cost more (in the form of a lowered sale price) than it is worth.

16. On the U.N. monitors, see *Maclean's*, 29 Aug. 1988, 10–17. Another important function of the U.N. troops was to raise the cost of violating the agreement, since violation might entail casualties among noncombatant nations.

In addition to the problems of unobservability and shortness of interaction, which hinder voluntary cooperation even between two persons, several other difficulties appear or are exacerbated when the number of players grows beyond two. The most straightforward of these difficulties is illustrated by the steady erosion of incentives to contribute to public goods that often occurs as a group's size increases. Theoretically, this erosion follows in models in which the importance of individual contributions declines with group size (cf. Hardin 1982).

An institutional response to the problem of maintaining voluntary cooperation in large groups is illustrated in the Hutterite communes. The Hutterites have developed an elaborate procedure for regularly splitting their communities whenever a certain optimal size (sixty to one hundred individuals, or about six to ten families) is exceeded (Bullock and Baden 1977). A similar emphasis on smallness (plus a bit of isolation) characterizes other successful communal life-styles (e.g., that of the Israeli kibbutzim).

A second difficulty, which appears in two-person prisoners' dilemmas but is more troublesome in n-prisoners' dilemmas, is the problem of multiple equilibria. This can be explained by adverting to one of the more remarkable results in the theory of repeated games: the "folk theorem." The folk theorem, so called because it is widely known to game theorists but of obscure authorship, deals with repeated noncooperative games (games in which the players cannot make binding agreements). Let G be a noncooperative game in normal form (two examples of such games are given in figures 4 and 5). Denote by G^* the "supergame" of G—that is, the game that consists of an infinite sequence of plays of the "stage game" G. Roughly put, the folk theorem states that, if the players of G^* have enough information (in particular, at the end of each stage they are informed of the strategy chosen by all other players in that stage), then *any* outcome that is individually rational can be supported by some Nash equilibrium.[17] An outcome is individually rational if the payoff each player gets is not less than his *security level,* defined as the worst payoff that can be forced on him by the remaining players. Thus, very little restriction is placed on the outcomes that one

17. The only other restriction to note is that the outcome be feasible—that is, that it be possible to attain such an outcome via some strategy n-tuple in G (or via some *correlated* strategy n-tuple in G). See Aumann 1981.

might predict in a repeated noncooperative game by the notion of Nash equilibrium alone.

An industry has arisen in game theory in response to the problem of multiple equilibria, devoted to finding such refinements of Nash equilibria as perfect or sequential equilibria (see, for example, Selten 1975; Kreps and Wilson 1982; Kalai and Samet 1982; Banks and Sobel 1987). It is safe to say, however, that none of the refinements of Nash equilibria proposed so far produces unique and widely accepted equilibrium predictions for all games; indeed, the only refinement concept offered that produces unique predictions is that of Harsanyi and Selten (1988). The multiplicity of equilibria, however, means that a coordination problem similar to the standardization game discussed in section 1 can arise over which of the many equilibrium outcomes will be selected. One view of leadership (or central agency, in our terms) is as a mechanism for preventing any efficiency losses through lack of coordination (see Calvert 1985; Kavka 1987, 247).

A third impediment to decentralized cooperation in large groups is what we refer to as the "group punishment" feature. In a repeated prisoner's dilemma, if one player defects at time t, the only way to punish her (if side payments are not allowed) is for some other player to defect at some future date. But any such future defection unavoidably hurts not just the original defector but all other players as well. It is therefore questionable whether collective action can hold together on a purely voluntary basis, just on the threat of in-kind retaliation. Should Ms. A really resume polluting in order to punish Mr. B's act of pollution, given that she thereby also punishes C, D, and E? Can a group of laborers gain a reputation for reliability if shirking by any one of them is punished by retaliatory shirking? When the strategy of in-kind retaliation is carried to its logical extreme—in the so-called grim trigger strategy—cooperation is enforced by the threat of universal and perpetual defection. But if this threat ever gets carried out (in a model in which mistakes are possible, presumably), the result is tantamount to the utter dissolution of the organization. We do not believe that many important organizations are held together by a well-understood threat of dissolution should any member defect. It is possible to bypass the group punishment feature if side payments—transfers of private goods—are feasible. In this case, a defector can be punished directly: whipped, fined, ostracized, frowned on, whatever. But the other problems—unobservability of actions, shortness of time horizons, insignificance of individual contributions, and multiplicity of equilibria—remain and may be severe.

5. PROBLEMS WITH CENTRAL AUTHORITY

Although institutionalizing central authority can in principle be effective in overcoming collective dilemmas, there is no guarantee that it will do so in practice.[18] Central authority can be either too weak or too strong. It is *too weak* when the selective incentives at its command are insufficient to deter noncooperative behavior, so that the potentially capturable benefits of cooperation are not in fact captured. This is the case of the king who cannot maintain internal order or of the businessman who cannot prevent shirking by his employees. Central authority is *too strong* when the selective incentives at its command allow the incumbent central agent to appropriate all of the rents produced by collective effort, to deter any attempt to remove him, and even to extract resources produced by individual (noncollective) action. This is the case of the strong queen who maintains order but extracts taxes so high that each citizen is nearly indifferent between that order and war of all against all.

These twin problems besetting the institution of central authority recall Madison's comment in *Federalist* 51: "In framing a government to be administered by men over men, the great difficulty lies in this: you must first enable the government to control the governed; and in the next place oblige it to control itself." When Madison wrote this passage, the American people had recently experienced both government too strong (under George III) and government too weak (under the Articles of Confederation); his dictum simply recalled this experience to his readers' minds. From a contractarian perspective, Madison's statement haunts any institution that relies on central authority to solve collective dilemmas: what is the point of central authority if it fails, through weakness or through strength, to effect a Pareto improvement?

The rest of this section is devoted to a discussion of one horn of this dilemma: the problem of authority that is too strong. This problem has, of course, received considerable attention from political theorists through the years. In particular, the motivation of the central agent has received extensive attention. Alchian and Demsetz (1972) argue that the central agent will have the incentive to monitor at the efficient level only if he has a claim to all revenues above a certain fixed level (i.e., only if he is the residual claimant). Frohlich and Oppenheimer (1978) are less precise but emphasize the importance of the central agent having a substantial stake in the collective action being organized. Some versions of

18. See Kiewiet and McCubbins 1991.

the theory of absolute monarchy emphasize the monarch's ultimate ownership of all land (cf. Hirschman 1977).

It is not clear that any of these techniques of motivating the central agent—giving him the residual claim, a substantial stake, or reversionary ownership rights—adequately deals with the problem at hand. A residual claimant may profitably devote her time to driving her employees' wages down rather than helping to increase their productivity.[19] A political entrepreneur may sell his followers out. A king may prosecute ruinous foreign wars or pursue a luxurious life-style rather than sticking to his Hobbesian functions. These defects of motivation have given rise to a variety of institutional supplements. We discuss them here under three headings: establishing mechanisms for the central agent's removal, lengthening the agent's time horizon, and putting central authority into commission.

The first of these techniques is the most straightforward. If the central agent can be removed by the group whose agent he is, then actions detrimental to all or most of the group should presumably be discouraged. From this perspective, it is an important part of the total compensation package that CEOs can be removed by the stockholders (but not, usually, by the workers); that legislative party leaders are elected (usually by members of their party serving in parliament); that pirate captains (Ritchie 1986) and the kings of the ancient Germanic tribes were elected by their followers; and that the right of the people to overthrow "unjust" monarchs, even those reigning by divine right, was clearly understood. Of course, the practical importance of the possibility of removal depends on how real the possibility is. At one extreme, if removal requires revolution and revolutionaries can be hung, then a substantial prisoner's dilemma arises over who is to bear the cost of providing the collective good of removing the tyrant. At the other extreme, competitive and regular elections with low costs to losing challengers may impose a substantial constraint on the incumbent central agent.

Another technique of shoring up the incentives of the central agent, which complements the possibility of removal, is to lengthen his time horizon. This strategy makes the threat of removal more potent because there is more to be lost in the future. Time horizons of monarchs can be extended by making monarchy hereditary.[20] Time horizons of corporate managers can be extended by the development of marketable reputations and "good will" (Kreps 1990) or by the posting of bonds (Jensen and Meckling 1976). Time horizons of politicians can be extended by attractive (but forfeitable) pension schemes, such as peerages and knighthoods in England or retirement benefits in the U.S. Congress, and by putting no limit on the number of terms that may be served.

19. There is a substantial literature in economics on managerial incentives in large corporations, much of it concerned with managers who maximize their own utility rather than firm profits. Major examples include Baumol's (1962) sales maximization hypothesis and Williamson's (1967) managerial discretion model.

20. English law considered kings and bishops to be corporations sole, with infinite lifetimes.

A third and quite common technique of getting around the problem of too-strong central authority is to make it collective. The institutionally simplest way to do so is to put central authority into commission. Examples include plural executives, such as the Roman triumvirates or the Swiss Federal Council, and corporations, whether civil, eleemosynary, business, or municipal. Institutionally more elaborate schemes fall under the rubric of "checks and balances": the independent judiciary, the separation of executive and legislative powers, bicameralism, the independent comptroller in business firms, and so forth (Lijphart 1984; Baylis 1989; Watts and Zimmermann 1983).

The simpler examples of collective authority, where central power is shared but not institutionally divided and balanced, raise an obvious trade-off. On the one hand, the more central agents there are, the less likely that all will collude in schemes of corruption or oppression. Ideally, each will watch the others. On the other hand, the more members there are, the smaller the stake and say of each in collective affairs; hence, it is less likely that the collective action problem among the central agents will be overcome. The trick is to replace a single central agent, who ideally can convert a latent into a privileged group but who cannot quite be trusted, with a group of central agents that is (1) small enough, and in frequent enough interaction, so that voluntary cooperation in sharing the costs of monitoring can emerge; (2) composed in such a way that it can be trusted; and (3) large enough or given a large enough stake in the success of collective action so that it is viable in Schelling's sense (that is, each member of the group will benefit if the group cooperates in policing collective action, even if they bear all the costs themselves; see Schelling 1978).[21]

21. Other factors promote "good behavior" by central agents but are not endogenous to a single group. An example is the market for top corporate managers. Fama (1980) notes that a manager's future remuneration depends substantially on the past performance of the firms managed. Thus, each individual firm does not need to solve the problem of managerial motivation solely by internal means; it is helpful by the existence of a properly functioning market for managerial talent. A political analogue is implicit in Schlesinger's (1966) idea of "progressive ambition."

6. CONCLUSION

This chapter has surveyed theories of organizational design from several fields: the theory of political entrepreneurship from the political economy literature; the theory of the firm in the industrial organization literature; and the Hobbesian theory of the state. From this survey we have pieced together a common view of the origin and functioning of organizations.

In rough outline, this view goes as follows: Collective action in any field of endeavor can produce a surplus, in the sense that collective output exceeds the sum of individual outputs. This surplus appears in firms, for example, whenever the production process is such that what worker A does increases the marginal productivity for worker B, and it appears in armies whenever what soldier A does increases the marginal effectiveness of soldier B. Such a surplus from collective action is an incentive to collective action. Unfortunately, even if the product is private (widgets or plunder) instead of public (national defense), a substantial free-rider problem stands in the way of voluntary cooperation. Absent unusual conditions, any single-period contract based solely on sharing the collective output leaves substantial incentives to shirk and free ride (Holmström 1982); and any multiperiod contract based solely on in-kind retaliation for shirking is implausible in large organizations. Thus, simple sharing rules and in-kind retaliation rules cannot sustain large organizations. Some attention to the actual actions taken by the various workers, soldiers, political activists, and so on is needed.

This necessity for keeping track of the actual effort and actions taken leads to the creation of specialists in monitoring—and gives rise to the profusion of auditors, managers, and supervisors observable in all real-world organizations of any size. But *quis custodiet ipsos custodes?* The answer has always been to arrange the incentives of auditors so that they will in fact ameliorate problems of collective action. The two basic forms this tinkering with incentives has taken are checks and balances (getting the auditors somehow to watch one another as well as those they audit) and hierarchy (placing auditors above the auditors). The latter solution, of course, leaves the top auditor unwatched, and here the solution has been twofold: to give the top auditor—whether general, CEO, or prime minister—a substantial personal stake in the success of the collective enterprise; and to provide a mechanism—coup, proxy fight, election, or whatever—for his or her removal.

FIVE

A Theory of Legislative Parties

Definitions of political parties have been offered from two main perspectives, one emphasizing structure, the other purpose. The structural perspective defines parties according to various observable features of their organization. Studies of the historical development of parties, for example, take pains to distinguish "premodern" parties from "modern" ones, typically by pointing to the increasing elaboration of extraparliamentary structures in the latter (Duverger 1954; LaPalombara and Weiner 1966). The purposive approach, by contrast, defines and categorizes political parties by the goals that they pursue. Typical examples include Edmund Burke's definition of a party as a group of men who seek to further "some particular principle in which they are all agreed" (Burke 1975, 113); Schattschneider's definition of a political party as "an organized attempt to get . . . control of the government" (Schattschneider 1942, 35); and Downs's definition of a political party as "a team of men seeking to control the governing apparatus by gaining office in a duly constituted election" (Downs 1957, 25).[1]

Neither the structural nor the purposive definitions of parties are suited to the needs of this chapter. The structural definitions take as defining features the kinds of things that we hope to explain. Moreover, these definitions generally turn on extraparliamentary organization rather than on the intraparliamentary organization that is our main concern.

1. Another approach, which defines parties in terms of the actions that they take, is pursued in Panebianco 1988.