COVENANTS WITH AND WITHOUT A SWORD: SELF-GOVERNANCE IS POSSIBLE

ELINOR OSTROM, JAMES WALKER, and ROY GARDNER
Indiana University, Bloomington

Contemporary political theory often assumes that individuals cannot make credible commitments where substantial temptations exist to break them unless such commitments are enforced by an external *agent. One such situation may occur in relation to* common pool resources, *which are natural or man-made resources whose yield is subtractable and whose exclusion is nontrivial (but not necessarily impossible). Examples include fisheries, forests, grazing ranges, irrigation systems, and groundwater basins. Empirical evidence, however, suggests that appropriators in common pool resources develop credible commitments in many cases without relying on external authorities. We present findings from a series of experiments exploring (1) covenants alone (both one-shot and repeated communication opportunities); (2) swords alone (repeated opportunities to sanction each other); and (3) covenants combined with an internal sword (one-shot communication followed by repeated opportunities to sanction each other).*

And Covenants, without the sword, are but words, and of no strength to secure a man at all.
—Hobbes, *Leviathan*

Contemporary political theories frequently presume that individuals cannot make credible ex ante commitments where substantial ex post temptations exist to break them unless such commitments are enforced by an external agent. Hobbes justified the necessity of Leviathan on the frailty of mere words. Hobbes calls a contract that involves a promise by at least one of the parties to perform in the future a *covenant* (1960, 87). When both parties promise future performance, it is a *covenant of mutual trust* (p. 89). A covenant of mutual trust in a state of nature is void in Hobbes's view if either has a reasonable suspicion that the other will not perform: "For he that performeth first, has no assurance the other will perform after; because the bonds of words are too weak to bridle men's ambition, avarice, anger, and other passions, without the fear of some coercive power; which in the condition of mere nature, where all men are equal, and judges of the justness of their own fears, cannot possibly be supposed" (pp. 89–90). On the other hand, a covenant made "where there is a power set up to constrain those that would otherwise violate their faith" is likely to be fulfilled (p. 90). Thus, Hobbes argued for the necessity of a "coercive power, to compel men equally to the performance of their covenants, by the terror of some punishment, greater than the benefit they expect by the breach of their covenant" (p. 94).

The weakness of mere words and the necessity of external agents to enforce contracts is also a foundation upon which the powerful edifice of noncooperative game theory has been constructed. John Nash (1950, 1951) distinguished between cooperative and noncooperative games. In cooperative games, players can communicate freely and make enforceable agreements; in noncooperative games, they can do neither. Some theorists particularly stress the inability to make enforceable agreements: "The decisive question is whether the players can make enforceable agreements, and it makes little difference whether they are allowed to talk to each other. Even if they are free to talk and to negotiate an agreement, this fact will be of no real help if the agreement has little chance of being kept. An ability to negotiate agreements is useful only if the rules of the game make such agreement binding and enforceable" (Harsanyi and Selten 1988, 3).[1] In other words, communication without a change in the payoff function does not eliminate a Nash equilibrium (the major solution concept used in noncooperative game theory). Not all Nash equilibria are efficient. Equilibrium payoffs are lower in a finitely repeated, full-information dilemma game than if the participants could make credible ex ante commitments. Hobbes's state of nature has frequently been represented as a social dilemma game.[2]

The assumption that what is called a *sword* wielded by an "external enforcer" is necessary before individuals can make credible ex ante commitments in social dilemma situations has important implications for common pool resources. Common pool resources (CPRs) are natural or man-made resources whose yield is subtractable and the exclusion from which is nontrivial (but not necessarily impossible).[3] Because CPRs can be represented as social dilemmas, particular policy recommendations follow almost immediately. Ophuls argued, for example, that "environmental problems cannot be solved through cooperation . . . and the rationale for government with major coercive powers is overwhelming" (1973, 228). His conclusion was that "even if we avoid the tragedy of the commons, it will *only* be by recourse to the tragic necessity of Leviathan" (p. 220).

Empirical evidence suggests, however, that individuals facing social dilemmas in many cases develop credible ex ante commitments without relying on external authorities. *Appropriators* from CPRs (e.g., fishers, irrigators, and herders) have repeatedly shown their capacity to organize themselves, establish credible commitments, monitor each others' behavior, and impose sanctions on those who break their commitments.[4] Self-organized CPR institutions have been devised without reference to central authorities and sustained over long periods of time without enforcement by external agents.[5] Experimental studies have repeatedly found that individuals placed in laboratory social dilemmas and allowed to communicate consistently achieve better outcomes than predicted by noncooperative game equilibrium. While findings concerning the positive effect of communication in laboratory settings (covenanting without a sword) lead to optimistic predictions about the capacities of individuals to solve social dilemmas, many cases of self-organized arrangements in field settings rely on internal sanctioning mechanisms (E. Ostrom 1990).

Past research has produced three anomalies from the standpoint of predictions derived from both Hobbes's state-of-nature theory and noncooperative game theory:

1. In one-shot social dilemma experiments, communication alone leads to substantial improvement in outcomes.[6]
2. In repeated social dilemma experiments, repeated communication alone leads to substantial improvements in joint outcomes.[7]
3. In field settings of repeated social dilemmas, participants invest substantial time and effort monitoring and imposing sanctions on one another.[8]

The results from field settings show that participants in social dilemmas do not rely entirely on communication. They monitor each other closely and impose sanctions on those who do not conform to the rules they have devised. If communication alone were a fully reliable mechanism to overcome the gap between ex post temptations and ex ante promises, then one should not see time and effort devoted to monitoring and sanctioning efforts in the field. Thus, prior research has generated interesting puzzles. Given that social dilemmas lie at the foundation of the theory of the state and the theory of collective action, it is important to explore the independent and interactive effects of (1) communicating (covenanting) alone, both one-shot and repeated; (2) sanctioning (the sword) alone; and (3) communicating with options to sanction (covenants with a sword). We shall manipulate experimental treatments to examine all three mechanisms. We shall construct a common constituent game as the basis for all these manipulations. This game, an n-person CPR appropriation game, will now be described and solved.

GAME-THEORETICAL PREDICTIONS

The Common Pool Resource Constituent Game

We will first specify the class of constituent CPR games from which we draw our designs. Assume a fixed number n of appropriators with access to the CPR. Each appropriator i has an endowment of resources e which can be invested either in the CPR or a safe, outside activity. The marginal payoff of the outside activity is normalized equal to w. The payoff to an individual appropriator from investing in the CPR depends on aggregate group investment in the CPR and on the appropriator investment as a percentage of the aggregate. Let x_i denote appropriator i's investment in the CPR, where $0 \le x_i \le e$. The group return to investment in the CPR is given by the production function $F(\Sigma x_i)$, where F is a concave function, with $F(0) = 0$, $F'(0) > w$, and $F'(ne) < 0$. Initially, investment in the CPR pays better than the opportunity cost of the forgone safe investment [$F'(0) > w$]; but if the appropriators invest all resources in the CPR the outcome is counterproductive [$F'(ne) < 0$]. Thus, the yield from the CPR reaches a *maximum net level* when individuals invest some, but not all, of their endowments in the CPR.[9] This environment is much richer and more complex than the ubiquitous repeated prisoner's dilemma game. While no formal game or laboratory experiment ever captures all the nuances of field settings, this n-person CPR game is a far more realistic environment in which to investigate the questions we have posed than many of the dilemma games previously explored.

Let $x = (x_1, \ldots, x_n)$ be a vector of individual appropriators' investments in the CPR. The payoff to an appropriator, $u_i(x)$, is given by

$$u_i(x) = we \text{ if } x_i = 0$$
$$= w(e - x_i) + (x_i/\Sigma x_i)F(\Sigma x_i) \text{ if } x_i > 0 \,. \qquad (1)$$

Equation 1 reflects the fact that if appropriators invest all their endowments in the outside alternative, they get a sure payoff (we), whereas if they invest some of their endowments in the CPR, they get a sure payoff $w(e - x_i)$ plus a payoff from the CPR, which depends on the total investment in that resource $F(\Sigma x_i)$ multiplied by their share in the group investment ($x_i/\Sigma x_i$).[10] Let the payoffs (represented in equation 1) be the payoff functions in a symmetric, noncooperative game. Since our experimental design is symmetric, there is a symmetric Nash equilibrium, with each player investing x_i^* in the CPR, where

$$-w + (1/n)F'(nx_i^*) + F(nx_i^*)[(n - 1)/x_i^*n^2] = 0 \,. \qquad (2)$$

At the symmetric Nash equilibrium, group investment in the CPR is greater than optimal, but not all yield from the CPR is wasted.

Compare this deficient equilibrium to the optimal solution. Summing across individual payoffs $u_i(x)$ for all appropriators i, one has the group payoff function

$$u(x) = nwe - w\Sigma x_i + F(\Sigma x_i) \,,$$

which is to be maximized subject to the constraints $0 \leq \Sigma x_i \leq ne$. Given the specified productivity conditions on F, the group maximization problem has a unique solution characterized by the condition:

$$-w + F'(\Sigma x_i) = 0. \quad (3)$$

According to equation 3, the marginal return from a CPR should equal the opportunity cost of the outside alternative for the last unit invested in the CPR.

It is worth noting that both the Nash equilibrium investment and the optimum group investment do not depend on the endowment parameter e as long as e is sufficiently large. Out of equilibrium, however, a larger e means players are capable of making larger mistakes. If the time required to converge to an equilibrium is large, then the Nash equilibrium may be an inappropriate behavioral theory for CPRs with high endowments.

Finite Repetition of a Common Pool Resource Constituent Game

Denote the constituent game by X and let X be played a finite number of times. Typically, a repeated game has many equilibria to choose from. Two equilibrium refinement principles are subgame perfection and subgame consistency. An equilibrium is subgame perfect if it prescribes equilibrium play on every subgame. An equilibrium is subgame consistent if it prescribes identical play on identical subgames. If the constituent game has a unique equilibrium, then the finitely repeated game has a unique subgame perfect and subgame consistent equilibrium (Selten 1971). Thus, equation 2 characterizes a finite sequence of equilibrium outcomes. This prediction is based on the assumption of a finite game of complete information. Our subjects know the game is finite.[11] Although we do not have complete control over our subjects' understanding of their decision task, we make all information readily available to them. Failure to induce complete information on the part of our subjects jeopardizes the uniqueness of this refined equilibrium (Kreps et al. 1982).

Communication and the Constituent Game

When the constituent game X has a unique equilibrium x^*, neither repetition nor communication creates new equilibrium outcomes. Let c denote a communication strategy, in the communication phase C, available to any player. As long as saying one thing and doing another has no payoff consequences, then any strategy of the form (c, x^*) is an equilibrium of the one-shot game (C, X), and finitely repeated x^* is a subgame consistent equilibrium outcome of one-shot communication (C, X, X, . . . , X) or repeated communication (C, X, C, X, . . . , C, X). In this situation, subgame consistency is deaf to covenants. However, as we shall show, communication makes a big difference in behavior.

Sanctioning and the Constituent Game

Our sanctioning institution is represented formally using the following construction. Let s be a matrix of zeros and ones, where $s_{ij} = 1$ means that player i has sanctioned player j, and $s_{ij} = 0$ means that i has not sanctioned j. Row i of the matrix s codes all of player i's sanctioning behavior. As before, let x be a vector of individual investments in the CPR and $u_i(x)$ be i's payoff function in the game without sanctioning. Player i's payoff function in the game with sanctioning, $u_i(x, s)$, is given by

$$u_i(x, s) = u_i(x) - f1 \sum_j s_{ij} - f2 \sum_j s_{ji}. \quad (4)$$

The parameters $f1$ and $f2$ represent the cost of fining and being fined, respectively.[12] The sum Σs_{ij} is the total number of fines j levied by player i, costing him $f1$ each; the sum Σs_{ji} is the total number of times player i is fined, costing him $f2$ each.

Adding this sanctioning mechanism to our constituent game X produces a game X–S with a unique subgame consistent equilibrium. In a one-shot game with a unique Nash equilibrium x^*, any sanctioning activity is costly and cannot lead to higher payoffs. Thus, the equilibrium of the one-shot game with sanctioning is the pair $(x^*, S^*) = (x^*, 0)$; that is, the equilibrium sanctioning matrix is the zero matrix. At equilibrium, no one sanctions. Now suppose that the one-shot CPR game with sanctioning is to be repeated a finite number of times T. This finitely repeated game has a unique subgame consistent equilibrium given by strategy 1:

STRATEGY 1. *In every round, play (x^*, 0). In the event of any deviation from prescribed play, resume playing (x^*, 0) after the deviation.*

This equilibrium follows from backward induction. At the last round, T, no deviation is profitable. At the next to last round, $T - 1$, given that no deviations will occur in the last round, then no deviation is profitable—and so on. Repeating the game should not lead to sanctioning, either. There is, however, compelling evidence that backward induction must be learned through repeated play.

Besides the unique subgame consistent equilibrium, there are many imperfect equilibria as well. Let $z_i < x_i^*$ be the same for all i. Consider the repeated game strategy, strategy 2:

STRATEGY 2. *In every round except T, play (z, 0). In the event of any deviation, play ($x_i = e$, s = I) for one round, then resume playing (z, 0). If no deviation took place in round $T - 1$, play (x^*, 0) in round T.*

This represents a trigger strategy. All players agree to invest less than they would according to strategy 1. If some player cheats, then every player dumps all his resources into the CPR ($x_i = e$) and every player issues one sanction for one round. Then play returns to normal. In the final round, everyone plays the one-shot Nash equilibrium. We claim that strategy 2 represents an imperfect equilibrium. To show this, it suffices to show that no deviation from prescribed play pays. Let $F(ne)$ be a very large negative number. For $f1$ and $f2$ large enough, a player who deviated optimally for one round would gain some positive amount, depending on the level of z_i but in the next round would lose $(1/n)F(ne) + f1 + f2$ due to punishment from overinvestment and sanctions, as in strategy 2. This threat we call the *dire threat*, since it is the worst threat imaginable for one round in our design. Given such a threat, it does not pay to deviate, even for one round. Finally, if a punishment is not called for in the last round, the endgame equilibrium is played in that round. This shows that strategy 2 is an equilibrium. Its imperfection lies in the fact that the trigger punishment—dumping all tokens into the resource, everybody placing a fine—is too harsh to be credible at the end of the game.[13]

There is a large set of equilibria along the lines of strategy 2 involving variation of the length of punishment (1 or more rounds), the base level of investments z_i, and the direness of the one-period threat (dump not quite all tokens in the CPR, levy fines with some probability). In particular, by varying $f1$ and $f2$, we hoped to allow the subjects to find equilibria of the strategy 2 family involving punishments of the form (z_i, I) (i.e., reduced investment in the CPR) but sanctions for everyone if a deviation occurs (see Jankowski 1990).

Communication, the Constituent Game, and Sanctioning

We investigate the combination of communication with sanctioning in two ways. Our first design allows for a one-shot communication period, which is then followed by a sequence of constituent games with a sanctioning mechanism imposed. In our second design, we impose a one-shot communication period in conjunction with an opportunity for the subjects to choose whether or not they want a sanctioning mechanism. In both designs, the payoff functions are still given by equation 4, since communication per se has no payoff consequences and sanctioning does. Without loss of generality, let c be a communication strategy. Then appending c to strategy 1 yields a subgame consistent equilibrium, and every subgame consistent equilibrium has the same payoffs as strategy 1. In addition, as in repeated X–S, imperfect equilibria exist yielding higher payoffs than equilibria which are subgame consistent.

THE LABORATORY DECISION ENVIRONMENT

Design

In our experimental investigation we have operationalized this CPR environment with eight appropriators ($n = 8$) and quadratic production functions $F(\Sigma x_i)$, where

$$F(\Sigma x_i) = a\Sigma x_i - b(\Sigma x_i)^2 ,$$

with $F'(0) = a > w$ and $F'(nw) = a - 2bnw < 0$. For this quadratic specification, one has from equation 3 that the group optimal investment satisfies $\Sigma x_i = (a - w)/2b$. The CPR yields 0% on net when investment is twice as large as optimal, $\Sigma x_i = (a - w)/b$. Finally, solving equation 2, the symmetric Nash equilibrium group investment is given by

$$\Sigma x_i = [n/(n + 1)](a - w)/b .$$

This level of investment is between maximal net yield and zero net yield, approaching the latter as n gets large. One additional constraint that arises in a laboratory setting is that the x_i be integer-valued. This is accomplished by choosing the parameters a, b, d, and w in such a way that the predictions associated with Σx_i are all integer-valued.

In particular, we use the parameters shown in Table 1. These parameters lead to the predictions that 36 tokens invested in Market 2 yields the level of investment that maximizes group earnings and that each subject investing 8 tokens in Market 2 is a unique Nash equilibrium. At this equilibrium, subjects earn approximately 39% of maximum net yield from the CPR. Once again, note that the Nash equilibrium and optimal investment are not affected by the level of endowments.

Subjects and the Experimental Setting

Our experiments used subjects drawn from the undergraduate population at Indiana University. Students were volunteers recruited from principles-of-economics classes. Prior to recruitment, potential volunteers were given a brief explanation in which they were told only that they would be making decisions in an "economic choice" environment and that the money they earned would be dependent upon their own investment decisions and those of the others in their experimental group. In all experiments reported here, subjects were randomly recruited from a pool of subjects with prior experience in a CPR decision environment. All experiments were conducted on the NOVANET computer system at Indiana University. The computer facilitates the accounting procedures involved in the experiment, enhances across-experiment subject control, and allows for minimal experimenter involvement.

We shall represent the state of nature with the minimal institution baseline game (no covenants/no swords), then present findings from three environments with institutional configurations: (1) an imposed communication mechanism, (2) an imposed sanctioning mechanism, and (3) an imposed communication mechanism with either an imposed sanctioning mechanism or an opportunity to choose a sanctioning mechanism. Figure 1 gives an overview of the experimental design. Table 2 presents a summary of results across designs, focusing on the average net yield from the CPR (Market 2) as a percentage of maximum possible net yield.[14]

TABLE 1

Experimental Design Baseline Parameters for a Given Decision Period

EXPERIMENT CHARACTERISTICS	EXPERIMENT TYPE: LOW-ENDOWMENT	EXPERIMENT TYPE: HIGH-ENDOWMENT
Number of subjects	8	8
Individual token endowment	10	25
Production function, Market 2[a]	$23(\Sigma x_i) - .25\ (\Sigma x_i)^2$	$23(\Sigma x_i) - .25(\Sigma x_i)^2$
Market 2 return/unit of output	1¢	1¢
Market 1 return/unit of output	5¢	5¢
Earnings/subject at group max.[b]	91¢	83¢
Earnings/subject at Nash equil.	66¢	70¢
Earnings/subject at 0 net yield	50¢	63¢

[a] Σx_i = the total number of tokens invested by the group in Market 2. The production function shows the number of units of output produced in Market 2 for each level of tokens invested therein. Market 2 represents a CPR because the total output from Market 2 (and each subject's share of that output) is a function of the investment levels of all subjects. [b] In the high-endowment design, subjects were paid in cash one-half of their "computer" earnings. This maintained potential experimental profits at near equal levels across designs. Amounts shown are potential cash payoffs.

THE BASELINE GAME: NO COVENANTS AND NO SWORDS

For comparison with the designs discussed later, the baseline experiments can be represented as an iterated series (of 20 rounds) of the constituent game X. At the beginning of each experimental session, subjects were told that (1) they would make a series of investment decisions, (2) all individual investment decisions were anonymous to the group, and (3) they would be paid their individual earnings (privately and in cash) at the end of the experiment. Subjects then proceeded at their own pace through a set of instructions summarized as follows:

> Subjects faced a series of decision rounds in which they were endowed with a specified number of tokens, which they invested between two markets. Market 1 was described as an investment opportunity in which each token yielded a fixed (constant) rate of output and each unit of output yielded a fixed (constant) return. Market 2 (the CPR) was described as a market that yielded a rate of output per token dependent upon the total number of tokens invested by the entire group. The rate of output at each level of group investment was described in functional form, as well as tabular form. Subjects were informed that they would receive a level of output from Market 2 that was equivalent to the percentage of total group tokens they invested. Further, subjects knew that each unit of output from Market 2 yielded a fixed (constant) rate of return.[15]

Subjects knew with certainty the total number of decision makers in the group, total group tokens, and that endowments were identical. Subjects knew that the more the group invested in Market 2, the lower the average and marginal returns from Market 2 would be. They did not know the exact number of decision rounds.

In the baseline experiments, subjects participated in a series of 20 decision rounds. After each round, subjects were shown a display that recorded (1) their profits in each market for that round, (2) total group investment in Market 2, and (3) a tally of their cumulative profits for the experiment. During the experiment, subjects could request, through the com-

FIGURE 1.

Experimental design

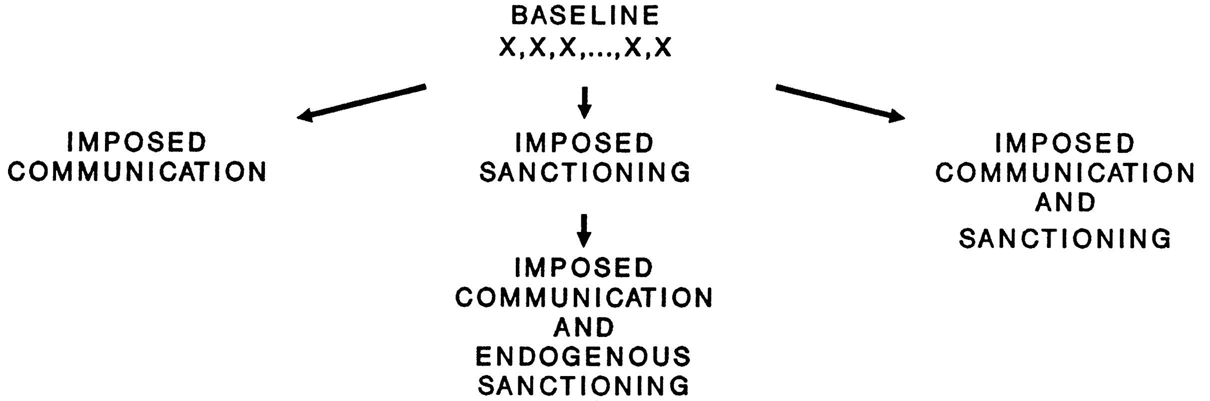

TABLE 2

Summary Results: Average Yield as a Percentage of Maximum

	ROUND					
EXPERIMENTAL DESIGN	1–5	6–10	11–15	16–20	21–25	26+
Baseline						
10TK[a]	51.5	34.7	34.4	35.6	37.1	29.6
25TK	−42.5	−12.4	10.3	32.0	—	—
One-shot communication 25TK	−40.9	−12.7	74.1	45.4	42.5	58.6
Repeated communication						
10TK	32.6	27.3	97.2	98.4	100.0	—
25TK	32.5	−14.4	74.1	75.0	68.9	—
Sanction 25TK	−35.7	−39.6	40.1	38.8	28.7	—
One-shot communication sanction 25TK	−.7	−27.0	86.8	86.3	82.5	77.8
One-shot communication 25TK						
No sanction chosen[b]	46.4	41.2	91.7	61.9	14.7	—
Sanction chosen[b]	−16.9	−5.1	92.5	91.6	89.9	93.8
Sanction chosen[c]	—	—	96.8	97.0	96.7	90.4

[a]TK corresponds to tokens per subject.
[b]Communication and sanctioning choice occurred after round 10.
[c]Communication and sanctioning choice occurred after round 1; the table displays this data beginning in round 11 for comparison purposes.

puter, this information for all previous rounds. Subjects received no information regarding other subjects' *individual* investment decisions or the number of rounds.

As shown in Table 2, in the first 10 rounds, the low-endowment treatment leads to an average net yield of 43%. In the high-endowment treatment, net yield averages −28%. This shows the importance of the endowment treatment out of equilibrium. In the last five rounds of the baseline experiments, the two treatments are very close—36% and 32%, respectively. These percentages are fairly close to the 38% predicted by the subgame consistent equilibrium, evidencing strategic learning on the part of the subjects.

Several characteristics of the individual experiments are important. Investments in Market 2 are characterized by a pulsing pattern: yield is reduced, investors reduce their investments in Market 2, and yields increase. For the high-endowment experiments, the low points in the pulsing pattern were at yields far below zero. There was some tendency for the variance in yields to decrease over the course of the experiments. At the aggregate level, the Nash prediction best describes our data. However, at the individual level we observed no experiment in which individual investments stabilized at the Nash equilibrium. In short, as predicted by both Hobbes's state-of-nature theory and subgame consistent equilibrium, individuals acting independently in a CPR game without communication or sanctioning do not solve the collective action problem they confront.

COVENANTS WITHOUT A SWORD: COMMUNICATION ALONE

Prior experimental evidence strongly supports the conclusion that communication in social dilemmas increases the frequency with which players choose and sustain joint income-maximizing strategies even when individual incentives conflict with such strategies. We shall examine the results from several types of communication experiments.

One-Shot Communication

Our first communication design was parallel to that of the high-endowment baseline game for the first 10 repetitions of the constituent game.[16] At the end of the tenth round, the subjects were informed that they would have a single opportunity of 10 minutes to discuss the decision problem. The instructions were:

> Some participants in experiments like this have found it useful to have the opportunity to discuss the decision problem you face. You will be given ten minutes to hold such a discussion. You may discuss anything you wish during your ten-minute discussion period, with the following restrictions: (1) you are not allowed to discuss side payments; (2) you are not allowed to make physical threats; (3) you are not allowed to see the private information on anyone's monitor.

After this opportunity to communicate, the subjects returned to the constituent game, which was then repeated up to 22 more times. The structure of the one-shot communication experiment is

$$X, X, X, \ldots, X, C, X, X, \ldots, X.$$

The subgame consistent and subgame perfect equilibrium outcome for the one-shot communication game is for each individual to invest 8 tokens in the CPR, the same as in the baseline. The maximum net yield is obtained if a total of 36 tokens are invested. Subjects are not allowed to invest fractional tokens; and the symmetric strategy to obtain the maximum return is halfway between everyone's investing 4 tokens and everyone's investing 5 tokens. If the subjects were to decide to invest either 4 or 5 tokens each, they would obtain 99% of maximum net yield in either case.

The transcripts of the discussion during the communication round reveal that subjects perceived their problem as involving two tasks: determining the maximal yield available and agreeing upon a strategy to achieve that yield. In all three groups, an agreement was reached; but no group found the optimal solution. Averaging over all experiments, yields were negative throughout the first 10 rounds (−27%). Once communication was allowed, net yields jumped to an average of 74% for the five rounds immediately following the communication time-out, then gradually decayed thereafter. The average net yield in the postcommunication phase was 55%. Defining a *defection* as an individual Market 2 investment greater than that agreed upon, we observe a defection rate of 25% during the decision rounds following communication. These results suggest that a single communication period enables participants to begin the process of adopting a joint strategy. However, the inability to communicate more often limits the durability of their agreements.

Repeated Communication

Our second design involves repeated communication in both the low- and high-endowment settings. At the outset, the constituent game was repeated for 10 rounds. After round 10, the subjects read an announcement similar to that for one-shot communication but informing them they would have an opportunity for discussion after *each* subsequent round. The subjects left their terminals and sat facing one another.[17] The structure of the repeated communication experiment is

$$X, X, \ldots, X, C\text{–}X, C\text{–}X, \ldots, C\text{–}X.$$

These experiments provide strong evidence for the power of repeated face-to-face communication. Subjects successfully used the opportunity to (1) calculate coordinated yield-improving strategies, (2) devise verbal agreements to implement these strategies, and (3) deal with nonconforming subjects through verbal statements. When allowed to communicate repeatedly, subjects greatly enhanced their joint yield and sustained this enhancement. In the *low*-endowment environment, net yields averaged 99% of optimum in the repeated communication phase with a 5% defection rate.[18] The *high*-endowment CPR game is a more challenging decision environment than the low-endowment one. While the subgame consistent equilibrium outcomes of the two games are identical, the disequilibrium implications of the 25-token (high-endowment) game change considerably. With 25 tokens, as few as three subjects investing all of their tokens can essentially ruin the CPR (bring returns below w), while with 10 tokens, it takes seven out of eight subjects to accomplish this much damage. Net yields in the repeated communication phase averaged only 73% of optimum, with a 13% defection rate.

Repeated communication enabled subjects to discuss defections and to cut the defection rate on agreements significantly. In all communication experiments, subjects offered and extracted promises of cooperation, thereby increasing their joint yield significantly above that obtained prior to communication. Discussions went well beyond discovering the level of investments that would generate maximum yields. A striking aspect of the discussion rounds was how rapidly the subjects, who had not had an opportunity to establish a well-defined community with strong internal norms, were able to devise their own agreements and verbal punishments for those who broke those agreements. These verbal sanctions had to be directed at unknown defectors, since subjects' decisions were anonymous. Subjects detected defection solely through aggregate investments. In many cases, statements like "Some scum-bucket is investing more than we agreed upon" were a sufficient reproach to change defectors' behavior. However, verbal sanctions were less effective in the 25-token environment.

The evidence from the one-shot and the high-endowment designs suggests why individuals in many field settings may not rely only on face-to-face communication. When repeated discourse is infeasible or when the actions of one or a few individuals can be a strong disequilibrating force, individuals who have the capacity to agree to sanction one another, as well as communicate with one another, might well want to reinforce their covenants with their own swords. While subgame consistency predicts that individuals in such settings would not sanction one another, endogenous sanctioning is frequently observed in field settings. We now turn to an examination of whether sanctioning behavior will occur in our laboratory environment and its effects on behavior.

SWORD WITHOUT COVENANTS: SANCTIONING ALONE

Experiments in this design began like high-endowment baseline experiments with the exception that after each round, subjects received *individual* data of all decisions.[19] This information was given by subject number, thus maintaining anonymity. Our sanctioning mechanism required that each subject incur a cost (a fee) in order to sanction another. In our first sanctioning design, after round 10, subjects were given an announcement summarized as follows:

> Subjects were informed that in all remaining rounds each would be given the opportunity to place a fine. Each subject could levy one fine at a specified fee. The subject fined would pay a fine of a specified amount. It was possible for a single subject to be charged multiple fines. After each round, each subject filled out a fining form. These forms were collected and tallied by the experimenter, who then reported the results privately to each subject. Note that any subject who was fined did not know the identity of those who imposed the fine. At the end of the experiment, the experimenters subtracted from subjects' total profits the total of all fees and all fines.

The actual fees ranged from 5¢ to 40¢, the fines from 10¢ to 80¢ cents. The fee-to-fine ratio was either .25 or .50. After subjects read the announcement, we answered questions regarding the implementation of the procedure. No discussion was held on why the subjects might want to use the procedure or its possible consequences. This created an experimental setting as close as possible to the noncooperative assumptions of no communication and no capacity to engage in enforceable agreements. The structure of the imposed sanctioning design is

X, X, . . . , X–S, X–S, X–S, . . . , X–S, X.

The principle results from our sanctioning experiments are summarized as follows.[20] Across all eight experiments, net yield rose from −38% before the imposition of sanctioning to 37% after. When one subtracts the costs of fees and fines, however, average net yield increased to only 9%. Thus, sanctioning alone is not an efficient institution. Besides these quantitative results, we draw the following qualitative conclusions:

1. Significantly more sanctioning occurs than predicted by subgame consistency, and the frequency is inversely related to cost.
2. Sanctioning is primarily focused on heavy CPR investors.
3. There is a nontrivial amount of sanctioning that can be classified as error, lagged punishment, or "blind" revenge.

We observed 176 instances of sanctioning across the eight experiments. In no experiment did we observe fewer than 10 instances. The frequency of sanctions was inversely related to the cost of imposing the fine and dramatically increased with the stiffness of the fine. Further, our results, although reminiscent of strategy 2, did not strictly support the conclusion that subjects were playing an equilibrium of this form. Except for one experiment, where net yield was over 95%, yields were too low, and sanctioning levels too high, to be consistent with imperfect equilibria.

The second and third results relate to the reasons for sanctioning. From postexperiment interviews and personal observations, we offer four explanations for the higher-than-predicted level of sanctioning:

1. *One-round punishment.* The person fined was the highest, or one of the highest, investors in the previous round.
2. *Lagged punishment.* The person fined was one of the highest investors in the CPR in a round prior to the previous one.
3. *Blind revenge.* The person fined was a low CPR investor and was fined by a person fined in a previous round.
4. *Error.* No obvious explanation can be given for the action (trembling hand).

In summary, 77% of all sanctioning was aimed at investors who in the previous round were above-average investors in the CPR. An additional 7% was aimed at subjects who had been heavy investors in the CPR in earlier (but not the most recent) rounds. We classify 5% as blind revenge and the remaining 11% as errors.[21]

The evidence from these experiments suggests why individuals in field settings might not want to rely on sanctioning alone. Individuals who have the capacity to sanction one another without the ability to communicate about joint strategies and the use of sanctions face an insuperable handicap to increasing efficiency. Finally, we return to the question whether communication and sanctioning together foster sustainable high yields.

COVENANTS WITH A SWORD: COMMUNICATION AND SANCTIONING

Our last two decision environments investigate the consequences of combining a one-shot opportunity to communicate with either an experimenter-imposed sanctioning mechanism or an opportunity to decide whether or not to adopt a sanctioning mechanism endogenously. These experiments began like those in the design with sanctioning alone. After round 10, subjects were given an announcement that they would have a single 10-minute discussion period. In experiments with an imposed-sanctioning mechanism, subjects were also given an announcement (prior to discussion) similar to that of the sanctioning-alone environment. In experiments where subjects had an opportunity to choose a sanctioning mechanism, the announcement informed them that at the end of 10 minutes they would vote on whether to institute a sanctioning mechanism and if so, what the level of fines would be. The only restriction on the sanctioning mechanism was that the fee-to-fine ratio was ½. The voting rule was strict majority, with the status quo a repeated baseline experiment without a sanctioning mechanism.

Imposed Communication and Sanctioning

The structure of the three experiments in this design is

X, X, . . . , X, C, X–S, X–S, . . . X–S, X.

In the first experiment, the participants rapidly focused on the problem of deciding upon a joint investment strategy. They spent most of their 10 minutes calculating various options to ensure that they had discovered an optimal strategy. They decided to invest 4 tokens each in Market 2 and the remaining 21 tokens in Market 1. Further, they agreed to fine one another if anyone put more than 4 tokens in Market 2. One subject said, "If everyone puts in 4 tokens, we are going to be making 42 cents more money in the individual accounts. This is the highest." Another said, "Does everyone agree to this? OK, now we have agreed that everyone will put 4 tokens in and if anyone puts any more in, we are all going to fine them. Is that all agreed, now?" With this specific agreement to which everyone nodded assent, the subjects returned to their terminals and made investments for 16 more rounds without a single defection nor any use of the sanctioning mechanism. They obtained 98% of maximum net yield from the CPR and did not waste resources on fees or fines.

In the second experiment, the subjects did not find the optimal strategy but devised a complex rotation system to ensure that they all received what they thought would be maximal returns. They decided that each subject would invest 6 tokens in Market 2 and that the right to invest 2 more tokens each round would be rotated first to Subject 1, then to Subject 2, on through Subject 8. One subject suggested that they not fine at all, but another argued: "No, let's fine anyone who breaks our rules. If they break our rules, then we should fine 'em!" After further discussion, the subjects agreed that they would use the sanctioning mechanism to fine anyone who deviated from "their rules." For two rounds they kept their agreement. On round 13, one subject invested 7, rather than 6, and was immediately fined by one of the other subjects. On round 19, two subjects invested one more token than agreed upon and each was immediately sanctioned by another subject. No more defections were attempted and no more fees were paid to assess fines. In this experiment, the subjects achieved 86% of maximum net yield (since they had miscalculated the optimum). Their net return dropped to 79% when fees and fines were deducted from their earnings.

In the third experiment, the subjects never discussed the possibility of devising a joint strategy even though they mentioned how the overinvestment of some of the subjects during the first 10 rounds had made it difficult for the rest of them. The closest they got to an agreement was to discuss fining those who were obviously overinvesting, for example, "those who invested over 21 tokens in Market 2." A considerable amount of their discussion time was wasted in awkward silence. They finally asked whether they had to sit there the entire 10 minutes. After verifying that no subject wanted to use the remaining 2 minutes of their time for further discussion, the experimenters let the subjects return to their terminals. Following the communication period, the subjects achieved an average of 70% of net yield from the CPR, up from −14% in prediscussion rounds. A total of twenty 40¢ fees were paid to impose the same number of 80¢ fines on other subjects. The fines were directed toward subjects who had invested heavily in Market 2 in the prior round. Net yield fell to 24% with fees and fines deducted from earnings.

Imposed Communication and Endogenous Sanctioning

In order to make the choice of a sanctioning mechanism meaningful, subjects in this decision environment were randomly drawn from the pool of subjects from our imposed sanctioning design. The structure of the four experiments in this design is

X, X, . . . , X, C, X, X, . . . , X or

X, X, . . . , X, C, X–S, X–S , . . . , X–S, X.

Subjects decided upon a joint investment strategy and established a sanctioning mechanism in only two of the four experiments. In one they chose a fee-to-fine ratio of 10¢/20¢ and in the other a ratio of 20¢/40¢. Net yield averaged 91% of maximum following the communication round. The level of defection from their agreements was very low. Net yield fell slightly to 86% when fees and fines were deducted.

In one of the experiments where subjects rejected the adoption of a sanctioning mechanism, they did adopt an agreed-upon joint strategy to invest four tokens in the CPR (a joint strategy very close to optimal). In the first few rounds following discussion, they sustained their agreement; but by round 16, four subjects invested more than their agreement. From then on, the level of defections steadily rose. Overall, they had an average net yield of 36% and a defection rate of 42%, representing the highest defection rate of any communication experiment where an agreement was reached. The other experiment in which a sanctioning mechanism was not chosen is an outlier. Across all of our CPR experiments with the baseline institution, this is the only experiment where yields in the first 10 rounds were essentially maximal. When given the opportunity to discuss the decision problem and choose a sanctioning mechanism, the group agreed that they did not need a mechanism and that no one should try to get "greedy," that is, invest too much in the CPR. The group held together for a few rounds, after which yields began a gradual decline. This decline was due primarily to a gradual increase in Market 2 investments by two subjects. By round 25, net yield had dropped to 56% of optimum. Net yield for all rounds following the discussion session averaged 76%, compared to 90% prior to round 10.

We have traced back to the specific sanctioning/no communication experiment in which each of these subjects participated. Of the 32 subjects in these four experiments, 18 voted for, and 14 voted against, the implementation of a sanctioning mechanism. Of the 14 who voted *no*, 11 had previously participated in a sanctioning experiment with a fee-to-fine ratio of 20¢/80¢. Of the 18 who voted *yes*, only 3 had been in a 20¢/80¢ design. We infer from this result that the high level of sanctioning activity in the 20¢/80¢ design, the lack of overall efficiency gains and the presence of blind revenge combined to impede the willingness of participants to choose a sanctioning mechanism.

It is possible that the experience of the first 10 rounds of the constituent game had an effect on mechanism choice. To examine this possible hysteresis effect, two additional experiments were conducted. In these two experiments, the opportunity to communicate and to adopt a sanctioning mechanism was available at the outset. The structure of these two experiments is

C, X, X, . . . , X or C, X–S, X–S, . . . , X–S, X.

In both of these experiments, the subjects quickly agreed to an investment strategy and a sanctioning mechanism to punish defectors. Across the two experiments, net yields averaged 95%—94% with fees and fines included.

The payoff consequences of selecting or not selecting a sanctioning mechanism were very different across the experiments in this design. The groups choosing some form of sanctioning institution earned average yields of 93% in the postdiscussion phase. Indeed, yields this high suggest that this set of institutions, endogenously chosen, approximate the conditions necessary for a cooperative game. The groups not choosing some form of sanctioning institution earned average yields of only 56%, with serious decay.

The results from this set of communication and sanctioning experiments suggest that some subjects can find yield-improving joint strategies, design a sanctioning mechanism, use the sanctioning mechanism, and achieve a high rate of conformance to their joint strategies. On the other hand, prior negative experience with institutions that individuals view as punitive and inefficient is conducive neither to the design of better institutions nor to a willingness to use them.

CONCLUSIONS: SELF-GOVERNANCE IS POSSIBLE

The inconsistency between predicted results and observed behavior in prior research stimulated this research. In one-shot and finitely repeated games, communication alone is not predicted to have an effect on behavior. Earlier experimental studies of social dilemmas have, however, shown that communication alone leads to more efficient outcomes. We confirm these results in complex CPR environments.

With regard to communication alone, we obtain the following results (summarized in Table 3):

1. In the low-endowment CPR environment, average net yield increased from 35% (when no communication was allowed) to 99% (when communication was allowed on a repeated basis).
2. In the high-endowment CPR environment, average net yield increased from 21% (when no communication was allowed) to 55% (when communication was allowed only once) to 73% (when communication was allowed on a repeated basis).

These results raise several puzzles. Because the subgame consistent equilibrium prediction for games with communication (one-shot or repeated) is the same as that for games without communication, opportunities for "mere jawboning" should make no difference; yet they do. Further, the equilibrium prediction is the same for both low- and high-endowment environments. The high-endowment environment, however, exhibits lower net yield and fosters less effective communication. The finding that CPR appropriators in the field invest substantial resources in sanctioning activities stimulated our exploration of sanctioning behavior in the laboratory. With an imposed sanctioning institution and no communication we find the following:

3. Subjects are willing to pay a fee to place a fine on another subject far more than was predicted.
4. In the high-endowment environment, average net yield increases from 21% with no sanctioning to 37% with sanctioning. When the costs of fees and fines are subtracted from average net yield, however, net yield drops to 9%.

Thus, subjects overuse the sanctioning mechanism, and sanctioning without communication reduces net yield. The finding that CPR appropriators in the field invest substantial resources in both communication and sanctioning activities stimulated our exploration of these joint behaviors in the laboratory. Examining only the high-endowment environment, we find the following:

5. With an imposed sanctioning mechanism and a single opportunity to communicate, subjects achieve an average net yield of 85%. When the costs of fees and fines are subtracted, average net yield is still 67%. These represent substantial gains over the baseline, where the net yield averaged 21%.
6. With the right to choose a sanctioning mechanism and a single opportunity to communicate, subjects who adopt a sanctioning mechanism achieve an average net yield of 93%. When the costs of fees and fines are subtracted, average net yield is still 90%. In addition, the defection rate from agreements is only 4%.
7. With the right to choose a sanctioning mechanism and a single opportunity to communicate, subjects who do not adopt a sanctioning mechanism achieve an average net yield of only 56%. In addition, the defection rate from agreements is 42%.

Thus, subjects who use the opportunity to communicate to agree to a joint investment strategy and

TABLE 3

Aggregate Results All Designs (%)

EXPERIMENTAL DESIGN	AVERAGE NET YIELD CPR	AVERAGE NET YIELD CPR (MINUS FEES AND FINES)	DEFECTION RATE
Baseline			
10TK[a]	34	—	—
25TK	21	—	—
One-shot communication 25TK	55	—	25
Repeated communication			
10TK	99	—	5
25TK	73	—	13
Sanction 25TK	37	9	—
One-shot communication sanction 25TK	85	67	1
One-shot communication 25TK			
No sanction chosen	56	—	42
Sanction chosen	93	90	4

Note: All computations are for periods in which the treatment was in effect. Nash equilibrium for all designs is 39% net CPR yield.
[a]TK corresponds to tokens per subject.

choose their own sanctioning mechanism achieve close-to-optimal results. This is especially impressive in the high-endowment environment, where defection by a few subjects is very disruptive. For those who predict cooperation in repeated settings based on trigger strategies, our findings are not supportive. In no experiment where one or more subjects deviated from an agreed-upon joint strategy did the subjects then follow a trigger strategy of substantially increasing their investments in the CPR.[22] In fact, in some experiments where one or more subjects deviated from an agreed-upon joint strategy, some subjects subsequently *reduced* their investments in the CPR. When subjects discussed the problem of how to respond to one or more free-riders, they overtly rejected the idea of dumping all of their tokens into the CPR.[23] To return to our starting point, namely, Hobbes's assertion, these experiments suggest that covenants, even without a sword, have some force, while swords without a covenant may be worse than the state of nature. Best of all the conditions we examined are covenants with an *internal* sword, freely chosen or made available as an institutional option.[24]

Two major implications follow from the results. The first relates to policy analysis. Policymakers responsible for the governance and management of small-scale, CPRs should *not* presume that the individuals involved are caught in an inexorable tragedy from which there is no escape. Individuals may be able to arrive at joint strategies to manage these resources more efficiently. To accomplish this task, they must have sufficient information to pose and solve the allocation problems they face. They must also have an arena where they can discuss joint strategies and perhaps implement monitoring and sanctioning. In other words, when individuals are given an opportunity to restructure their own situation, they frequently—but not always—use this opportunity to make credible commitments and achieve higher joint outcomes without an external enforcer. We cannot replace the determinate prediction of *no cooperation* with a determinate prediction of *always cooperate*. Our findings challenge the Hobbesian conclusion that the constitution of order is only possible by creating sovereigns who then must govern by being above subjects, by monitoring them, and by imposing sanctions on all who would otherwise not comply.[25]

The second major implication relates to behavioral theory. In finitely repeated social dilemma experiments, a wide variety of treatments that do not change the theoretically predicted subgame consistent equilibrium outcomes do change subjects' behavior. This raises a *substantial* question whether our subjects conceptualize their decision task in the way theorists do. For instance, if subjects believe the game is being repeated according to some exogenous probabilistic mechanism, then there are equilibria supporting more cooperative behavior if the subjective continuation probability is not too low. Or it may be that subjects are acting as boundedly rational players in the sense of Selten, Mitzkewitz, and Uhlich (1988). In this case, the observed improvement in yield could be the result of boundedly rational equilibrium, as Selten, Mitzkewitz, and Uhlich observe in a duopoly context. We intend to explore these intriguing possibilities in future research.

Notes

Earlier versions of this paper were presented during the 1991 meetings of the Midwest Political Science Association, the Political Economy Forum, the Bonn Workshop in Mathematical Economics, the American Political Science Association, the International Association for the Study of Common Property, and the Economic Science Association. We would like to thank Ray Battalio, William Bianco, Randall Calvert, Robert Dorsey, Werner Güth, Roberta Herzberg, John Ledyard, Thomas Palfrey, Charles Plott, Hugh Ward, and Rick Wilson for their useful comments. Financial support from National Science Foundation grant SES–4843901 and the U.S. Department of Agriculture (Cooperative Agreement 43–3AEM–1–80078) is gratefully acknowledged. The views expressed here do not necessarily reflect the views of the funding agencies. The assistance of Dean Dudley in the experimental lab is also much appreciated. All data are stored on permanent NOVANET disk files. Send inquiries to Professor James M. Walker, Department of Economics, Ballantine 901, Indiana University, Bloomington, Indiana 47405.

1. Harsanyi and Selten add that in real life, "agreements may be enforced externally by courts of law, government agencies, or pressure from public opinion; they may be enforced internally by the fact that the players are simply unwilling to violate agreements on moral grounds and know that this is the case" (1988, 3). To model self-commitment using noncooperative game theory, the ability to break the commitment is removed by trimming the branches that emanate from a self-commitment move to remove any alternative contrary to that which has been committed. In a lab setting, this would mean changing the structure of the alternatives made available to subjects after an agreement, which was not done.

2. See, e.g., Campbell 1985; Hardin 1971, 1982; McLean 1981; Moore 1987; Taylor 1987. The term *social dilemma* is introduced by Dawes 1975, 1980.

3. See Bianco and Bates 1990 for a theoretical analysis of the capabilities and limits of assigning leaders strong sanctioning powers and Samuelson et al. 1986 for an experimental investigation of the choice of a Leviathan-like mechanism to solve social dilemmas.

4. See Gardner, Ostrom, and Walker 1990 and E. Ostrom, Gardner, and Walker n.d. for a detailed exposition of the terms we use and the models we have developed.

5. See Berkes 1989; National Research Council 1986; E. Ostrom 1990; Pinkerton 1989; and Wade 1988 for a discussion of successful, failed, and fragile efforts to self-organize and govern, small-scale CPRs. Michael Hechter's (1987) recent synthesis of self-organizing capabilities within many social groups provides evidence across other domains (see also Williamson 1975, 1985). See Kreps 1990 for a useful synthesis of literature relevant to endogenous commitments within firms.

6. The extensive literature on one-shot communication in dilemma games includes Bornstein and Rapoport 1988; Bornstein et al. 1989; Braver and Wilson 1984, 1986; Caldwell 1976; Dawes, McTavish, and Shaklee 1977; Dawes, Orbell, and van de Kragt 1984; Dawes, van de Kragt, and Orbell 1988; Edney and Harper 1978; Jerdee and Rosen 1974; Kramer and Brewer 1986; Orbell, Dawes, and van de Kragt 1990; Orbell, van de Kragt, and Dawes 1988, 1991; van de Kragt et al. 1983, 1986.

7. For a discussion of relevant literature, see Isaac and Walker 1988, 1991; E. Ostrom and Walker 1991.

8. See n. 5.

9. Investment in the CPR beyond the maximum net level is termed *rent dissipation* in the literature of resource economics. One can interpret this environment as a limited access CPR (see, e.g., Clark 1980; Cornes and Sandler 1986; Negri 1989).

10. This specification actually has a number of other possible interpretations. For instance, if one defines $F(\Sigma x_i)/\Sigma x_i = y$ and y to be a public good, then one has the payoff functions for a voluntary contribution mechanism as in Isaac and Walker 1988. Alternatively, one can define y in the same expression to be an externality, in which case one has payoff functions for Plott's experiments on externalities in product markets (Plott 1983). For further details, see Ledyard 1991.

11. During recruitment, subjects are told they will participate in a one-to-two hour decision-making experiment. Although the exact endpoint is not revealed, it is explicitly bounded from above. Further, all subjects have experienced the boundedness of an experiment that lasted between 10 and 30 rounds.

12. We use the word *fine* not in the context of redistribution. What is crucial here is that real resources are used up, not just redistributed, by efforts to sanction.

13. Besides the symmetric imperfect equilibria we have given, there are many asymmetric equilibria. Take any permutation of the identity matrix, e.g., 1 sanctions 2, 2 sanctions 3, etc. Then these permuted sanctions also support the same outcome as strategy 2 does. Notice that if mistakes are what cause deviations, then an equilibrium like strategy 2 will generate n fines every time a mistake takes place—considerably more than the zero fines generated by the subgame consistent equilibrium strategy 1.

14. Net yield accrued as a percentage of maximum = (return from Market 2 − opportunity costs of tokens invested therein)/(return from Market 2 at the optimum − opportunity costs of tokens invested therein). Opportunity costs equal the potential return that could have been earned by investing the tokens in Market 1.

15. A complete set of instructions is available upon request. In high-endowment experiments, subjects were informed that their cash payoff would be one-half of the "lab dollars" earned in the experiment. This was done so that total payments in high- and low-endowment treatments would be approximately equal. At the end of all experiments, subjects were paid privately (in cash) their individual earnings. All subjects had participated previously in an experiment using the constituent game environment. The number of rounds in earlier experiments had varied from 10 to 20. Subjects were recruited randomly from this pool of experienced subjects to ensure that no prior experimental group was brought back intact. We provide a detailed account of behavior in the constituent game environment in Walker, Gardner, and Ostrom 1990, 1991.

16. In the one-shot communication experiments, subjects received information on individual decisions after each round to facilitate our comparing results in designs with sanctioning either alone or in conjunction with one-shot communication added. Information was given by subject number, thereby preserving anonymity. This added information had no significant impact on observed yields.

17. As in the one-shot communication setting, each person was identified with a badge that was unrelated to their player number. This facilitated player identification in our transcripts. If unanimous, players could forgo discussion.

18. We also conducted a series of costly communication experiments, in which subjects had to pay in advance for opportunity to communicate. See E. Ostrom and Walker 1991 for a detailed discussion. See also Isaac and Walker 1988, 1991.

19. Earlier experiments focusing on sanctioning mechanisms without communication include Yamagishi 1986, 1988.

20. A comparison of the initial 10 rounds to the initial 10 rounds of baseline suggests that the addition of anonymous information about individual decisions had no impact on investments.

21. A second set of sanctioning experiments was conducted as a check on the robustness of our original design. Readers of our earlier results conjectured that the lack of a significant improvement in net yield with the introduction of a sanctioning mechanism in our initial design could be due to a hysteresis tied to the decisions in the first 10 periods, periods in which there was no sanctioning mechanism. In our second design, three new experiments were conducted, with the sanctioning mechanism introduced prior to the first decision period. The fee-to-fine ratio was 40¢/80¢. Subjects used fines repeatedly in all three experiments. The results from our second set of experiments are consistent with those from our first design. There was no persistent yield-improving behavior that could be tied to the introduction of the sanctioning mechanism. In fact, when costs of fees and fines are incorporated we found a negative impact on net benefits.

22. It is well known that inducing trigger strategy behavior in subjects is extremely difficult. For a recent attempt, see Sell and Wilson 1991.

23. One player resisted the suggestion made by another

player to dump all of the tokens into the CPR by stating, "We screw ourselves, too."

24. We did not explore the effect of an external agent assigned responsibility to monitor and sanction behavior because in this research we were interested in the feasibility of internal enforcement.

25. See V. Ostrom 1987, 1989, 1991 for an elucidation of an alternative theory to Hobbes's.

References

Berkes, Fikret, ed. 1989. *Common Property Resources, Ecology, and Community-based Sustainable Development*. London: Belhaven.

Bianco, William T., and Robert H. Bates. 1990. "Cooperation by Design: Leadership, Structure, and Collective Dilemmas." *American Political Science Review* 84:133–47.

Bornstein, Gary, and Amnon Rapoport. 1988. "Intergroup Competition for the Provision of Step-level Public Goods: Effects of Preplay Communication." *European Journal of Social Psychology* 18:125–42.

Bornstein, Gary, Amnon Rapoport, Lucia Kerpel, and Tani Katz. 1989. "Within- and Between-Group Communication in Intergroup Competition for Public Goods." *Journal of Experimental Social Psychology* 25:422–36.

Braver, Sanford L., and L. A. Wilson. 1984. "A Laboratory Study of Social Contracts as a Solution to Public Goods Problems: Surviving on the Lifeboat." Presented at a meeting of the Western Social Science Association, San Diego.

Braver, Sanford L., and L. A. Wilson. 1986. "Choices in Social Dilemmas: Effects of Communication within Subgroups." *Journal of Conflict Resolution* 30(1):51–62.

Caldwell, Michael D. 1976. "Communication and Sex Effects in a Five-Person Prisoners' Dilemma Game." *Journal of Personality and Social Psychology* 33(3):273–80.

Campbell, R. 1985. "Background for the Uninitiated." In *Paradoxes of Rationality and Cooperation*, ed. R. Campbell and L. Sowden. Vancouver: University of British Columbia Press.

Clark, Colin W. 1980. "Restricted Access to Common-Property Fishery Resources: A Game-theoretic Analysis." In *Dynamic Optimization and Mathematical Economics*, ed Pan-Tai Liu. New York: Plenum.

Cornes, Richard, and Todd Sandler. 1986. *The Theory of Externalities, Public Goods, and Club Goods*. Cambridge: Cambridge University Press.

Dawes, Robyn M. 1975. "Formal Models of Dilemmas in Social Decision Making." In *Human Judgement and Decision Processes*, ed. Martin F. Kaplan and Steven Schwartz. New York: Academic.

Dawes, Robyn M. 1980. "Social Dilemmas." *Annual Review of Psychology* 31:169–93.

Dawes, Robyn M., Jeanne McTavish, and Harriet Shaklee. 1977. "Behavior, Communication, and Assumptions about Other People's Behavior in a Commons Dilemma Situation." *Journal of Personality and Social Psychology* 35(1): 1–11.

Dawes, Robyn M., John M. Orbell, and Alphons J. C. van de Kragt. 1984. "Normative Constraint and Incentive Compatible Design." University of Oregon, Eugene. Typescript.

Dawes, Robyn M., Alphons J. C. van de Kragt, and John M. Orbell. 1988. "Not Me or Thee but WE: The Importance of Group Identity in Eliciting Cooperation in Dilemma Situations—Experimental Manipulations." *Acta Psychologica* 68: 83–97.

Edney, Julian J, and Christopher S. Harper. 1978. "The Commons Dilemma: A Review of Contributions from Psychology." *Environmental Management* 2(6): 491–507.

Gardner, Roy, Elinor Ostrom, and James Walker. 1990. "The Nature of Common-Pool Resource Problems." *Rationality and Society* 2:335–58.

Hardin, Russell. 1971. "Collective Action as an Agreeable *N*-Prisoner's Dilemma." *Behavioral Science* 16(5): 472–81.

Hardin, Russell. 1982. *Collective Action*. Baltimore: Johns Hopkins University Press.

Harsanyi, John C., and Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge: Massachusetts Institute of Technology Press.

Hechter, Michael. 1987. *Principles of Group Solidarity*. Berkeley: University of California Press.

Hobbes, Thomas. 1960. *Leviathan*. Oxford: Basil Blackwell.

Isaac, R. Mark, and James M. Walker. 1988. "Communication and Free-riding Behavior: The Voluntary Contribution Mechanism." *Economic Inquiry* 24(4): 585–608.

Isaac, R. Mark, and James M. Walker. 1991. "Costly Communication: An Experiment in a Nested Public Goods Problem." In *Laboratory Research in Political Economy*, ed. Thomas R. Palfrey. Ann Arbor: University of Michigan Press.

Jankowski, Richard. 1990. "Punishment in Iterated Chicken and Prisoner's Dilemma Games." *Rationality and Society* 2(4): 449–70.

Jerdee, Thomas H., and Benson Rosen. 1974. "Effects of Opportunity to Communicate and Visibility of Individual Decisions on Behavior in the Common Interest." *Journal of Applied Psychology* 59(6): 712–16.

Kramer, R. M., and Marilyn M. Brewer. 1986. "Social Group Identity and the Emergence of Cooperation in Resource Conservation Dilemmas." In *Experimental Social Dilemmas*, ed. Henk A. Wilke, David M. Messick, and Christel G. Rutte. Frankfurt: Lang.

Kreps, David M. 1990. "Corporate Culture and Economic Theory." In *Perspectives on Positive Political Economy*, ed. James E. Alt and Kenneth A. Shepsle. New York: Cambridge University Press.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory* 27:245–52.

Ledyard, John O. 1991. "Is There a Problem with Public Good Provision?" California Institute of Technology, Pasadena. Mimeo.

McLean, Iain. 1981. "The Social Contract in Leviathan and the Prisoner's Dilemma Supergame." *Political Studies* 29(3): 339–51.

Moore, Clement Henry. 1987. "Prisoners' Financial Dilemmas: A Consociational Future for Lebanon?" *American Political Science Review* 81:201–18.

Nash, John F. 1950. "The Bargaining Problem." *Econometrica* 18:155–62.

Nash, John F. 1951. "Non-cooperative Games." *Annals of Mathematics* 54:286–95.

National Research Council. 1986. *Proceedings of the Conference on Common Property Resource Management*. Washington: National Academy Press.

Negri, D. H. 1989. "The Common Property Aquifer as a Differential Game." *Water Resources Research* 25:9–15.

Ophuls, W. 1973. "Leviathan or Oblivion." In *Toward a Steady State Economy*, ed. H. E. Daly. San Francisco: Freeman.

Orbell, John M., Robyn M. Dawes, and Alphons J. C. van de Kragt. 1990. "The Limits of Multilateral Promising." *Ethics* 100(4): 616–27.

Orbell, John M., Alphons J. C. van de Kragt, and Robyn M. Dawes. 1988. "Explaining Discussion-induced Cooperation." *Journal of Personality and Social Psychology* 54(5): 811–19.

Orbell, John M., Alphons J. C. van de Kragt, and Robyn M. Dawes. 1991. "Covenants without the Sword: The Role of Promising in Social Dilemma Circumstances." In *Social Norms and Economic Institutions*, ed. Ken Koford. Ann Arbor: University of Michigan Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.

Ostrom, Elinor, Roy Gardner, and James Walker. N.d. *Rules and Games: Institutions and Common-Pool Resources*. Ann Arbor: University of Michigan Press.

Ostrom, Elinor, and James Walker. 1991. "Communication in a Commons: Cooperation without External Enforcement."

In *Laboratory Research in Political Economy*, ed. Thomas R. Palfrey. Ann Arbor: University of Michigan Press.
Ostrom, Vincent. 1987. *The Political Theory of a Compound Republic: Designing the American Experiment*. 2d rev. ed. Lincoln: University of Nebraska Press.
Ostrom, Vincent. 1989. *The Intellectual Crisis in American Public Administration*. 2d ed. Tuscaloosa: University of Alabama Press.
Ostrom, Vincent. 1991. *The Meaning of American Federalism: Constituting a Self-governing Society*. San Francisco: Institute for Contemporary Studies Press.
Pinkerton, Evelyn, ed. 1989. *Co-operative Management of Local Fisheries: New Directions for Improved Management and Community Development*. Vancouver: University of British Columbia Press.
Plott, Charles R. 1983. "Externalities and Corrective Policies in Experimental Markets." *Economic Journal* 93:106–27.
Samuelson, Charles D., David M. Messick, Henk A. M. Wilke, and Christel G. Rutte. 1986. "Individual Restraint and Structural Change as Solutions to Social Dilemmas." In *Experimental Social Dilemmas*, ed. Henk A. M. Wilke, David M. Messick, and Christel G. Rutte. Frankfurt: Lang.
Sell, Jane, and Rick Wilson. 1991. "Levels of Information and Contributions to Public Goods." *Social Forces* 70(1): 107–24.
Selten, Reinhard. 1971. "A Simple Model of Imperfect Competition Where Four Are Few and Six Are Many." *International Journal of Game Theory* 2:141–201.
Selten, Reinhard, Michael Mitzkewitz, and Gerald R. Uhlich. 1988. "Duopoly Strategies Programmed by Experienced Players." University of Bonn, Discussion Paper B-172. Special Research Project 303.
Taylor, Michael. 1987. *The Possibility of Cooperation*. Cambridge: Cambridge University Press.
van de Kragt, Alphons J. C., et al. 1983. "The Minimal Contributing Set as a Solution to Public Goods Problems." *American Political Science Review* 77:112–22.
van de Kragt, Alphons J. C., et al. 1986. "Doing Well and Doing Good as Ways of Resolving Social Dilemmas." In *Experimental Social Dilemmas*, ed. Henk A. M. Wilke, David M. Messick, and Christel G. Rutte. Frankfurt: Lang.
Wade, Robert. 1988. *Village Republics: Economic Conditions for Collective Action in South India*. Cambridge: Cambridge University Press.
Walker, James, Roy Gardner, and Elinor Ostrom. 1990. "Rent Dissipation in a Limited-Access Common-Pool Resource: Experimental Evidence." *Journal of Environmental Economics and Management* 19:203–11.
Walker, James, Roy Gardner, and Elinor Ostrom. 1991. "Rent Dissipation and Balanced Deviation Disequilibrium in Common Pool Resources: Experimental Evidence." In *Game Equilibrium Models II*, ed. Reinhard Selten. Berlin: Springer.
Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.
Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.
Yamagishi, Toshio. 1986. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology* 51(1): 110–16.
Yamagishi, Toshio. 1988. "Seriousness of Social Dilemmas and the Provision of a Sanctioning System." *Social Psychology Quarterly* 51(1): 32–42.

Elinor Ostrom is Arthur F. Bentley Professor of Political Science and co-director of the Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington, IN 47505-3186.

James Walker is Associate Professor of Economics and Faculty Associate, Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington, IN 47405-3186.

Roy Gardner is Professor of Economics, Cooperating Scientist ERS-USDA, and Faculty Associate, Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington, IN 47405-3186.