E 627 Comments No. 7[•] by Frank P. Maier-Rigaud

"Covenants with and without a Sword" is an extension of "Rent Dissipation in a Limited-Access Common-Pool Resource: Experimental Evidence" (abbreviated as WGO in the following) in which the same CPR games were conducted using a non-cooperative institutional structure. The main emphasis of the present paper is to discuss alternative CPR institutions introducing communication and sanctioning mechanisms. The experimental results described in WGO were used as baseline experiment to test the following 5 variations of the baseline game:

1. Covenants without a Sword: Communication alone
   a) One-shot Communication
   b) Repeated Communication
2. Sword without Covenants: Sanctioning alone
3. Covenants with a Sword: Communication and Sanctioning
   a) Imposed Communication and Sanctioning
   b) Imposed Communication and Endogenous Sanctioning

Before presenting the results of these variations and the ensuing conclusion, it is useful to recall possible theoretical predictions.

Since the Pareto efficient solution, obtained by maximizing the wealth of all participants as a whole, as well as the total rent dissipation (TRD) solution, obtained in open-access CPR environments, have already been discussed in detail in my discussion of WGO, I will focus here on the specific assumptions of non-cooperative game theory and the expected Nash equilibrium. Non-cooperative game theory, as well as Hobbes's *Leviathan*, emphasize the weakness of mere words and the necessity of an external authority to enforce agreements. In cooperative games, players can communicate freely and make enforceable agreements; in non-cooperative games, they can do neither. This leads to the game theoretic prediction that as long as the payoff function is not changed, the Nash equilibrium cannot be eliminated by mere communication. These theoretical assumptions have far reaching consequences for local governance issues as empirically demonstrated by Ostrom (1990) in which individuals facing social dilemmas are capable of developing credible ex ante commitments without relying on external authorities. This empirical evidence contrasting to theoretical predictions, as well as experimental evidence showing major divergences from theoretical predictions led to the series of experiments presented in this paper.[1]

Communication without a sword

Based on previous experimental results showing 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, two "pure" communication experiments were run. The first one-shot communication experiment was identical to the baseline high-endowment game with the exception that after 10

[•] Critical summary of: Covenants With and Without a Sword: Self-Governance is possible, by J. Walker, R. Gardner and E. Ostrom, American Political Science Review 86(2), 1992, p.404-417. Literature used:

Walker, J., Gardner, R., Ostrom, E. (1990) Rent Dissipation in a Limited-Access Common-Pool Resource: Experimental Evidence, in Journal of Environmental Economics and Management, 19, p. 203-211

Ostrom, E. (1990) Governing the Commons, The Evolution of Institutions for Collective Action, Cambridge University Press

[1] The experimental anomalies from the point of view of non-cooperative game theory are the nontrivial improvement possibilities of outcomes through communication even in one-shot social dilemma experiments and the improvements in joint outcomes in repeated games reached solely through repeated communication. The third anomaly stems mainly from field settings where considerable investments in monitoring and sanctioning have been observed.

rounds the players had a single opportunity to discuss the decision problem.[2] After this communication opportunity of 10 minutes, the participants returned to the game and played for up to 22 more rounds. The subgame consistent and subgame perfect equilibrium outcome for the one-shot communication game was for each participant to invest 8 tokens in the CPR.[3] The Pareto optimum is at a total of 36 tokens and therefore the symmetric Pareto strategy is between 4 and 5 tokens invested by every participant. Since fractional tokens were not allowed, subjects could "only" obtain 99% of the maximum net yield investing either 4 or 5 tokens.

The second design involved repeated communication in both high- and low-endowment settings. After ten rounds identical to the baseline game, subjects were informed that they would have the possibility to talk after each subsequent round, following the same instructions as in the one-shot communication period. This institutional setting provides strong support for the power of face-to-face communication. Subjects used the time to communicate coordinated yield maximizing strategies and deal with defectors through verbal statements. As a result, 99% of maximum net yield was appropriated in the low-endowment setting with a defection rate of 5%. The performance was lower in the high-endowment game because each subject was endowed with 25 tokens as compared to 10 and therefore as few as three subjects (as opposed to seven) investing all their tokens in the CPR could essentially ruin it.[4] With a 13% defection rate, net yields averaged only 73% of the optimum in the high-endowment game.

The results in those two distinct institutional settings, only allowing communication, suggests why individuals in field settings may not want to rely solely on face-to-face communication even though social pressure in small communities, as a more subtle form of sanction, may already provide a sufficiently high incentive to keep people from defecting.

Sword without Covenants

The experiments conducted with sanctioning alone were set up in analogy to the previously discussed experiment, with the difference that communication was replaced by the possibility to sanction. After the tenth round of the high-endowment game, subjects received individual data on the decisions made by all subjects in the previous round and the possibility to sanction for all consecutive rounds.[5] Sanctioning was costly and in order to sanction another participant, half or one fourth of the sanction sum, depending on the design, was incurred as sanctioning cost by the initiator. Actual fees ranged from 5 to 40 cents, fines ranged from 10 to 80 cents accordingly.

Experimental results showed that sanctioning alone is not an efficient institution in this environmental setting. Even though across all experiments net yield rose from -38% to 37%, most of the gains were lost when sanctioning costs and fees were subtracted (net yield increased to only 9%). Besides these quantitative results, the authors drew the following qualitative conclusions: Significantly more sanctioning occurs than predicted by subgame consistency with a frequency inversely related to cost. Sanctions are primarily focused on heavy CPR investors (77%), but there remains a non-trivial amount of sanctioning possibly explained as error (11%) (trembling hand), lagged punishment (7%) or "blind" revenge (5%). An average of 22 sanctions per experiment with no observations below 10 led to yields that were too low to be consistent with imperfect equilibrium except for one outlier.

[2] During the communication opportunity the players were explicitly told that they were not allowed to discuss side payments, to make physical threats or to see other participants' private information.
[3] Subgame perfection and subgame consistency are two equilibrium refinement principles. If the game has a unique equilibrium, it also has a unique subgame perfect and subgame consistent equilibrium.
[4] This is due to the fact that the subgame consistent equilibrium outcome does not change with a change in the token endowment.
[5] Individual data on decisions made was given by subject number in order to maintain anonymity.

The evidence from the conducted experiments suggests that individuals with the capacity to sanction one another but without the ability to communicate and devise joint strategies face an insurmountable handicap to increasing net yields.

Covenants with a Sword

These 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. The experiments all began like those with "pure" sanctioning. After round ten, subjects were told they would have a single 10 minute discussion period. In experiments where subjects could choose the sanctioning mechanism themselves (as long as the fee to fine ratio of .5 was kept) were informed prior to the discussion period that at the end of the ten minutes they would have to vote (strict majority) on whether or not to institute a sanctioning mechanism and what the level of the fines should be.

In the first of the three experiments with imposed communication and sanctioning, the participants rapidly focused on the problem of deciding upon a joint investment strategy. They decided to invest 4 tokens in the CPR (Market 2) and 21 tokens in Market 1. They also agreed to fine one another if anyone put more than 4 tokens in Market 2. With this specific agreement, the subjects made investments for 16 more rounds without defection nor any use of the sanctioning mechanism with a result of 98% of maximum yield from the CPR. The second experiment was less successful because the participants did not find the optimal maximizing returns. They agreed, however, to fine everybody who would deviate from the agreed upon investment as dictated by "their rules". In the following rounds, a few defections were attempted but punished directly so that from round 19 on no more defections occurred. Due to their calculation error and the few defections, subjects achieved only 86% of the maximum net yield. Their return dropped to 79% when the fines and fees were subtracted. The third experiment is interesting insofar as it shows that the factor "subject" or "participant" cannot be held constant in experiments. The group constellation already played a role in the previous experiment where obviously nobody knew how to calculate the optimal amount of token contributions to Market 2, but in this group, the greatest deficiency was the incapability of the group to interact, probably due to lacking personalities capable of starting a discussion. As a result, the group could improve as compared to prediscussion rounds (-14%) but only reached 24% after fines and fees were deducted from earnings. This particular experiment makes clear that the possibility of communication is a necessary but not sufficient condition for improved results through coordination and formation of mutually accepted rules. As long as this possibility is not used efficiently by the group, results are not improved substantially.

The last experiments conducted left the choice to the group to decide whether to use a sanctioning mechanism or not and if yes, with what amounts. The structure of the experiments was therefore partly in the hands of the participants who could decide between a baseline game with a ten minute one-shot communication period after round ten or a baseline game with a ten minute one-shot communication period after round ten and consecutive sanctioning possibilities at the agreed upon fines and fees (again the ratio was given) after each following round. Interestingly, subjects decided upon a joint investment strategy and established a sanctioning mechanism (fee to fine relationships of 10/20cents and 20/40cents) in only two of the four experiments. In order to explain this behavior, the authors traced the individual subjects back to the specific sanctioning/no communication experiment in which each of the subjects participated. Of the 32 subjects in the four experiments, 18 voted for and the rest voted against the implementation of a sanctioning mechanism. Of the 14 who voted against a sanctioning mechanism, 11 had

previously participated in a sanctioning experiment with a fee-to-fine ratio of 20/80 cents. Of the 18 who voted yes, only three had been in this design. The authors explain the low interest in a sanction mechanism by the fact that the experience gained in an environment of high sanctioning activity combined with blind revenge and therefore low net yields probably led the individuals to shy away from sanctioning in general. This hypothesis is very important since it suggests participant conditioning. Subject's constitutional preferences seem to be based on a trial and error search process as opposed to rational calculations. As the authors point out, prior negative experience with institutions that individuals view as punitive and inefficient is neither conducive to design better institutions nor to a willingness to use them. In order to test for a similar conditioning effect in the previous ten rounds on the mechanism choice and to be able to rule out this type of "path dependent" behavior, two additional experiments were conducted in which communication and sanctioning was available right from the beginning. Quick agreement on investment strategy and sanctioning mechanisms to punish defectors was found and net yields averaging 94-95% including fees and fines were reached.

All in all, the experiments conducted lead to the conclusion that the role of sanctioning increases as the game progresses away from the round where an agreement was reached. In the two experiments where no sanctioning mechanism was installed[6], an increase in investments in Market 2 is observed after a certain amount of rounds with almost no defections. Once this process is realized by all the participants the net yield decreases steadily in a "domino-effect" in which more and more people defect.

Conclusion

Generally one can conclude that Pareto optimal results are approximated by introducing an institution where communication is possible. In contrast to game theoretical assumptions, communication leads to the highest improvements if a choice between sanctioning and communication has to be made.[7] Only after higher gains are reached through communication, a sanctioning mechanism might be useful in preventing the erosion of this high level and guaranteeing the ongoing conformity to the agreed upon contributions. Indeed, cheap talk is a very cheap instrument, capable of raising returns substantially at zero costs (at least in the experiments conducted here), despite and in contradiction to the fact that theoretical predictions do not vary with the introduction of communication. This is equally true for high- and low-endowment environments. The high-endowment environment, however, exhibits lower net yield and fosters less active communication that raises another puzzle. Another hypothesis potentially motivated by the observed pulsing pattern of investments in Market 2 analyzed in WGO, could not be sustained. Trigger mechanisms to achieve cooperation as for example the famous Tit for Tat from Axelrod could not be observed. As a matter of fact, reactions to overinvestment by some subjects led others to reduce their investment below the agreed upon level.

All in all, the experiments suggest that the best outcome can be achieved in a game of covenants with an internal sword, freely chosen or made available as an institutional option.

Two major implications follow from the results. The first one adds experimental evidence to the general thrust of empirical evidence presented in Ostrom (1990). Namely that policymakers responsible for the governance and management of small-scale CPRs should not presume that the individuals involved are caught in an inexorable

[6] In one experiment the subjects voted against one and in the other they derived from good initial experiences that such a mechanism would not be necessary. This again is particularly interesting because it shows that individuals tend to rely more on experience than on the rationale of the situation.

tragedy as described as a general case by Hardin, but that individuals may be capable to arrive at joint strategies to manage these resources more efficiently. The second major implication is of fundamental importance to behavioral theory. A wide variety of treatments that did not change the theoretically predicted equilibrium outcomes did change subjects' behavior and therefore empirical outcomes. This raises the question whether rational choice theory is a rational choice in modeling such behavior or if alternative concepts as, for example, Selten's bounded rationality or Skinner's operant conditioning combined with Popper's active search process guided by conjectures and refutations represent viable alternatives.

[7] Actually, due to oversanctioning when communication is not allowed, the sanction mechanism reduces net yield. Oversanctioning might be explained by the hypothesis that people try to communicate using the sanction mechanism in order to reach a common strategy.