W99-20 10/14/99

COLLECTIVE ACTION AND THE EVOLUTION OF SOCIAL NORMS

Elinor Ostrom

Co-Director
Workshop in Political Theory and Policy Analysis
Center for the Study of Institutions, Population, and Environmental Change
Indiana University, Bloomington

Earlier versions of this paper have been presented at seminars at the Workshop in Political Theory and Policy Analysis, Indiana University; Department of Political Science, Gothenburg University; and the Beijer Institute of Ecological Economics, at the Royal Swedish Academy of Sciences in Stockholm. I appreciate the helpful comments made by Iris Bohnet, Juan-Camilo Cardenas, Bruno Frey, Roy Gardner, Steffen Huck, Fabrice Lehoucq, Frank Maier-Rigaud, Mike McGinnis, Jimmy Walker, and the outstanding editing by Patty Dalecki. Support by the Ford Foundation and the MacArthur Foundation is gratefully acknowledged.

Workshop in Political Theory and Policy Analysis

Indiana University, 513 North Park
Bloomington, IN 47408-3895 USA

Phone: (812) 855-0441
Fax: (812)855-3150

workshop@indiana.edu
www.indiana.edu/~workshop

Elinor Ostrom

COLLECTIVE ACTION AND THE EVOLUTION OF SOCIAL NORMS

> . . . unless the number of individuals in a group is quite small, or unless there is coercion or some other special device to make individuals act in their common interest, *rational, self-interested individuals will not act to achieve their common or group interests.* (Olson, 1965: 2, author's emphasis)

With the publication of *The Logic of Collective Action* in 1965, Mancur Olson challenged a cherished foundation of modern democratic thought that groups would form and take collective action whenever members jointly benefitted. Olson's provocative assertion that no one would contribute to the provision of a public good—the *zero contribution thesis*—was soon derived as the predicted equilibrium of an N-person Prisoner's Dilemma (PD) game (Hardin, 1971, 1982). The N-person PD game—and social dilemmas more generally—are viewed as the canonical representations for collective-action problems (Lichbach, 1996). The zero contribution thesis, however, contradicted observations of everyday life in mature democracies that many people voted, did not cheat on their taxes, and contributed effort to voluntary associations.

Extensive fieldwork has by now established that many individuals in all walks of life and all parts of the world voluntarily organize themselves to gain the benefits of trade, to provide mutual protection against risk, and to create and enforce rules that protect natural resources.[1] Empirical research in natural settings thus challenges the validity of the *zero* contribution thesis. On the other hand, field research confirms that the temptation to free ride on the provision of collective benefits is a universal problem. In all recorded, long surviving, self-organized resource governance regimes, participants invest resources in monitoring the actions of each other so as to reduce the probability of free riding (Ostrom, 1990).

For all of the work, empirical findings have not yet been integrated into a revised theory of collective action. Thirty-five years of extensive empirical research could be summarized with the weak statement that "some groups do and some groups do not succeed in overcoming social dilemmas to achieve collective action." A theory for how self-organized groups can survive for long periods of time, if they develop governance arrangements consistent with a set of design principles, has been proposed and is supported by further empirical research (Ostrom, 1990; Morrow and Hull, 1996; Asquith, 1999; Bardhan, 1999; Lam, 1998). A theory for how groups overcome the base problems they face has not, however, achieved general acceptance.

An immense number of contextual variables are identified by empirical researchers as conducive or detrimental to endogenous collective action. Among those proposed are: the type of production and allocation functions; the predictability of resource flows; the relative scarcity of the good; the size of the group involved; the heterogeneity of the group; the dependence of the group on the good; common understanding of the group; the size of the total collective benefit; the marginal contribution by one person to the collective good; the size of the temptation to free ride; the loss to cooperators when others do not cooperate; having a choice of playing or not; the presence of leadership; past experience and level of social capital; the autonomy to make binding rules; and a wide diversity of rules that are used to change the structure of the situation.

Several scholars have proposed theoretical syntheses of empirical findings regarding the factors affecting the likelihood of groups overcoming social dilemmas (Baland and Platteau, 1996; Ostrom, 1992, forthcoming; Marwell and Oliver, 1993; de Janvry, McCarthy, and Sadoulet, 1998). Substantial theoretical dispute, however, remains over the impact of contextual variables. In particular, the impact of the size of a group and internal heterogeneity are frequently considered important contextual variables, but the direction of their impact and how they operate is strongly contested (Olson, 1965; Hardin, 1982; Keohane and Ostrom, 1995; Agrawal, forthcoming; Molinas, 1998; Leach, Mearns, and Scoones, 1999; Agrawal and Gibson, 1999). While these contested variables are repeatedly identified as important in field research, they have not systematically been integrated into mainstream theoretical analyses. Size, for example, plays no role in the analysis of a linear public good game. The prediction of zero contribution holds whether there are 10, 100, or 1,000 participants.

Consequently, a substantial hiatus exists between the micro foundations of noncooperative game theory and the rich world of empirical research in field settings. Many scholars agree that something is wrong with the theoretical foundations, and serious work is underway to provide an alternative micro-theory of individual choice that can explain anomalous findings.[2] In this article I will discuss the necessity of assuming the existence of multiple types of players—rational egoists, as well as players who use social norms—in models of nonmarket behavior. I will use an indirect evolutionary approach to explain how multiple types of players could survive, and even flourish, in social dilemma situations. The success of those who adopt social norms strongly depends on their capacity to identify one another. Thus, contextual variables that enhance the knowledge that players have about each other's past behavior are theoretically strong candidates to include in future efforts to explain the origin of collective action.

I will first address the sufficiency of the standard model of rational choice to explain behavior in collective-action situations. To do this I will examine how predictions from this theory have repeatedly been tested in laboratory experiments. Experimental research is an important tool for examining the micro-foundations of a theory because one can design experiments to include specific variables while carefully controlling for others. Experimental research has now generated a body of facts that cannot be explained using the standard theory of rational choice.

These anomalies can be explained, however, when one assumes the existence of multiple types of players, including those who adopt social norms of behavior. In the anonymously competitive environment of an open market, rational egoists are the only type of players to survive. Thus, a theory of individual choice based entirely on one type of player has been long successful in predicting and explaining behavior in competitive markets (Alchian, 1950). In nonmarket settings, when users of social norms can identify one another, norm-users can survive and even flourish. Thus, focusing on how contextual variables help those who use social norms identify one another, is the theoretical lense that will eventually facilitate the integration of micro-theory and findings from empirical resource about the effect of contextual variables on overcoming social dilemmas.

The Sufficiency of the Standard Model of Rational Choice to Explain Behavior in Collective-Action Situations

The Standard Assumptions of Individual Choice

Most studies by political economists assume a standard model of rational action—what I will call a rational egoist. Using this model to generate predicted outcomes in a linear, public good game, for example, the normal assumption is that utility is a linear function of individual earnings:

$$(1) \qquad U_i = U_i\,[(E - x_i) + A \cdot P\,(\textstyle\sum x_i)]$$

where E is an individual endowment of assets, x, is the amount of this endowment contributed to provide the good, A is the allocation formula used to distribute the group benefit to individual players, and P is the production function. In a linear public good game, A is specified as 1/N and $0 > P > 1$ (but both of these functions vary in other types of collective action). So long as $P < 1$, contributing to the collective good is never an optimal strategy for a fully self-interested player.

The optimal outcome for the group of players as a whole is for everyone to contribute all of their endowments to provide the public good. The unique equilibrium for a single-shot game, however, is that everyone contributes zero since each individual has access to benefits without paying any costs. Contributing zero in every round is also the predicted equilibrium for the finitely repeated game. These predictions are based on the assumptions that all players are fully rational and interested only in their own immediate financial payoff. It is also assumed that (1) all players have complete information and common knowledge of the exogenously fixed structure of the game, (2) all players believe that all of the other players are fully rational, and (3) no external actor is available to enforce agreements among players.

When used to model outcomes of collective-action situations, the standard model of a rational egoist generates clear and precise predictions. It is thus possible to design experiments that closely match the objective condition of the theoretical models. Experimental economics is based on an assumption that a experimenter can induce the preferences of a formal game by paying subjects according to a payoff schedule that is common knowledge (unless an experiment is specifically designed to examine the effect of private information) (Smith, 1982). Experiments that have used this assumption in testing predictions based on the rational egoist model in auctions and competitive market situations have substantiated the explanatory power of the theory (see Kagel and Roth, 1995 for a summary). In regard to various types of collective-action situations, on the other hand, the results have been entirely different.

Testing Predictions in Linear Public Good and Common-Pool Resource Experiments

The first public good experiments were undertaken by Dawes, McTavish, and Shaklee (1977) who found that subjects contributed significantly more than zero. These experiments were soon followed by those of Marwell and Ames (1979, 1980, 1981) who found that subjects in one-shot experiments contributed about one-half of their endowment to the public good and retained the rest for themselves. In response to these findings, Isaac, Walker, and Thomas (1984) and Isaac, McCue, and Plott (1985) designed the voluntary contribution mechanism, which has become the standard for conducting experiments of linear public goods. The extensive number of studies using variations of this design have been synthesized in Davis and Holt (1993), Ledyard (1995), and Offerman (1997). Consistently, subjects contribute between 40 and 60% of their endowments to the public good in a one-shot game as well as in the first round of finitely repeated games. After the first round, contribution levels remain higher than predicted but tend to decay downward. While the average contributions never reach zero, a repeated finding is that the vast majority (over 70%) of subjects contribute nothing in the announced last round of a finitely repeated sequence (Fehr and Schmidt, forthcoming).

The decay toward the predicted outcome led to a speculation that this simply reflected initial subject confusion and that by the tenth round (which was the last round of the initial experiments), players learn the equilibrium (see, for example, Ledyard, 1995). The speculation is not supported by further evidence. In experiments where the game is halted after some rounds to change design conditions, contribution levels substantially rise (rather than continuing to fall) immediately after the experiment is re-started (Andreoni, 1988; Croson, 1996; Keser and van Winden, 1996). In a clear test of this speculation, Isaac, Walker, and Williams (1994) repeated the same game for 10 rounds, 40 rounds, and 60 rounds with experienced subjects who were specifically told the end point of each design. As shown in Figure 1, the rate of decay is inversely related to the number of decision rounds. Subjects learn something other than the predicted equilibrium strategy. They appear to learn how to cooperate at a moderate level for ever longer periods of time!

[Figure 1 about here]

While outcomes in an institutionally stark common-pool resource game are more consistent with standard theory than outcomes in public good experiments, subjects in both common-pool resource and public good games approach optimal outcomes once they are allowed to communicate on a face-to-face basis (Ostrom, Walker, and Gardner, 1992). The strong effect of communication is not consistent with currently accepted theory. Verbal agreements in these experiments are not enforced. Thus, communication is only as "cheap talk" and makes no difference in predicted outcomes in social dilemmas (Farrell, 1987). When individuals are allowed to talk, the presumption is that they will use the opportunity to try to convince others to contribute but then not contribute at all themselves (Farrell and Rabin, 1996).

And yet, one of the strongest and most frequently replicated findings is that substantial increases in the levels of cooperation happen when individuals are allowed to communicate on a face-to-face basis (see Ledyard, 1995). Instead of using this opportunity to fool others into cooperating, subjects use the time to discuss the optimal strategy, to extract promises from one another, and to give verbal tongue-lashings when aggregate contributions fall below promised levels. Sally (1995) conducted a meta-analysis of more than 100 experiments involving over 5,000 subjects and found that opportunities for communication in one-shot experiments significantly and dramatically raised cooperation rates. When communication is implemented by allowing subjects to signal promises to cooperate through their computer terminals, however, much less cooperation occurs than in experiments allowing face-to-face communication (Sell and Wilson, 1991; Rocco and Warglien, 1996; Rocco, 1998; Palfrey and Rosenthal, 1988; Roth and Murnighan, 1982). On the other hand, allowing subjects to see the person with whom they were playing, as contrasted to a completely anonymous situation, reduced the choice of dominant strategies (no contribution) from 64% to 41% in a four-person, PD game (Bohnet and Frey, 1999). Face-to-face communication was even more effective, leading only 4% of the subjects to follow the dominant strategy.

A rational egoist in a public good game should not in any way be affected by a belief regarding the contribution levels of others. The dominant strategy is a zero contribution no matter what others do. An oft-replicated finding, however, is that the willingness of subjects to contribute to the provision of a public good is strongly correlated with their expectations of the behavior of others (Dawes, McTavish, and Shaklee, 1977; Dawes, Orbell, and van de Kragt, 1986; Orbell and Dawes, 1991; Brandts and Schram, 1995: Croson, 1998). In a provision point experiment, for example, all those who contributed sufficient funds to produce the benefit predicted that sufficient funds would be provided by others to make their own contribution superfluous (Messick, 1999). It is also the case that when subjects are given an opportunity to choose to enter or not enter a dilemma situation, those who intend to contribute are more willing to enter these transactions (Orbell and Dawes, 1993).

Another strong finding is the impact of endogenous sanctioning. Standard theory makes a clear prediction related to this second-order social dilemma (Oliver, 1980). No one is predicted to spend anything to punish others since the positive impact of such an action is shared equally with others whether or not they contribute. In the laboratory, the reverse is true. In both public good (Fehr and Gachter, 1998) and common-pool resource (Ostrom, Gardner, and Walker, 1994) experiments, subjects given an opportunity to use costly sanctioning mechanisms exercise this option frequently. The use of sanctions increases the level of cooperation. Some subjects even punish low contributors in the announced final round of an experiment when punishment could no longer affect future behavior (Fehr and Gachter, 1998). In CPR experiments, subjects reacted to endogenous punishment options as a good with a downward-sloping demand curve. Punishment rates were higher when the cost to the sanctioner was lower or when the effectiveness (amount of the fine) was higher (Ostrom, Gardner, and Walker, 1994).

McCabe, Rassenti, and Smith (1996) show that players are willing to receive a lower payoff in order to punish partners who do not reciprocate a signalled intention to arrive at an outcome where benefits are shared equally. Yamagishi (1986) found that cooperation levels among Japanese subjects rose significantly when subjectscould contribute to a "punishing fund" used to fine individuals who contributed the least. Also interesting is the finding that subjects who had the lowest trust scores on a previously administered survey gave significantly more to the punishment fund and then were able to achieve the highest level of cooperation. By the last round they were contributing 90% of their endowment to the public good (see Yamagishi, 1988 for a replication with U.S. subjects). Thus, experiments conducted in the U.S., Switzerland, and Japan show that individuals are willing to contribute to a second-order public good, that endogenous sanctioning systems make a difference in the first-order public good, and that those who are initially the least trusting are more willing to contribute to sanctioning systems and respond more to a change in the structure of the game.

Testing Predictions of the Standard Model of the Individual

In addition to experimental research that has examined the level of cooperation achieved in collective-action settings, research on ultimatum, dictator, and trust games directly examines whether most individuals behave as the standard model of a rational egoist predicts. In an ultimatum experiment, subjects are assigned two roles: Proposer and Responder. The subjects are told that a sum of money will be divided between them if they both agree to this distribution. The Proposer is instructed to make an offer of how to divide this sum. The Responder can then accept or reject the offer. If the offer is rejected, neither player receives any money. Two predictions are derived from the standard noncooperative analysis of this game. The Proposer should offer the Responder as small a unit as is feasible (thus, retaining most of the funds). The Responder should accept any amount above zero.

Neither prediction finds empirical support in repeated experiments conducted in Germany, Indonesia, Israel, Japan, Slovenia, Slovakia, and the United States (Camerer and Thaler, 1995). While modest differences do occur in the average sum offered or rejected in different cultures (Girth, Schmittberger, and Schwarze, 1982; Roth et al., 1991; but see Okada and Riedl, 1999), the overwhelming finding is that Proposers tend to offer around half of the sum and Responders tend to reject low but positive offers. In a meta-analysis of ten experiments involving 875 Proposers, Fehr and Schmidt (forthcoming) report that over 70% of the Proposers make offers between 40 and 50% of the allocated sum. Increasing the amount of money involved does not produce the predicted behavior, but the proportion of "fair" splits is modestly reduced. Hoffman, McCabe, and Smith (1996), for example, found that 83% of the Proposers splitting ten dollars offered between four and five dollars, while 74% given the opportunity to split $100 offered between $40.00 and $50.00. When a sample of Indonesian subjects were assigned a provisional sum of three months wages (200,000 Rp), 56% of the Proposers allocated between 40 and 50% of this very substantial sum to the Responder (Cameron, 1995).

The structure of a dictator experiment is similar, but the Responder can no longer reject an offer. The prediction of standard theory is precise: Proposers should keep all of the money for themselves. And, again, behavior is not consistent with the prediction. The proportion of the sum that is shared in dictator experiments tends to be less than in ultimatum experiments, but most Proposers give away a substantial portion of the funds they are allocated (Davis and Holt, 1993: 263-69). Forsythe, Horowitz, and Savin (1994), for example, found that only 20% of Proposers assigned zero to the Responders, while 80% gave between 10 and 50% of the funds they were assigned to the Responder, Even in experiments where extreme care is taken to ensure anonymity, as well as to reduce doubts that subjects may have about the veracity of the experimental set up itself, only 40% of the Responders took all of the funds. About one-third of the Proposers allocated between 40 and 60% to the Responders (Frohlich, Oppenheimer, and Moore, 1998).

In a trust or gift-exchange game, the trustor can voluntarily give resources, which increase by some multiple, to a trustee. After knowing whether the trustor has extended trust or not, the trustee can keep the funds or transfer funds, which again increase by some multiple, back to the trustor. These games resemble a sequential PD game where each player has more than two actions. The prediction for such a game is that both players will give the other zero. The vast majority of trustors, however, do transfer some funds and many of the trustees reward them by returning funds to them (Fehr, Kirchsteiger, and Riedl, 1993; Berg, Dickhaut, and McCabe, 1995). In a sequential PD game, Watabe et al. (1996), Smith (1997), and Hayashi et al. (1999) all found that a substantial proportion of first movers (around one-half) trusted the second mover enough to take a cooperative first move and that the trust was most frequently repaid by the selection of a cooperative move by the second player (around three-quarters reciprocating). None of these findings can explained by the standard model.

The Facts to be Explained

These strong results have been replicated so frequently under controlled circumstances that further theoretical analyses need to explain how the following could occur.

- About half of the subjects cooperate in one-shot, or the first round of public good games.
- Subjects keep cooperation levels relatively high for long sequences of finitely repeated public good games even though over 70% of them contribute zero in the last period whenever it is scheduled.
- Face-to-face communication in all types of social dilemmas produces substantial increases in cooperation that are sustained across all periods including the last period.

- Those who believe others will cooperate in social dilemmas are more likely to cooperate themselves.
- Subjects expend personal resources in order to punish those who make below-average contributions to a collective benefit, including the last period of a finitely repeated game.
- Proposers tend to offer Responders about half of the funds in ultimatum games and are frequently turned down if they offer below one-fourth.
- Proposers in dictator games give away substantial amounts even under double-blind conditions.
- Subjects in games of trust allocate substantially more to others than predicted.
- All of these rates are affected by various ways of framing the situation and rules used for assigning participants, increasing competition among them, allowing communication or sanctions, or allocating benefits.

These facts are hard to explain using the standard theory that all individuals who face the same monetary payoffs evaluate decisions the same way based entirely on expected consequences for their own, immediate, material welfare. On the other hand, these facts do not rule out the likelihood that many individuals behave in a manner consistent with the standard theory. Thus, one is forced by well-substantiated facts to adopt a more eclectic (and classical) view of human behavior as involving the possibility of multiple types of players, some of whom derive some utility from intrinsic payoffs.

Adding Multiple Types of Players

Assuming that all individuals follow the advice of the standard theory is no longer tenable given the massive, inconsistent evidence. Assuming the existence of "norm-using" players, in addition to rational egoists, enables one to start making more coherent sense out of the findings. Two norm-using players —"conditional cooperators" and "willing punishers"—appear to have participated in the experiments discussed above along with rational egoists. *Conditional cooperators* are individuals who use a norm of reciprocity and are willing to initiate cooperative action when they estimate others will reciprocate and to repeat these actions as long as a sufficient proportion of the others involved reciprocate. Tit-for-tat is an example of this type of strategy in a two-person social dilemma (Axelrod, 1984, 1986). *Willing punishers* are conditional cooperators who are willing, if given an opportunity, to punish others through verbal rebukes or to use costly material payoffs when available. Willing punishers may become willing rewarders if the circle of relationships allows them to reward those who have contributed more than the minimal level.

Conditional cooperators are the source of the relatively high levels of contributions in one-shot, and the initial rounds of PD and public good games. Their initial contributions may encourage some rational egoists to contribute in order to obtain higher returns in the early rounds of the game (Kreps et al., 1982). Conditional cooperators are more likely to cooperate depending on their belief about the behavior of others, but their tolerance for free riding varies. Those who are the most sensitive to free riding reduce contributions first. Their reduction leads others to curb their contributions. By the last round of a stark public good experiment, all of the rational egoists have stopped contributing and only a few of the most determined conditional cooperators make positive contributions.

In a recent study, Fischbacher, Gachter, and Fehr (1999) motivated subjects in a one-shot linear public good game to complete a contribution schedule as well as making a contribution decision. On the schedule, each subject specified the number of tokens they would give depending on how many tokens everyone else gave. Three types of players were evident in their data. About half of the players were conditional cooperators whose own proposed contributions monotonically increased with the average contributed by others. About one-third were rational egoists who planned to contribute zero tokens no matter what others did. A small set of players proposed to increase their contributions until the average contributions of others was around 10 tokens (out of 20). They then proposed to decrease their contributions steadily as the average contributions of others was posited to increase from 11 to 20. One might interpret this third set as rational egoists encouraging others to cooperate at low levels of contributions and free riding if contributions were relatively high. The conditional cooperators in the Fischbacher, Gachter, and Fehr study tend to have a "self-serving bias." Except for the very lowest levels, they tend to contribute slightly below the average of other's contributions. In a simulation based on these contribution schedules, Fischbacher, Gachter, and Fehr show that strategies of self-serving, conditional cooperators cannot be stable in a repeated situation without communication or sanctioning options. At any expected level of others' contributions, such players will contribute less than the average—thus, drawing down the average. Simulated outcomes based on the contribution schedules obtained in this study replicate the shape of the decay functions shown in Figure 1 very closely.

When participants can communicate on a face-to-face basis, conditional cooperators make judgments about the reliability of others, try to extract promises of coordinated cooperation from everyone as well as threaten to stop their own cooperation unless others reciprocate. Frank, Gilovich, and Regan (1993) found that allowing subjects to have a face-to-face discussion enabled them to predict who would play cooperatively at a rate significantly better than chance. Kikuchi, Watanabe, and Yamagishi (1996) found that those who were rated as having a high degree of trust were able to predict others' behavior more accurately than those with low levels of trust. If some conditional cooperators are also willing punishers, communication provides an opportunity for verbal sanctions and other overt signs of disapproval. Thus, the finding that face-to-face communication is more efficacious than computerized signaling is probably due to the richer language structure available and the added intrinsic costs involved in hearing the intonation and seeing the body language of those who are genuinely angry at free riders. If provided an opportunity to use material resources to punish (or reward) free-riding (or extra effort), willing punishers (or rewarders) will expend extra resources to reinforce the higher outcomes that full cooperation can generate for all and perhaps to teach others that reciprocity is a better overall strategy.

In ultimatum games, Proposers who are conditional cooperators will split the sum in an equitable fashion. All types of players would accept an offer that is close to an equal split. Proposers, who are rational egoists, do not know whether they face another rational egoist who would accept any positive sum, or a conditional cooperator who will reject a low split. Past experience with conditional cooperators would lead a rational egoist to fear a rejection if the offer is too low, and thus to make an offer closer to a fair split to ensure that the proposal is not rejected. Behavior in games of trust can be similarly explained. The most challenging fact to be explained is the large sums assigned by Proposers in dictator games. One can only assume that some conditional cooperators view all life as a repeated game and are willing to allocate a substantial portion of the sum to the helpless Responder, in the sense that they themselves have been and are likely to be the recipient of some future act of kindness.

Evolution. Culture, Learning, and Social Norms

Assuming the presence of norm-using players in addition to rational egoists is a substantial step forward in being able to explain the empirical findings in the field and in the lab. But how are we to account for the presence of multiple types of players? Recent developments in evolutionary theory—including the study of cultural evolution—begin to provide a useful set of answers (Alexander, 1987; Hirshleifer, 1977; Boyd and Richerson, 1985, 1988; Trivers, 1971; Barkow, Cosmides, and Tooby, 1992). Human evolution occurred mostly during the long Pleistocene era that lasted for about 3 million years to about 10,000 years before modern times. This was an era when humans roamed the earth in small bands of hunter-gatherers who were dependant on each other for mutual protection, sharing of food, and providing for the young. Survival was dependent not only on aggressively seeking individual returns but also on solving many day-to-day collective-action problems. Whether these groups were structured in stronger dominance hierarchies or by looser social exchange networks, learning the social norms of a particular group and how to recognize who was using deceit and who was a trustworthy reciprocator was an essential skill. Those of our ancestors who solved these problems most effectively would have had a selective advantage over those who did not.

Evolutionary psychologists who study the cognitive structure of the human brain conclude that humans do not develop general analytical skills that are then applied to a variety of specific problems. Humans are not terribly skilled at general logical problem solving (as any scholar who has taught probability theory to undergraduates can attest). Rather, the human brain appears to have evolved a domain-specific, human-reasoning architecture (Clark and Karmiloff-Smith, 1991). For example, humans use a different approach to reasoning about deontic relationships—what is forbidden, obligated, or permitted—as contrasted to reasoning about what is true and false. When reasoning about deontic relationships, humans tend to check for violations—or cheaters (Manktelow and Over, 1991). When reasoning about whether empirical relationships are true, they tend to use a confirmation strategy (Oaksford and Chater, 1994). This deontic effect in human reasoning has been repeatedly detected and in children as young as three years old, and is not associated with overall intelligence or educational level of the subject (Cummins, 1996; Harris and Nunez, 1996).

The human being appears to have an inherited propensity to learn social norms similar to our inherited propensity to learn grammatical rules (Pinker, 1994). Social norms are shared understandings about actions that are obligatory, permitted, or forbidden (Crawford and Ostrom, 1995; Cummins, 1996). Which norms are learned, however, varies from one culture to another, across families, and with exposure to diverse social norms expressed within various types of situations. Someone who has played competitive team sports or served in the military is likely to have learned about actions that others expect of them and that a failure to use the norm brings internally and externally triggered intrinsic costs. The anguish that an individual suffers from failing to use a social norm, such as telling the truth or keeping a promise, is referred to as guilt, if entirely self-inflicted, or as shame, when the knowledge of the failure is known by others (Posner and Rasmusen, forthcoming).

Recent developments in evolutionary theory and supporting empirical research provide strong support for the assumption that modern humans have inherited a propensity to learn social norms. A strict evolutionary approach, however, is difficult to apply when trying to understand how individuals with a propensity to learn social norms, interact, adapt, and learn within shorter time frames. An evolutionary game-theoretical model of a collective-action situation would, for example, treat all individuals as having inherited a particular type of strategy and unable to change their own strategy. Over time, those carrying the more successful strategies for that environment would reproduce at a higher rate than those who were less successful. After many iterations the more successful strategies would come to prominence in the population. These models provide substantial insight into the initial mix of strategies that enable some strategies to come to dominance and when a single strategy is likely to survive versus multiple strategies (see, for example, Nowak and Sigmund, 1998; Sethi and Somanathan, 1996; Epstein, 1998; Epstein and Axtell, 1996). Strategies consistent with reciprocity are repeatedly part of an evolutionary stable solution in many social dilemma games. Strict evolutionary models, however, focus as much on the results of *past* interactions as strict rational choice models focus on expected results of *future* interactions. In ongoing collective-action settings, however, most individuals learn from the past as well as plan for the future.

An Indirect Evolutionary Approach

Recent work on an indirect evolutionary approach to the study of human behavior offers a rigorous theoretical approach for understanding how preferences—including those associated with social norms—evolve or adapt (Girth and Yaari, 1992; Giith, 1995). In an indirect evolutionary model, players receive objective payoffs, but make decisions based on the transformation of these material rewards into intrinsic preferences. Those who value reciprocity, fairness, and/or trust add a subjective change parameter to actions (of themselves or others) that are consistent or not consistent with their norms.

Social dilemmas associated with games of trust (sequential PD games) are particularly useful games for discussing the indirect evolutionary approach given recent empirical studies. In a 2-person game of trust, Player 1 must decide whether to trust Player 2 to perform an act that will return Player 2 a lower pecuniary payoff than not performing the act. The starred payoffs in Figure 2 (based on Giith and Kliemt, 1998) represent standardized pecuniary outcomes of this game. Player 2 can either trust, *T,* or not trust, *NT,* Player 2. If Player 1 does not trust Player 2, both receive a material reward of $s*$ which is less than

r^*. If Player 1 trusts Player 2, then Player 2 can perform the trusted act, *P,* in which case they both receive r^*. Player 2, however, faces the temptation not to perform, *NP,* and to receive the higher material reward of 1 while Player 1 receives only zero. Most contractual relationships have this basic structure. Contracts may relate to strictly private goods, involve the provision of a public good, or concern the maintenance of common-pool resources.

[Figure 2 about here]

For a rational egoist playing this game, material outcomes are linearly related to utility. Dropping the stars in Figure 2 will capture a rational egoist's preferences. For rational egoists, the equilibrium strategy is for Player 1 to choose *NT* because Player 1 can predict that Player 2 will choose *NP.* Thus, both players end up with s^* in objective terms when they could have realized $r^* > s^*$. To represent the utility function of individuals who have adopted a social norm, such as to reciprocate trust with trustworthiness, it is necessary to transform the material payoffs to reflect the internal benefits and costs of taking actions consistent or inconsistent with the norm. The intrinsic payoff represented by the change parameter may be positive (a reward for a good deed) or negative (a punishment for breaking a social norm) and be triggered by external monitoring (shame or glory) or strictly imposed by the individual on themselves (guilt or pride) (Crawford and Ostrom, 1995). The size of the intrinsic payoff may be larger than, equal to, or less than the material payoff. Thus, simply assuming that humans adopt social norms does not eliminate the problem of collective action.

In Figure 3, which represents the trust game with payoffs in utility space, a change parameter, **m,** has been added to the material payoffs obtained through nonperformance to reflect intrinsic payoffs felt when breaching a trust. The parameter, **m,** can now be used to represent a trustworthy, m-type, for whom $m + 1 < r$ $(m < r-1 < 0)$ and a rational egoist type, m for whom $m + 1 > r$ $(m > r-1)$. For the trustworthy type, the choice of *P* is rational, while the choice of *NP* is rational for the rational egoist. The value of **m** can vary substantially from one person to another, but all **m**-types will fall somewhere within the two inequalities leading to the two types of players.

[Figure 3 about here]

While the value of **m** affects the choice that an individual makes, it does not affect the objective payoff obtained after a move is made. Thus, social norms may lead individuals to behave differently in the same objective situation depending on how strongly they value conformance with (or deviance from) a norm. By their behavior and resulting interaction, however, different types of players are more or less likely to gain higher objective returns. In an evolutionary process, one can posit a game where nature chooses which of the two **m**-types will hold the position of Player 1 or Player 2. If this game is one of complete information so that Player 1 knows the type of Player 2, then all Player 1's (no matter which type they are) will play *AT* when faced with a rational egoist and *T* when facing a trustworthy type. With complete information regarding types, trustworthy types will receive more opportunities to perform and receive *r,* while rational egoists will consistently receive a lower payoff.

In an evolutionary process, only the trustworthy type would survive in a complete information setting (Guth and Kliemt, 1998: 386). Viewed as a cultural evolutionary process, new entrants to the population would be more likely to adopt the preference ordering of those who obtained the higher material payoffs in the immediate past (Boyd and Richerson, 1985). Viewed as a learning process, those who were less successful would tend to learn the values of those who had achieved higher material rewards (Borgers and Sarin, 1997).[3] Where a player's type is common knowledge, rational egoists would not survive, but full and accurate information about all player's types is a very strong assumption.

On the other hand, in a large population without any signals about player type, preferences will evolve to be a linear transformation of material payoffs.[4] If only information about the proportion, q, of a population that are trustworthy is known, and no information is known about the type of a specific player, then Guth and Kliemt (1998) show that first players will trust Player 2's so long as $qr > s$. In such a setting the share of the population held by the trustworthy types is bound to decline. On the other hand, if there is a noisy signal (Frank, 1987) about a player's type that is at least more accurate than random, the trustworthy type will survive as a substantial proportion of the population. Noisy signals may result from seeing one another, face-to-face communication, and various mechanisms that humans have designed to monitor each other's behavior.

Guttman (1999) has applied the indirect evolutionary approach to a finitely repeated PD game in which players can decide to stay with the same partner with whom they are randomly matched or exit after a transaction and be matched with a new partner. He finds that under these conditions, and even starting with a small proportion of contingent cooperators, rational egoists are encouraged to reciprocate during most of their transactions, but to defect in the last period. Thus, the important result of Kreps et al. (1982), that cooperation can be an equilibrium in a finitely repeated PD game, has now been embedded in an endogenous process whereby contingent cooperators survive in a competitive, evolutionary environment and their presence induces rational egoists to mimic their behavior (and gain a reputation for reciprocity) until the last period of the game.

Evidence Consistent with Indirect Evolutionary Processes

An indirect evolutionary approach explains how a mixture of norm-users and rational egoists would emerge in settings where standard rational-choice theory predicts the presence of rational egoists alone. The evidence cited earlier is consistent with this explanation and is in large part the stimulus for the development of the indirect evolutionary approach. Recent studies provide evidence that supports this theory.

One prediction derived from the theory is that some actors in a game with the structure of a PD game would prefer to obtain the outcome achieved when both cooperate (C,C) over the outcome achieved when the actor defects and a partner cooperates (D,C), even though the latter provides higher financial rewards. Further, the proportion of individuals ranking (C,C) above (D,C) should change with the immediate experiences of the subjects. This prediction has empirical support. In a ranking that occurred after subjects made their own decision in a one-shot, sequential, double-blind, PD experiment, but before they knew their partner's decision, 40% of a pool of 136 subjects ranked (C,C) higher than (D,C), and 27% were indifferent between these outcomes (Ahn, Ostrom, and Walker, 1998). Thus, two-thirds of this pool of respondents appear to have a change parameter in their preference function leading them to rank (C,C) higher than (or equal to) a larger financial reward. A group of 181 undergraduates completed a similar questionnaire on the first day of classes at Indiana University in January of 1999. In this nondecision setting, 27% ranked the outcome (C,C) over (D,C) and 25% were indifferent. Thus, over half reflected the presence of a change parameter when presented with a hypothetical payoff matrix. A third ranking was obtained after 72 subjects had played 12 rounds of a finitely repeated PD game where partners were randomly matched each round and rates of cooperation were low. After experiencing multiple instances where partners defected, only 19% of the respondents ranked (C,C) above (D,C) while 17% were indifferent (Ahn et al., 1999).

If actors use norms of reciprocity, framing dilemma games in a manner that stresses reciprocity and providing nonfinancial encouragement of cooperative behavior should generate more cooperation. In a series of three-person, depletable common-pool resource dilemmas, Martichuski and Bell (1991) framed one design with the following advice. "Here is a way to make a lot of points: When you make your choices, make them exactly the way that you would want other people to make their choices." In one design condition, players who withdrew a sustainable number of points from the pool were sent a written message that everyone saw on their screen of "Good choice, Player X" where X was the subject's first name. In a punishment design, the feedback of "Bad Choice, Player X" was directed to those who withdrew an excessive number of points. These feedback messages are an external, nonmaterial payoff that also alerted all participants as to the level of withdrawals made by each player. This simple framing and nonfinancial rewards and punishments led participants to gain significantly more points and to keep the pool active longer than the same design without the framing and written messages. Subjects responding to these intrinsic rewards were thus actually able to gain higher objective payoffs than those not reacting to these rewards.

Players should use any information given to them about prior strategies adopted by others when making choices in dilemma situations. Further, those who have used strategies consistent with being a conditional cooperator should be more optimistic about other's behavior than rational egoists. Cain (1998) provided information about a player's choice in a dictator game to players subsequently engaged in a PD game. Stingy players—those who retained at least 70% of their endowment in the earlier dictator game—tended to predict that all players would defect. Nice players—those that gave away at least 30% of their endowment—tended to predict that the nice players would cooperate and stingy players would defect (Cain, 1998: 151). This information also affected the behavior of players—particularly those who were "nice" in the dictator game. Nice players chose cooperation 69% of the time when they were paired with other nice players and 39% of the time when they were paired with stingy players.

Bohnet, Frey, and Huck (1999) directly test an indirect evolution model when they examine whether levels of third-party enforcement "crowd out" or "crowd in" the norm of being trustworthy (see Frey, 1994, 1997; Frey and Oberholzer-Gee, 1997). They embed a basic game of trust in a regulatory regime where a litigation process is initiated if there is a breach of performance. As shown in Figure 4, Player 2 is held liable *(L)* or not liable *(NL)* with probability z. If held liable, both players split the pecuniary surplus as if Player 2 had performed but Player 2 has to pay for the cost of a trial $c^* > 0$. (This division is similar to the rules used in English, most European, and some U.S. courts for breach of contract settlements.) If not held liable, Player 2 profits from the breach and pays none of the litigation costs, while Player 1 receives no pecuniary payoffs from the contract and suffers a financial loss of $-a^*$ (a equal or greater than c*) because of the need to pay for the litigation.

[Figure 4 about here]

Two types of players are posited: a trustworthy m-type for whom $m + 1 < r$ and an untrustworthy m -type for whom $m + 1 > r$ (these subjective payoffs are not shown in Figure 4). Drawing on a stochastic model of individual preference adaptation, Bonnet, Frey, and Huck (1999) predict that norm-using behavior will be crowded out when m -types earn more than m-types. They design an experiment with three enforcement regimes. In the high-enforcement regime, litigation is successful in punishing those players who breach contracts—a high z case where $z > 1/(1 + c)$. In the high-enforcement case, both types of players choose *T* when in the position of Player 1 and P when in the position of Player 2. Regardless of the distribution of types of players, effective external enforcement guarantees high performance. The high-enforcement regime is similar to a perfectly working third-party enforcement regime. In the medium-enforcement regime, where $a/(r + a) < z < 1/(1 + c)$, m -types will, on average, earn more monetary payoffs than m-types. Over time, norm-users are crowded out.

In a full information, low-enforcement regime where $z < a/(r + a)$, the equilibrium strategies will be *(NT, NP)* when Player 2 is trustworthy and *(T, P)* if Player 2 is not trustworthy. In this case, norm-using types in the role of Player 2 always earn *r* and those who do not use the norm always earn *s* when in the position of Player 2. Returns to those in the role of Player 1 do not differ among types. Since $r > s$, norm-using behavior is more successful in a full-information, low-enforcement regime. Over time, trustworthy players would be the only kind to remain in the game. With partial information, players would discriminate between those expected to be trustworthy and those who expected not to be trustworthy.

Subjects played 9 rounds of a repeated trust game with monetary payoffs fitting the game structure of Figure 4. During the first 3 rounds, the publicly announced enforcement regime was either high, $z = .9$; medium, $z = .5$; or low, $z = .1$. After round 3, players in the medium and high z regimes were told that the regime would shift to a low z. Bohnet, Frey, and Huck predicted that the longer subjects played in a low-z regime, the more norm-using preferences (and resulting behavior) should be "crowded" in. Besides expecting a higher performance rate in the low-enforcement regime, they predicted that there would be no end-game reduction in cooperation in this regime. In the experiments where subjects were randomly matched with one another and received information about the aggregate number of moves made by Player 1's and Player 2's for past rounds, cooperation rates rose steadily in the last six rounds. Performance rates reached *their maximum* in the last round![5] Further, the longer that subjects had been in a low-enforcement regime, the higher the likelihood that they chose *P* when in the position of Player 2.

Bohnet, Frey, and Huck draw two main implications from their theory of general relevance to the theory of collective action. The first is that "it is impossible to predict behavior in a group of agents playing the contract game without knowing their history" (1999: 9). Since preferences change in light of experience, the behavior of individuals changes over time. "The longer a group of agents has played in a low-probability environment, the more individuals with a preference for honesty are present and the less breach is observed" (Ibid.). The second major implication is that rule enforcement does not exert a monotonic impact on behavior. In their model, the "worst regime is not one where contracts cannot be enforced but one with an intermediate level of contract enforceability" (Ibid.). This is true because with a low z, Player 1's are highly motivated to find out the tmstworthiness of Player 2's before engaging in an interaction, while in medium and high regimes they pay little or no attention to Player 2's type. While in a high effective enforcement regime, all players are forced to perform on their contracts, in a medium-enforcement regime, it is attractive for an m-type to breach contracts and risk being caught and their better economic performance over time crowds out norm-using types of players.

Cardenas, Stranlund, and Willis (1999) have also found that external monitoring and sanctioning imposed on an 8-person CPR dilemma "crowded out" more cooperative behavior. They conducted a series of 15 experiments in three rural villages of Colombia where local villagers daily confront CPR problems similar to those posed by the experiment. They imposed a regime that specified the number of endowed units each player should allocate to the CPR—equal to the optimal level of units per subject. Each subject was monitored each round with a probability of .06 and a small fine imposed if their action exceeded the imposed quota. Under this exogenous enforcement regime, individual behavior rapidly evolved toward the Nash best response, while in an alternative design where subjects could talk with one another each round, individual behavior slowly evolved away from the Nash best response. At the end of the experiment, subjects in the cheap talk condition overharvested at a lower rate and earned higher returns than those in the weakly enforced, external regulation regime. As the authors noted:

> The difference between the effects of communication and regulation on the balance between self and other-regarding behavior could not be more stark. While external regulation quickly crowded out group-oriented behavior in favor of greater self interest, the simple ability to communicate induced a shift from choices that were relatively group-oriented in the absence of communication to an even stronger group-orientation with communication. (Cardenas, Stranlund, and Willis, 1999: 19)

These findings are consistent with the earlier paradoxical findings of Frohlich and Oppenheimer (1996) from a 5-person PD game. One set of groups played a regular PD game (some with communication and some without). A second set of groups used an externally imposed, incentive-compatible mechanism designed to enhance efficiency. In the first phase of the experiment, the second set of groups gained higher monetary returns than the control groups as expected. In the second phase of the experiment, both groups played a regular PD game. To the surprise of the experimenters, a higher level of cooperation occurred in the control groups that played the regular 5-person PD in both phases (especially for those who communicated on a face-to-face basis). The greater cooperation that had occurred due to the incentive compatible mechanism appeared to be transient and disappeared after Phase 1. Further, the incentive compatible mechanism:

> seemed to undermine subsequent cooperation and leave the group worse off than those in the control group who had played a regular 5-PD. . . . By removing the opportunity to wrestle with the dilemma, the device may be cueing individuals simply to follow their individual interests, and may cause self-

interested behavior to be reinforced and carried over into subsequent decisions. (Frohlich and Oppenheimer, 1999: 180)

The Future of the Theory of Collective Action

The years since the publication of *The Theory of Collective Action* have seen extensive empirical research that challenges the strong version of the theory—the zero contribution thesis. Empirical research also demonstrates that free riding is a strong temptation that can lead to a failure to achieve collective action or a breakdown of existing efforts. Assuming that all individuals are rational egoists does not explain the level of cooperation achieved or how contextual variables affect patterns of cooperation or its failure. In addition to those individuals whose behavior is well captured by the standard theory of rational choice, it is necessary to assume that individuals who try to use social norms, such as reciprocity, trust, and fairness, are initially present in most situations. In competitive market situations, rational egoists are the major survivors over time. In nonmarket settings, it is possible for multiple types of players to survive.

Using an indirect evolutionary theory, it is possible to show that individuals who adopt social norms consistent with conditional cooperation are able to survive, and even flourish, in repeated dilemma situations so long as at least a small proportion of the other players start off as norm-using types and some information about individual player types is generally known. Second, the long-term growth of norm-using types in a population is more dependent on the development of mechanisms for revealing players' types than on the initial distribution of types in the beginning of a process. Developing low-cost methods to reveal the reliability of participants is thus a crucial step in enabling trustworthy participants to perform and return a higher outcome to themselves and the discriminating trustors. This is consistent with the historical findings related to trade (for example, see Milgrom, North, and Weingast, 1990) and to long-enduring governance of common-pool resources (Ostrom, 1990). Norm-using types will be crowded out in those settings where interacting with others does not produce higher payoffs in the long run. When this happens, only rational egoists survive. Rational egoists cannot by themselves overcome the temptation to free ride and thus collective action fails when social norms are crowded out.

Further theoretical work needs to ask how the large array of contextual variables identified in field research may affect the processes of teaching and evoking social norms, of informing participants of the behavior of others and their adherence to social norms, and of rewarding those who use social norms, such as reciprocity, trust, and fairness. We need to understand how institutional, cultural, and biophysical contexts affect the types of individuals who are recruited into and leave particular types of collective-action situations, the kind of information that is made available about past actions, and how individuals can themselves change structural variables so as to enhance the probabilities of norm-using types being involved and growing in strength over time. Further theoretical developments along these lines are essential for the development of public policies that enhance socially beneficial, cooperative behavior based in part on social norms rather than crowding them out. Past policy initiatives based primarily on changing payoff structures for rational egoists may have been misdirected and crowded out intrinsic preferences with counterintentional consequences. Increasing the capabilities of individuals to devise their own rules may well result in processes that allow social norms to evolve and thereby increase the probability of individuals solving problems of collective action.

Notes

1. See McCay and Acheson, 1987; Berkes et al., 1989; Bromley et al., 1992; Blomquist, 1992; Pinkerton, 1989; Feeny et al., 1990; Tang, 1992; Wade, 1994; Lichbach, 1996; and Lam, 1998 for a small sample of the vast corpus of research on self-organized resource regimes. An extensive bibliography on diverse institutions for dealing with common-pool resources is organized by Charlotte Hess and can be searched on the Web at: <http://www.indiana.edu/~workshop/wsl/wsl.htrnl> or obtained on a CD-ROM disk (Hess, 1999).

2. McCabe, Rassenti, and Smith, 1996; Elster, 1989; Frank, 1987; Frey, 1994, 1997; Rabin, 1993; Fehr and Schmidt, forthcoming; Selten, 1991; Sethi and Somanathan, 1996; Bowles, 1998; Ostrom, 1998.

3. Eshel, Samuelson, and Shaked (1998) develop a learning model where a population of Altruists who adopt a strategy of providing a local public good interact in a local neighborhood with a population of Egoists who free ride. In this local interaction setting, Altruist's strategies are imitated sufficiently often in a Markovian learning process to become one of the absorbing states. Altruists interacting with Egoists outside a circular local neighborhood are not so likely to survive.

4. This implies that in a game where players know only their own payoffs and not the payoffs of others, that they are more likely to behave like rational egoists. McCabe and Smith (1999) show that players tend to evolve toward the predicted, sub-game perfect outcomes in experiments where they have only private information of their own payoffs and to cooperative outcomes when they have information about payoffs and the moves made by other players (see also McCabe, Rassenti, and Smith, 1996).

5. In a control experiment, where subjects were permanently matched, and reputation effects such as modeled by Kreps et al. (1982) would dominate, cooperation levels were quite high over the course of the nine repetitions. As expected, however, they dropped from 100% to zero in the last round.

References

Agrawal, Arun, "Group Size and Successful Collective Action: Forest Management Institutions in Nine Himalayan Villages." In Clark Gibson, Margaret McKean, and Elinor Ostrom, eds., *People and Forests: Communities, Institutions, and Governance.* Cambridge, MA: MIT Press, forthcoming.

Agrawal, Arun, and Clark Gibson, "Community in Natural Resource Conservation," *World Development,* 1999, 27(4), 629-49.

Ahn, Toh-Kyeong, Elinor Ostrom, David Schmidt, and James Walker, "Dilemma Games: Game Parameters and Matching Protocols." Bloomington: Indiana University, Workshop in Political Theory and Policy Analysis, Working paper, 1999.

Ahn, Toh-Kyeong, Elinor Ostrom, and James Walker, "Trust and Reciprocity: Experimental Evidence from PD Games." Bloomington: Indiana University, Workshop in Political Theory and Policy Analysis, Working paper, 1998.

Alchian, Armen A. "Uncertainty, Evolution, and Economic Theory," *Journal of Political Economy,* 1950, *58(3),* 211-21.

Alexander, Richard D., *The Biology of Moral Systems.* New York: Aldine de Gruyter, 1987.

Andreoni, James, "Why Free Ride? Strategies and Learning in Public Goods Experiments," *Journal of Public Economics,* 1988, *37,* 291-304.

Asquith, Nigel M. *How Should the World Bank Encourage Private Sector Investment in Biodiversity Conservation?* A Report Prepared for Kathy MacKinnon, Biodiversity Specialist, The World Bank, Washington, D.C. Durham, North Carolina: Sanford Institute of Public Policy, Duke University, 1999.

Axelrod, Robert. *The Evolution of Cooperation.* New York: Basic Books, 1984.

Axelrod, Robert. "An Evolutionary Approach to Norms," *American Political Science Review,* December 1986, *80(4),* 1095-111.

Baland, Jean-Marie, and Jean-Philippe Platteau, *Halting Degradation of Natural Resources: Is There a Role for Rural Communities.* Oxford: Clarendon Press, 1996.

Bardhan, Pranab, "Water Community: An Empirical Analysis of Cooperation on Irrigation in South India." Berkeley: University of California, Department of Economics, Working paper, 1999.

Barkow, Jerome H., Leda Cosmides, and John Tooby, eds., *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* Oxford: Oxford University Press, 1992.

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, Reciprocity, and Social History," *Games and Economic Behavior,* July 1995, 70(1), 122-42.

Berkes, Fikret, David Feeny, Bonnie J. McCay, and James M. Acheson, "The Benefits of the Commons," *Nature,* July 1989, *340,* 91-93.

Blomquist, William, *Dividing the Waters: Governing Groundwater in Southern California.* San Francisco, CA: ICS Press, 1992.

Bohnet, Iris, and Bruno S. Frey, "The Sound of Silence in Prisoner's Dilemma and Dictator Games," *Journal of Economic Behavior and Organization,* January 1999, *38(1),* 43-58.

Bohnet, Iris, Bruno S. Frey, and Steffen Huck, "More Order with Less Law: On Contract Enforcement, Trust, and Crowding." Cambridge, MA: Harvard University, Working paper, 1999.

Borgers, Tilman, and Rajiv Sarin, "Learning Through Reinforcement and Replicator Dynamics," *Journal of Economic Theory,* 1997, **77,** 1-14.

Bowles, Samuel, "Endogenous Preferences: The Cultural Consequences of Markets and other Economic Institutions,'Vo«r«a/ *of Economic Literature,* March 1998, *36,* 75-111.

Boyd, Robert, and Peter J. Richerson, *Culture and the Evolutionary Process.* Chicago: University of Chicago Press, 1985.

Boyd, Robert, and Peter J. Richerson, "The Evolution of Reciprocity in Sizable Groups," *Journal of Theoretical Biology,* 1988,752,337-56.

Brandts, Jordi, and Arthur Schram, "Tacit Cooperation or Noise in Voluntary Contribution Mechanism Experiments: Applying the Contribution Function Approach." Amsterdam: CREED, Working paper, 1995.

Bromley, Darnel W., *etal,* eds., *Making the Commons Work: Theory, Practice, and Policy.* San Francisco, CA: ICS Press, 1992.

Cain, Michael, "An Experimental Investigation of Motives and Information in the Prisoner's Dilemma Game," *Advances in Group Processes,* 1998,75, 133-60.

Camerer, Colin, and Richard Thaler, "Anomalies: Ultimatums, Dictators, and Manners," *Journal of Economic Perspectives,* 1995, 9(2), 209-19.

Cameron, Lisa, "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia." Princeton, NJ: Princeton University, Discussion paper, 1995.

Cardenas, Juan-Camilo, John Stranlund, and Cleve Willis, "Local Environmental Control and Institutional Crowding-Out." Amherst: University of Massachusetts, Department of Resource Economics, Working paper, 1999.

Clark, Andy, and Annette Karrniloff-Smith, "The Cognizer's Innards: A Psychological and Philosophical Perspective on the Development of Thought," *Mind and Language,* Winter 1991, 5(4), 487-519.

Crawford, Sue E. S., and Elinor Ostrom, "A Grammar of Institutions," *American Political Science Review* September 1995, *89$),* 582-600.

Croson, Rachel, "Partners and Strangers Revisited," *Economics Letters,* 1996, *53,* 25-32.

Croson, Rachel, "Theories of Altruism and Reciprocity: Evidence from Linear Public Goods Games." Philadelphia: University of Pennsylvania, The Wharton School, 1998.

Cummins, Denise D., "Evidence of Deontic Reasoning in 3- and 4-Year-Olds," *Memory and Cognition,* 1996, *24,* 823-29.

Davis, Douglas D., and Charles A. Holt, *Experimental Economics.* Princeton, NJ: Princeton University Press, 1993.

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

Dawes, Robyn M., John M. Orbell, and Alphons van de Kragt, "Organizing Groups for Collective Action." *American Political Science Review,* December 1986, *80(4),* 1171-85.

de Janvry, Alain, Nancy McCarthy, and Elisabeth Sadoulet, "Endogenous Provision and Appropriation in the Commons." Berkeley: University of California, Department of Economics, 1998. Paper presented at the Allied Social Science Association Meeting, January 2-5, Chicago.

Elster, Jon, "Social Norms and Economic Theory," *Journal of Economic Perspectives,* 1989, 5(4), 99-117.

Epstein, Joshua M., "Zones of Cooperation in Demographic Prisoner's Dilemma," *Complexity* 1998, *4(2),* 36-48.

Epstein, Joshua M., and Robert Axtell, *Growing Artificial Societies: Social Science from the Bottom Up.* Cambridge, MA: MIT Press, 1996.

Eshel, Ilan, Larry Samuelson, and Avner Shaked, "Altruists, Egoists, and Hooligans in a Local Interaction Model," *American Economic Review,* March 1998,55(1), 157-79.

Farrell, Joseph, "Cheap Talk, Coordination, and Entry." *Rand Journal of Economics,* Spring 1987,*18,* 34-39.

Farrell, Joseph, and Matthew Rabin, "Cheap Talk," *Journal of Economic Perspectives,* Summer 1996, 70(3), 103-18.

Feeny, David, Fikret Berkes, Bonnie J. McCay, and James M. Acheson, "The Tragedy of the Commons: Twenty-Two Years Later," *Human Ecology,* 1990,75(1), 1-19.

Fehr, Ernst, and Simon Gachter, "Cooperation and Punishment." Zurich, Switzerland: University of Zurich, Working paper, 1998.

Fehr, Ernst, George Kirchsteiger, and Arno Riedl, "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics,* 1993,*108(2),* 437-60.

Fehr, Ernst, and Klaus Schmidt, "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics,* forthcoming.

Fischbacher, Urs, Simon Gachter, and Ernst Fehr, "Anomalous Behavior in Public Goods Experiments: How Much and Why? A Comment." Zurich: University of Zurich, Institute for Empirical Research in Economics, Working paper, 1999.

Forsythe, Robert, Joel L. Horowitz, and Nathan E. Savin, "Fairness in Simple Bargaining Games," *Games and Economic Behavior,* 1994, 6(3), 347-69.

Frank, Robert H, "If *Homo Economicus* Could Choose His Own Utility Function, Would He Want One with a Conscience?," *American Economic Review,* 1987, 77(4), 593-604.

Frank, Robert H, Thomas Gilovich, and Dennis T. Regan, "The Evolution of One-Shot Cooperation: An Experiment," *Ethology and Sociobiology,* July 1993,*14,* 247-56.

Frey, Bruno S., "How Intrinsic Motivation is Crowded Out and In," *Rationality and Society,* 1994, *6,* 334-52.

Frey, Bruno S., *Not Just for the Money: An Economic Theory of Personal Motivation.* Cheltenham, England: Edward Elgar, 1997.

Frey, Bruno S., and Felix Oberholzer-Gee, "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding Out," *American Economic Review,* 1997, 57(4), 746-55.

Frohlich, Norman, and Joe A. Oppenheimer, "Experiencing Impartiality to Invoke Fairness in the N-PD: Some Experimental Results," *Public Choice* 1996,*86,* 117-35.

Frohlich, Norman, and Joe A. Oppenheimer, "Values, Policies, and Citizen Competence: An Experimental Perspective." In Stephen L. Elkin and Karol E. Soltan, eds., *Citizen Competence and Democratic Institutions.* University Park: Pennsylvania State University Press, 1999, pp. 161-87.

Frohlich, Norman, Joe A. Oppenheimer, and J. Bernard Moore, "Some Doubts About Measuring Self-interest Using Dictator Experiments: The Cost of Anonymity." College Park, MD: University of Maryland, Department of Government and Politics, Working paper, 1998.

Giith, Werner, "An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives," *International Journal of Game Theory,* 1995, *24,* 323-44.

Giith, Werner, and Hartmut Kliemt, "The Indirect Evolutionary Approach: Bridging the Gap Between Rationality and Adaptation," *Rationality and Society,* August 1998, 70(3), 377-99.

Giith, Werner, Rolf Schmittberger, and Bernd Schwarze, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization,* December 1982, 3(4), 367-88.

Girth, Werner, and Menahem Yaari, "An Evolutionary Approach to Explaining Reciprocal Behavior in a Simple Strategic Game." In Ulrich Witt, ed., *Explaining Process and Change. Approaches to Evolutionary Economics.* Ann Arbor: University of Michigan Press, 1992, pp. 23-34.

Guttman, Joel M., "Self-Enforcing Agreements and the Evolution of Preferences for Reciprocity." Ramat-Gan, Israel: Bar-Ilan University, Working paper, 1999.

Hardin, Russell, "Collective Action as an Agreeable n-Prisoners' Dilemma," *Science,* September-October 1971, 76,472-81.

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

Harris, Paul L., and Maria Nunez, "Understanding of Permission Rules by Pre-School Children," *Child Development,* 1996, 67(4), 1572-91.

Hayashi, Nahoko, Elinor Ostrom, James Walker, and Toshio Yamagishi, "Reciprocity, Trust, and the Sense of Control: A Cross-Societal Study," *Rationality and Society,* February 1999, 77(1), 27-46.

Hess, Charlotte, *A Comprehensive Bibliography of Common Pool Resources.* CD-ROM. Bloomington: Indiana University, Workshop in Political Theory and Policy Analysis, 1999.

Hirshleifer, Jack, "Economics from a Biological Viewpoint," *Journal of Law and Economics,* 1977, *20,* 1-52.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith, "Social Distance and Other-Regarding Behavior in Dictator Games," *American Economic Review,* 1996, 86(3), 653-60.

Isaac, R. Mark, Kenneth McCue, and Charles R. Plott, "Public Goods Provision in an Experimental Environment," *Journal of Public Economics,* 1985, *26,* 51-74.

Isaac, R. Mark, James Walker, and Susan Thomas, "Divergent Evidence on Free Riding: An Experimental Examination of Some Possible Explanations," *Public Choice,* 1984, *43(2),* 113-49.

Isaac, R. Mark, James Walker, and Arlington W. Williams, "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups," *Journal of Public Economics,* May 1994, *54(1),* 1-36.

Kagel, John, and Alvin Roth, eds., *The Handbook of Experimental Economics.* Princeton, NJ: Princeton University Press, 1995.

Keohane, Robert O., and Elinor Ostrom, eds., *Local Commons and Global Interdependence: Heterogeneity and Cooperation in Two Domains.* London: Sage, 1995.

Keser, Claudio, and Frans van Winden, "Partners Contribute More to Public Goods than Strangers: Conditional Cooperation." Karlsruhe, Germany: University of Karlsruhe, Working paper, 1996.

Kikuchi, Masako, Yoriko Watanabe, and Toshio Yamagishi, "Accuracy in the Prediction of Others' Trustworthiness and General Trust: An Experimental Study," *Japanese Journal of Experimental Social Psychology,* 1996, 37(1), 23-36.

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

Lam, Wai Fung, *Governing Irrigation Systems in Nepal: Institutions, Infrastructure, and Collective Action.* Oakland, CA: ICS Press, 1998.

Leach, Melissa, Robin Mearns, and Ian Scoones, "Environment Entitlements: Dynamics and Institutions in Community Based Natural Resource Management," *World Development,* 1999, 27(2), 225-47.

Ledyard, John, "Public Goods: A Survey of Experimental Research." In John Kagel and Alvin Roth, eds., *The Handbook of Experimental Economics.* Princeton, NJ: Princeton University Press, 1995, pp. 111-94.

Lichbach, Mark Irving, *The Cooperator's Dilemma.* Ann Arbor: University of Michigan Press, 1996.

Manktelow, Ken I., and David E. Over, "Social Roles and Utilities in Reasoning with Deontic Conditionals," *Cognition,* 1991, *39,* 85-105.

Martichuski, Diane K. and Paul A. Bell, "Reward, Punishment, Privatization, and Moral Suasion in a Commons Dilemma," *Journal of Applied Social Psychology,* 1991, 27(16), 1356-69.

Marwell, Gerald, and Ruth E. Ames, "Experiments on the Provision of Public Goods I: Resources, Interest, Group Size, and the Free Rider Problem," *American Journal of Sociology,* 1979, *84,* 1335-60.

Marwell, Gerald, and Ruth E. Ames, "Experiments on the Provision of Public Goods II: Provision Points, Stakes, Experience and the Free Rider Problem," *American Journal of Sociology,* 1980, *85,* 926-37.

Marwell, Gerald, and Ruth E. Ames, "Economists Free Ride: Does Anyone *Ehel," Journal of Public Economics* 1981,*15,* 295-310.

Marwell, Gerald, and Pamela Oliver, *The Critical Mass in Collective Action: A Micro-Social Theory.* New York: Cambridge University Press, 1993.

McCabe, Kevin A., Stephen J. Rassenti, and Vernon L. Smith, "Game Theory and Reciprocity in Some Extensive Form Experimental Games." *Proceedings of the National Academy of Sciences,* November 1996, *93,* 13421-428.

McCabe, Kevin A., and Vernon L. Smith, "Strategic Analysis by Players in Games: What Information Do They Use." Tucson: University of Arizona, Economic Research Laboratory, Working paper, 1999.

McCay, Bonnie J., and James M. Acheson, *The Question of the Commons: The Culture and Ecology of Communal Resources.* Tucson: University of Arizona Press, 1987.

Messick, David M., "Alternative Logics for Decision Making in Social Settings," *Journal of Economic Behavior and Organization,* May 1999, 39(1), 11-28.

Milgrom, Paul R., Douglass C. North, and Barry R. Weingast, "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs," *Economics and Politics,* March 1990, 2(1), 1-23.

Molinas, Jose R., "The Impact of Inequality, Gender, External Assistance and Social Capital on Local-Level Cooperation," *World Development,* 1998, 26(3), 413-31.

Morrow, Christopher E., and Rebecca Watts Hull, "Donor-Initiated Common Pool Resource Institutions: The Case of the Yanesha Forestry Cooperative," *World Development,* 1996, 24(10), 1641-57.

Nowak, Martin A., and Karl Sigmund, "Evolution of Indirect Reciprocity by Image Scoring," *Nature,* 1998, 393(6685), 573-77.

Oaksford, Mike, and Nick Chater, "A Rational Analysis of the Selection Task as Optimal Data Selection," *Psychological Review,* 1994,*101(A),* 608-31.

Offerman, Theo, *Beliefs and Decision Rules in Public Goods Games: Theory and Experiments.* Dordrecht, the Netherlands: Kluwer Academic Publishers, 1997.

Okada, Akira, and Arno Riedl, "When Culture Does Not Matter: Experimental Evidence from Coalition Formation Ultimatum Games and Austria and Japan." Kyoto, Japan: Kyoto University, Institute of Economic Research, Discussion paper no. 497, 1999.

Oliver, Pamela, "Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations," *American Journal of Sociology,* 1980,55, 1356-75.

Olson, Mancur, *The Logic of Collective Action: Public Goods and the Theory of Groups.* Cambridge, MA: Harvard University Press, 1965.

Orbell, John M., and Robyn M. Dawes, "A 'Cognitive Miser' Theory of Cooperators' Advantage," *American Political Science Review,* June 1991, *85(2),* 515-28.

Orbell, John M., and Robyn M. Dawes, "Social Welfare, Cooperators' Advantage, and the Option of Not Playing the Game," *American Sociological Review,* December 1993, *58,* 787-800.

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

Ostrom, Elinor, "The Rudiments of a Theory of the Origins, Survival, and Performance of Common-Property Institutions." In Daniel W. Bromley, *et al.,* eds., *Making the Commons Work: Theory, Practice, and Policy.* San Francisco: ICS Press, 1992, pp. 293-318.

Ostrom, Elinor, "Self-Governance of Common-Pool Resources." In Peter Newman, ed., *The New Palgrave Dictionary of Economics and the Law,* vol. 3. London: Macmillan Press, 1998, pp. 424-33.

Ostrom, Elinor, "Reformulating the Commons." In Joanna Burger, Richard Norgaard, Elinor Ostrom, David Policansky, and Bernard Goldstein, eds., *The Commons Revisited: An Americas Perspective.* Washington, DC: Island Press, forthcoming.

Ostrom, Elinor, Roy Gardner, and James Walker, *Rules, Games, and Common-Pool Resources.* Ann Arbor: University of Michigan Press, 1994.

Ostrom, Elinor, James Walker, and Roy Gardner, "Covenants with and without a Sword: Self-Governance is Possible," *American Political Science Review,* June 1992, *86(2),* 404-17.

Palfrey, Thomas R., and Howard Rosenthal, "Private Incentives in Social Dilemmas," *Journal of Public Economics,* 1988, *35,* 309-32.

Pinker, Steven, *The Language Instinct.* New York: W. Morrow, 1994.

Pinkerton, Evelyn, ed., *Co-operative Management of Local Fisheries: New Directions for Improved Management and Community Development.* Vancouver: University of British Columbia Press, 1989.

Posner, Richard A., and Eric B. Rasmusen, "Creating and Enforcing Norms, with Special Reference to Sanctions," *International Review of Law and Economics,* forthcoming.

Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics," *American Economic Review,* 1993, *83,* 1281-302.

Rocco, Elena, "Trust Breaks Down in Electronic Contexts but Can Be Repaired by Some Initial Face-to-Face Contact." In *CHI 98 Conference Proceedings.* New York: Association for Computing Machinery, 1998, pp. 498-504.

Rocco, Elena, and Massimo Warglien, "Computer Mediated Communication and the Emergence of 'Electronic Opportunism.'" Trento, Italy: University of Trento, Working paper no. 1996-01, 1996.

Roth, Alvin E., and J. Keith Murnighan, "The Role of Information in Bargaining: An Experimental Study," *Econometrica,* 1982,50(5), 1123-42.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir, "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review,* December 1991, *81* (5), 1068-95.

Sally, David, "Conversation and Cooperation in Social Dilemmas. A Meta-Analysis of Experiments from 1958 to 1992," *Rationality and Society,* January 1995, 7(1), 58-92.

Sell, Jane, and Rick Wilson, "Levels of Information and Contributions to Public Goods," *Social Forces,* September 1991, 70(1), 107-24.

Selten, Reinhard, "Evolution, Learning, and Economic Behavior," *Games and Economic Behavior,* February 1991, 5(1), 3-24.

Sethi, Rajiv, and Eswaran Somanathan, "The Evolution of Social Norms in Common Property Resource Use," *American Economic Review,* September 1996, *86(4),* 766-88.

Smith, Vernon L., "Microeconomic Systems as an Experimental *Science," American Economic Review,* December 1982, 72, 923-55.

Smith, Vernon L., "The Two Faces of Adam Smith." Tucson: University of Arizona, Economic Research Laboratory, Working paper, 1997.

Tang, Shui Yan, *Institutions and Collective Action: Self Governance in Irrigation.* San Francisco, CA: ICS Press, 1992.

Trivers, Robert L., "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology,* March 1971, *46,* 35-57.

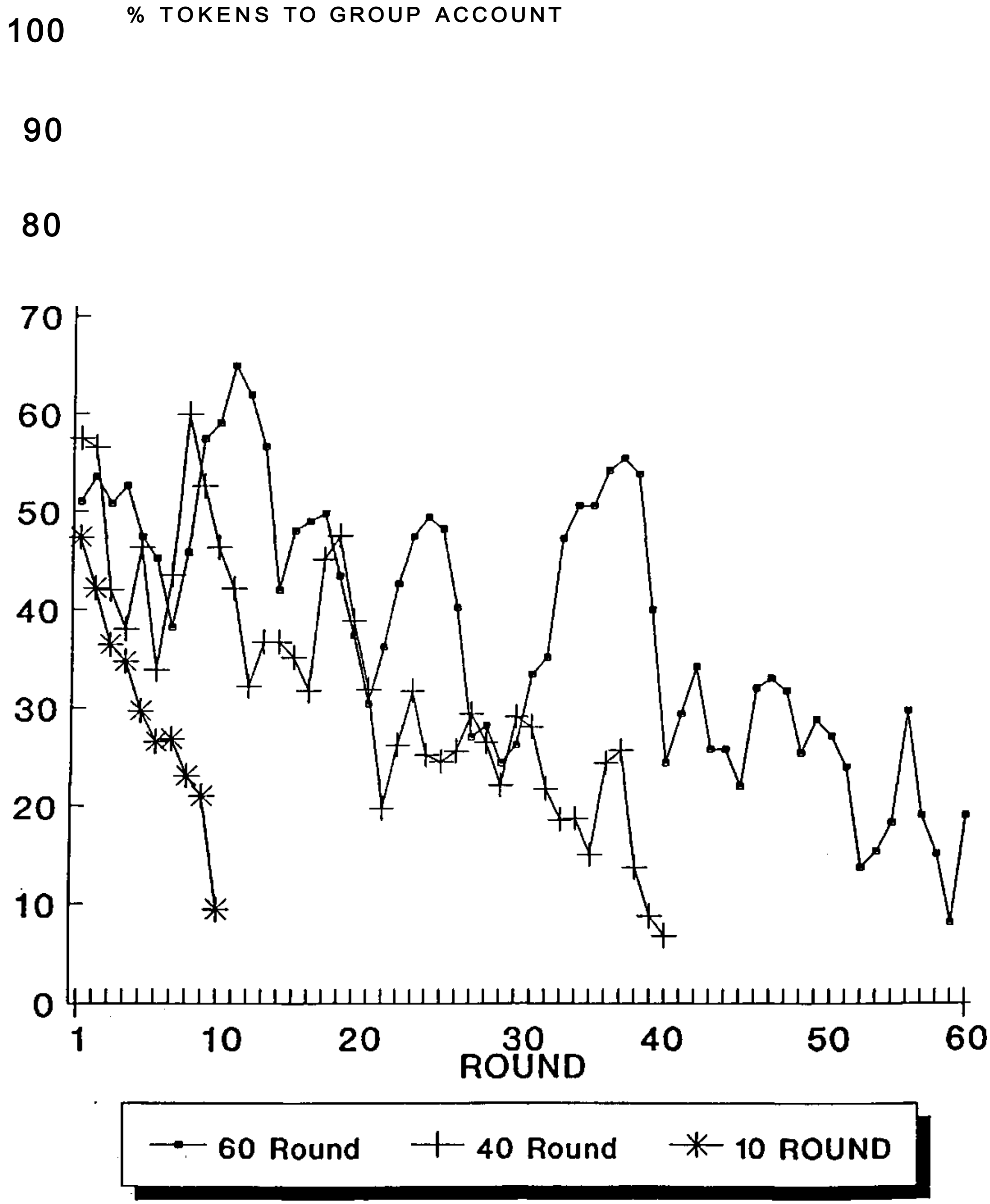
Wade, Robert, *Village Republics: Economic Conditions for Collective Action in South India.* San Francisco, CA: ICS Press, 1994.

Watabe, Motoki, S. Terai, Nahoko Hayashi, and Toshio Yamagishi, "Cooperation in the One-Shot Prisoner's Dilemma Based on Expectations of Reciprocity," *Japanese Journal of Experimental Social Psychology,* 1996, *36,* 183-96. [In Japanese with an English summary.]

Yamagishi, Toshio, "The Provision of a Sanctioning System as a Public Good," *Journal of Personality and Social Psychology,* 1986,57(1), 110-16.

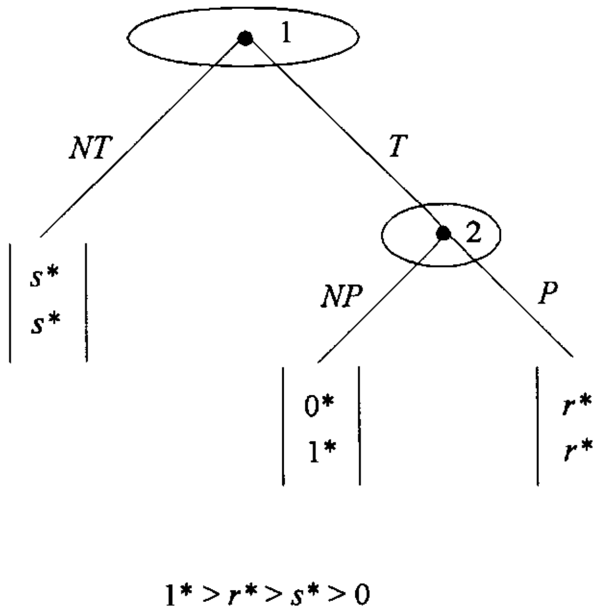
Yamagishi, Toshio, "The Provision of a Sanctioning System in the United States and Japan," *Social Psychology Quarterly,* 1988, 57(3), 265-71.

Figure 1. Contributions to Public Goods: 10, 40, and 60 Round Horizons



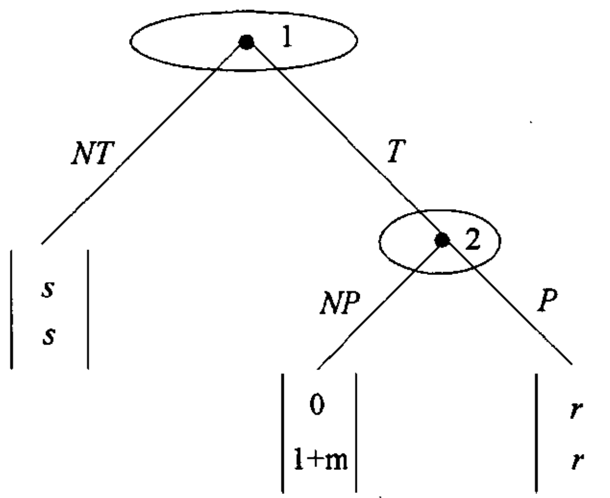
Source: **Isaac, Walker, and Williams (1994).**

Figure 2. The Trust Game: Material Outcomes



Source: Modified from Güth and Kliemt (1998: 381).

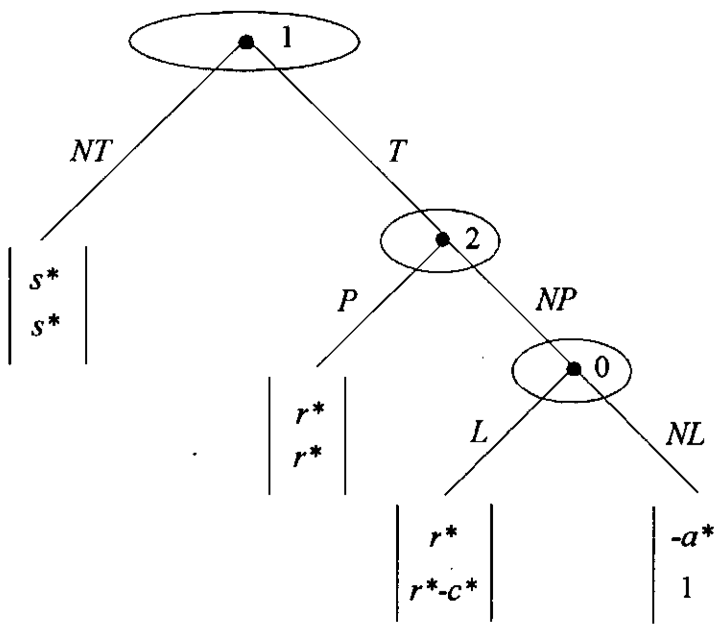
Figure 3. The Trust Game: Utilities



$1 > r > s > 0$

Source: Modified from Güth and Kliemt (1998: 386).

Figure 4. The Trust Game with Enforcement



$0 < a^* \leq c^* \leq s^* < r^* < 1$ and $ac^* < 1$

Source: Modified from Bohnet, Frey, and Huck (1999: 5).