W09-16 6/12/09

ANALYZING COLLECTIVE ACTION

by

Elinor Ostrom
Workshop in Political Theory and Policy Analysis
Indiana University

Center for the Study of Institutional Diversity
Arizona State University

Paper to be presented at the 2009 conference of the International Association of Agricultural Economists held in Beijing, China, August 16–22, 2009. It draws on extensive research undertaken with colleagues at the Workshop in Political Theory and Policy Analysis and on an earlier chapter for *The Oxford Handbook of Comparative Politics*, ed. Carles Boix and Susan Stokes (Oxford, UK: Oxford University Press). Presentations of earlier drafts were made at Duke University and at the Workshop in Political Theory and Policy Analysis, Indiana University. The excellent editing of Patty Lezotte and David Price has been of great help.

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 / http://www.indiana.edu/~workshop

ANALYZING COLLECTIVE ACTION

Elinor Ostrom

Considerable theoretical turmoil exists related to the underlying "problem" of collective action—how to model social dilemma situations in light of the repeated evidence that early theoretical predictions have not been supported. The most famous social dilemma is the Prisoner's Dilemma (PD). Traffic jams, residential flight, runs on scarce goods, extending and keeping trust in long-term relations, and the organizing of labor unions, work-teams, demonstrations, or any group seeking common interests—all can and have been modeled as social dilemmas generating collective action problems. Frequently, collective-action problems are modeled as public good games, common-pool resource games, games of trust, the dictator and ultimatum game, as well as a Prisoner's Dilemma game (Camerer 2003; Kagel and Roth 1995; Sandler 1992).

Collective action problems occur when individuals choose actions—such as whether to build and maintain an irrigation system—in an interdependent situation. If each individual in such situations selects strategies based on a calculus that maximizes short-term benefits to self, individuals will take actions that generate lower joint outcomes than could have been achieved. In other words, a collective action problem can be analyzed as a game where the Nash equilibrium for a single iteration of the game yields less than the socially optimal outcome. The socially optimal outcome could be achieved if those involved "cooperated" by selecting strategies other than those prescribed by the Nash equilibrium. Since the suboptimal joint outcome is an equilibrium, no one is independently motivated to change their choice, given the predicted choices of all others. Thus, the socially desirable outcome is predicted not to occur.

In addition to the assumption regarding the structure of payoffs leading to a deficient equilibrium, further assumptions made in almost all formal models of social dilemmas include:

1. All participants have common knowledge of the exogenously fixed structure of the situation and of the payoffs to be received by all individuals under all combinations of strategies.
2. Decisions about strategies are made independently and simultaneously.
3. No external actor (or central authority) is present to enforce agreements among participants about their choices.

When these assumptions are made for a game that is repeated only once, the theoretical prediction derived from noncooperative game theory is unambiguous—zero cooperation.

When uncertainty exists about the time or the number of rounds involved in a repeated game, such as would usually be case in field settings, two theoretical developments generate more optimistic predictions than backward induction in finitely repeated games. First, Kreps et al. (1982) posited that if some individuals in a game do *not* follow the prescriptions of full rationality involving the maximization of expected objective outcomes to self, other fully rational players might then adopt cooperative strategies at least in the early stages of a game so as to gain the benefits of engaging in reciprocal cooperation. Second, Fudenberg and Maskin (1986) posited that it was possible for subjects to eliminate free riding if some players made a firm commitment to follow a "grim trigger strategy." A grim trigger strategy involves a permanent switch from cooperation to defection once anyone fails to cooperate.

These theoretical results have held up over the years. Instead of generating a clear and better prediction, however, they have led to an explosion of the number of possible equilibria predicted by noncooperative game theory. Among the predicted equilibria are strategies yielding the suboptimal Nash equilibrium, the optimal outcome, and everything in between (Abreau 1988). Thus, while empirical evidence generates some optimism that collective action can be achieved in some settings, the *problem* of collective action remains: How can participants avoid the temptation of suboptimal equilibria and move closer to optimal outcomes—in other words, gain a "cooperators' dividend" (Lichbach 1996).

Developing a coherent theory of collective action related to the use of common-pool resources is a real challenge. At the individual level, individuals do take costly actions that effectively take the interests of others into account. Shivakumar (2005) and Gellar (2005) provide evidence of local and regional groups that are successfully engaging in collective action in Somaliland and in Senegal where little cooperation occurred earlier. On the other hand, individuals may callously ignore or viciously harm others depending on the setting in which they find themselves (see Fiske, Harris, and Cuddy 2004).

Thus, an important task for all social scientists is achieving a more coherent synthesis of theoretical work that posit variables affecting the likelihood of undertaking diverse forms of collective action. We must be able to explain success as well as failure of efforts to achieve collective action. Further, we need to recognize that forms of collective action differ in regard to the distribution of benefits and harms to those in a group and those who are external to it. Mobs, gangs, and cartels are forms of collective action as well as neighborhood associations, charities, and voting.

My line of attack on this immense topic will involve the following steps. First, I will discuss the growing and extensive theoretical literature positing a host of structural variables presumed to affect the likelihood of individuals achieving collective action to overcome social dilemmas. None of these structural variables, however, should really make any difference in the probability of successful collective action if we continue to treat the *model* of rationality that has proved successful in explaining behavior and outcomes in competitive market settings as a universal *theory* of human behavior. Thus, the second major section of the paper will examine how a theory of boundedly rational, norm-based human behavior is a better foundation for explaining collective action than a model of maximizing material payoffs to self. If one posits that individuals can use reciprocity and reputations to build trust in dilemma situations, then one can begin to explain *both* successful and unsuccessful efforts to overcome social dilemmas through collective action.

The third section of the paper will then briefly examine the linkage between the structural measures discussed in the first section with the core individual relationships discussed in the second. In conclusion, I will reflect on the challenge that social scientists face in testing collective-action theory in light of the large number of variables posited to affect outcomes.

Structural Variables Predicted to Affect the Likelihood of Collective Action

A rich array of theoretical speculations, formal game-theoretic models, and computer models of evolutionary processes have generated a long list of structural variables that are frequently postulated to affect the likelihood that a set of participants will be able to achieve outcomes greater than the deficient Nash equilibrium—or, the cooperators' dividend (Lichbach 1996). Let us first focus on structural variables that do not essentially depend on a situation being repeated. These include:

1. the number of participants involved;
2. whether benefits are subtractive or fully shared (i.e., public goods vs common-pool resources)
3. the heterogeneity of participants; and

4. face-to-face communication.

Then, we will focus on situations where repetition of the situation makes possible the impact of additional structural variables including:

6. information about past actions;
7. how individuals are linked; and
8. whether individuals can enter or exit voluntarily.

Let us turn to a brief discussion of these eight major variables and how they are posited to affect the possibility of collective action and the size of benefits achieved.

Situations Where Repetition is Not Relevant

Among the variables that are posited to affect the likelihood of participants overcoming a social dilemma are four variables considered to be important whether or not the situation is repeated: the number of participants, whether benefits are subtractive or fully shared, their heterogeneity, whether they can communicate, and the shape of the production function they face.

The Number of Participants Involved

In his influential book *The Logic of Collective Action*, Mancur Olson (1965) argued that as the size of a group increased, the probability of a group achieving a public good decreased and the extent of nonoptimality increased—for two reasons. First, as group size increases, the noticeability of any single input to the provision of a public good declines. It is then easier for the individual to think that their own free riding will not be noticed and thus it will not affect the likelihood that the good will be provided. Second, coming to an internal agreement about coordinated strategies in larger groups involves higher transaction costs. Thus, a core theoretical hypothesis has been that the number of participants will likely reduce the probability of achieving any form of collective action or at least diminish the amount of joint benefits that could be achieved.

Other scholars have been less positive about the effect of group size. Agrawal (2000), for example, posits a curvilinear relationship between size of group and collective action in light of his study of community forest regimes in India. If the group is very large, transaction costs and conflict may arise. If the group is too small, it is hard to generate the resources needed to engage effectively in collective action related to a forest. Thus, moderately sized groups are more able to solve these problems when related to the governance and management of many natural resources.

Chamberlin (1974) pointed out that differences in group size frequently affect other key variables including the marginal impact of an individual's contribution of a fixed amount (see also Frohlich and Oppenheimer 1970; Hardin 1982; Pecorino 1999). Thus, how size might affect the likelihood of cooperation depends on how other structural variables are affected by the size of a group (see also Ostrom 2001).

Subtractive versus Fully Shared Benefits

Olson originally included all dilemmas where it was difficult to exclude potential beneficiaries, whether or not they had contributed. This analysis confounded situations where the consumption of benefits by one individual subtracted benefits from others with situations where consumption was nonsubtractive in nature (characterized as having full jointness of supply—see V. and E. Ostrom 1999). In a public good environment, increasing the number of participants tends to bring additional resources that could be drawn on to provide a benefit that will be jointly enjoyed by all. It is because of the additional resources available in a larger group and the nonsubtractability characteristic of public goods, that Marwell and Oliver (1993, 45) conclude that when “a good has pure jointness of supply, group size has a *positive* effect on the probability that it will be provided.”

Goods that are subtractable in nature are better defined as common-pool resources (CPRs) (Ostrom, Walker, and Gardner 1992). Social dilemmas related to CPRs share with public good provision the problems of free riding, but they also include the problems of overharvesting and crowding. Important types of CPRs include forests, water systems, and pastures. In a CPR environment, an increase in the number of participants, holding other variables constant, is negatively related to achieving social benefits.

The Heterogeneity of Participants

Participants can be heterogeneous in many ways. Olson (1965) argued that if there were one or a few individuals who had much stronger interests in achieving a public good (in other words, they faced different payoff functions), the probability of a group achieving a public good increased even though the good was still likely to be underprovided. Others have speculated that heterogeneity in assets, information, and payoffs are negatively related to gaining a cooperators’ dividend due principally to increased transaction costs and the conflict that would exist over the distribution of benefits and costs to be borne. In fact, the literature contains many arguments that point to heterogeneity as a serious deterrent to cooperation (Hardin 1982; Johnson and Libecap 1982; Libecap and Wiggins 1984; Isaac and Walker 1988; Kanbur 1992; Bardhan 1993; Seabright 1993).

Face-to-Face Communication

Given that noncooperative game theory predicts that communication will make *no* difference in the outcome of social dilemmas, the repeated findings of a strong positive effect that communication has on the outcomes of collective-action experiments is a major theoretical puzzle (Sally 1995). The result has been replicated so many times, however, that contemporary scholars have to take it seriously.

Adolphs et al. (1996) posited that the brain of one person unconsciously processes information about the emotional state from the facial expressions of another person with whom they are interacting. Frohlich and Oppenheimer (1998) explain the effectiveness of communication in general related to the needs of individuals in such settings to express the desire to each other that they should forego their immediate self-interest for the benefit of the group. In other words, communication is used for "moral suasion." And, being able to look others directly in the eye while discussing such moral issues is substantially better than relying on written communication. Kerr and Kaufman-Gilliland (1994) conclude that communication in general helps a group gain a sense of "solidarity" and that face-to-face communication enhances the likelihood that individuals will keep their promises to cooperate. In general, the efficacy of communication appears to be related to the increased trust that individuals acquire when promises are made to them in a face-to-face setting. When they are in a repeated situation, they use the opportunity for communication to discuss deviations from promises made in a highly critical and moralistic tone (Ostrom, Gardner, and Walker 1994; Parks, Henager, and Scamahorn 1996; Valley, Moag, and Bazerman 1998).

Repetition of Interactions

With repeated interactions, at least three more structural variables are posited to affect the level of cooperation achieved in social dilemma situations: the level of information generated about past actions, how individuals are linked, and voluntary entry and exit.

Information about Past Actions

The amount of information that an individual can obtain about the earlier actions of others can make a substantial difference when choosing strategy in a repeated situation. In a two-person game where individuals know the structure of the game and learn accurate information about the outcomes achieved, the behavior of the other individual is also known. As soon as more than two individuals are involved, accurate information about outcomes alone is no longer sufficient to inform one player about the actions of others. In families and small farming neighborhoods, where interactions are repeated, reputations can be built over time and group members can build up a level of trust about other participants (Seabright 1993). Cooperation can grow over time in such settings. In large groups, the disjunction between an individual's actions and reputations is more difficult to overcome. In some situations, individuals can observe the actions of others and thus know what each individual did in the previous rounds. Various ways of monitoring the actions of participants increase or decrease the availability and accuracy of the information that individuals have concerning the particular actions of known individuals (or types of players) in the past (Janssen 2004).

How Individuals are Linked

Sociologists and social psychologists have stressed the importance of how individuals may or may not be linked in a network when confronting various types of social dilemmas (Granovetter 1973; Cook and Hardin 2001).[1] They have posited that individuals who are linked in a network where A contributes resources to B, and B contributes resources to C, and C contributes resources to A—or any similar unidirectional linking—are more likely to contribute to each other's welfare than individuals whose resource contribution goes to a generalized pool from which all individuals obtain benefits. The reason given for this expectation is that individuals in an undifferentiated group setting can expect to free ride for a longer period of time without reducing their own benefits than when contributions have to be delivered to someone in the chain of relationships in order for benefits to eventually come to them. Anyone in the chain who stops contributing faces a higher probability (so the argument goes) of the chain of benefit-enhancing contributions stopping and their losing out on obtaining a positive benefit. Creating a particular type of network may change the structure of the game from an n-person PD to an Assurance Game (Yamagishi and Cook 1993).

The Possibility of Choosing Whether to Play or Not (Entry and Exit)

Orbell and Dawes (1991) and Hauk and Nagel (2001) have argued that when individuals have a choice as to whether to play social dilemma games with others, and they can identify the individuals with whom they have played and have a memory of past history, that individuals will choose partners so as to increase the frequency with which cooperative outcomes are achieved. This gives individuals a third choice in a social dilemma game. Besides deciding whether to cooperate, they can decide whether to "opt out." If one player opts out, the decision round ends, and everyone receives a zero payoff. All players have an effective veto over the entire play of the game.

Janssen (2005) has developed an agent-based model of a two-person, prisoner's dilemma in which individuals can cooperate, defect or withdraw. Each agent carries symbols that can be identified by others. The symbols are used by participants to gain or lose trust that the other participant will cooperate. Given this capacity to recognize trustworthiness in others *and* the capacity to withdraw from playing a game at all, cooperation levels rise over time and reach relatively high levels in populations composed of 100 players. With 1,000 players, cooperation levels are lower unless the number of symbols that can be used to recognize trustworthy plays is increased—a somewhat counterintuitive result (see also Hauert et al. 2002).

Towards a More General Theory of Human Behavior

As is by now obvious from the above discussion, the earlier image of individuals stuck inexorably within social dilemmas has slowly been replaced in some theoretical work with a recognition that individuals face the *possibility* of achieving results that avoid the worst outcomes and, in some situations, may even approximate optimality. The clear and unambiguous predictions of earlier theories have been replaced with a broad range of predictions including some that are far more optimistic. The theoretical enterprise has, however, become more opaque and confused.

This is a particularly challenging puzzle for scholars who yearn for frameworks and theories of behavior that integrate across the social sciences. To have one theory—rational choice theory—that explains how individuals achieve close to optimal outcomes in markets, but fails to explain why anyone votes or contributes voluntarily to the provision of public goods, is not a satisfactory state of knowledge in the social sciences. Simply assuming that individuals are successfully socialized into seeking better group outcomes does not explain the obvious fact that groups often fail to obtain jointly beneficial outcomes (Dietz, Ostrom, and Stern 2003).

We need to recognize that what has come to be called rational choice *theory* is instead one *model* in a family of models useful for conducting formal analyses of human decisions in highly structured settings. It is a rather thin model of a broader theory of rational behavior. When it is used successfully, the rational choice model is largely dependent for its power of explanation on how the structure of the situations involved is modeled (Satz and Ferejohn 1994). In other words, the context within which individuals face social dilemmas is more important in explaining levels of collective action than relying on a single model of rational behavior as used in classical noncooperative game theory (see Orbell et al. 2004).

In highly structured and competitive environments, predictions generated from the *combination* of a model of the situation and a model of complete rationality are well-supported empirically. As Alchian (1950) demonstrated long ago, competitive markets eliminate businesses that do not maximize profits. Further, markets generate limited, but sufficient, statistics needed to maximize profits. The institutional structure of a market rewards individuals who make economically rational decisions and who can then be modeled as if they were determinate, calculating machines.

A broader theory of human behavior views humans as adaptive creatures (B. Jones 2001) who attempt to do as well as they can given the constraints of the situations in which they find themselves (or the ones that they seek out) (Simon 1955; 1957; 1999). Humans learn norms, heuristics, and full analytical strategies from one another, from feedback from the world, and from their own capacity to engage in self-reflection and imagine a differently structured world. They are capable of designing new tools—including institutions—that can change the structure of the worlds they face for good or evil purposes. They adopt both short-term and long-term perspectives dependent on the structure of opportunities they face. Multiple models are consistent with a theory of boundedly rational human behavior, including a model of complete rationality when paired with repetitive, highly competitive situations.

Heuristics and Norms

Many situations in life, however, do not generate information about all potential actions that one could take, all outcomes that could be obtained, and all strategies that others could take. One simply assumes this level of information when using a model of complete rationality. In most everyday situations individuals tend to use heuristics—rules of thumb—that they have learned over time regarding responses that tend to give them good (but, not necessarily optimal) outcomes in particular kinds of situations. In frequently encountered, repetitive situations, individuals learn better and better heuristics that are tailored to the particular situation. With repetition and sufficiently large stakes, individuals may learn heuristics that approach best-response strategies and thus approach local optima (Gigerenzer and Selten 2001).

In addition to learning instrumental heuristics, individuals also learn norms. By norms, I mean that the individual attaches an internal valuation—positive or negative—to taking particular types of action. Analytically, individuals can be thought of as learning norms of behavior that are relatively general and fit a wide diversity of particular situations. Crawford and Ostrom (2005) refer to this internal valuation as a delta parameter that is added to or subtracted from the objective costs of an action or an outcome. Andreoni (1989) models individuals who gain a "warm glow" when they contribute resources that help others more than they help themselves in the short term. Knack (1992) refers to negative internal valuations as "duty." The strength of the commitment (Sen 1977) made by an individual to take particular types of future actions (telling the truth, keeping promises), is reflected in the size of the delta parameter. After experiencing repeated benefits from their own and from other people's cooperative actions, individuals may resolve that they should always initiate cooperation in the future.[2] Or, after many experiences of being the "sucker" in such experiences, an individual may resolve never to initiate unilateral cooperation and to punish noncooperators whenever feasible.

James Cox and colleagues posit that individual behavior in a particular setting is affected by an individual's initial emotional or normative state and then by direct experience with others in a specific setting (Cox 2004). The underlying norms and direct experience in a particular setting combine to affect orientations toward reciprocity. "Instead of beliefs or type estimates we use emotional states based on actual experience: my attitude toward your payoffs depends on my state of mind, e.g., kind or vengeful, and your actual behavior systematically alters my emotional state" (Cox, Friedman, and Gjerstad 2004, 1).

Fairness is also one of the norms used by individuals in social dilemma settings. The maximal net return to a group may be obtained in a manner that is perceived to be fair or unfair by those involved—using the general concept that "equals should be treated equally and unequals unequally" (see Isaac, Mathieu, and Zajac 1991). When participants are symmetric in regard to all strategically relevant variables, the only real fairness issue relates to the potential capability of some to free ride on others (Dawes, Orbell, and van de Kragt 1986). When participants differ, however, finding an allocation formula perceived by most participants as fair is far more challenging (Rawls 1971). In both cases, however, theorists have argued that when participants think that a proposal for sharing costs and benefits is fair, they are far more willing to contribute (Isaac, Mathieu, and Zajac 1991).

Since norms are learned, they vary substantially across individuals, and within individuals across the different types of situations they face, and across time within any particular situation. As Brennan and Pettit (2004) stress, however, norms that help to solve social dilemmas need to be shared so that individuals who act contrary to the norm fear the reduction in esteem likely to occur. Once some members of a population acquire norms of behavior, they affect the expectations of others. When interacting with individuals who are known to use retribution against those who are not trustworthy, one is better off by keeping one's commitments.

Contingent Strategies and Norms of Reciprocity

Many theorists posit that one can explain behavior in social dilemmas better if one assumes that boundedly rational individuals enter situations with an initial probability of using reciprocity based either as a calculated strategy that contingent action leads one to be better off *or* based on a normative belief that this is how one should behave (Fehr and Gächter 2000; Bolton and Ockenfels 2000; Falk, Fehr, and Fischbacher 2002; Panchanathan and Boyd 2004). In either case, individuals learn to use reciprocity based on their own prior training and experience. The more benefits that they have received in the past from other reciprocators, the higher their own initial inclinations. The more they have faced retribution, the less likely they estimate that free riding is an attractive option. Their trust that others will also be reciprocators is highly correlated with their own norms but is affected by the information they glean about the reputation of other players and their estimate of the risk of extending trust given the structure of particular situation.

By and far the most famous contingent strategy—tit-for-tat—has been the subject of considerable study from an evolutionary perspective. Axelrod and Hamilton (1981) and Axelrod (1984) have shown that when individuals are grouped so that they are more likely to interact with one another than with the general population, and when the expected number of interactions is sufficiently large, reciprocating strategies such as tit-for-tat can successfully invade populations composed of individuals following an all-defect strategy.

Boyd and Richerson (1992) build a two-stage evolutionary model based on Hirshleifer and Rasmusen's (1989) model of a large population from which groups of size $n > 2$ are selected. The first stage is an n-person PD where an individual selects cooperate or defect. In the second stage, any individual can punish any other individual at a cost to the punisher and to the punished. The same group continues for the next round dependent on a probability function. Strategies are modeled as if they were inherited. They allow errors to occur in the execution of a cooperative strategy, but all other strategies are executed as intended. After the rounds of interaction are completed, the more successful strategies are reproduced at a higher rate than the less successful strategies.

In the Boyd and Richerson (1992) model, an increase in group size requires an offsetting linear increase in the number of interactions to achieve similar levels of collective action (see also Richerson and Boyd 2005). They also find that moralistic strategies, "which punish defectors, individuals who do not punish noncooperators, and individuals who do not punish nonpunishers can also overcome the problem of second-order cooperation" (1992, 184). When moralistic strategies are common, defectors and cooperators who do not punish are selected against due to the punishment directed at them. "In this way, selection may favor punishment, even though the cooperation that results is not sufficient to compensate individual punishers for its costs" (ibid.). These moralistic strategies can stabilize any behavior—a result that is similar to the famous "folk theorem" that any equilibrium can be stabilized by such punishing strategies as the grim trigger. Yamagishi and Takahashi (1994) explore in an evolutionary simulation whether linking sanctioning to cooperative actions so that cooperators punish defectors and defectors do not punish other defectors solves the problem of aggressive moralistic strategies or meta norms. When these strategies are linked, they find close to 100% cooperation.

Several of the heuristics or strategies posited to help individuals gain larger cooperators' dividends depend upon the willingness of participants to use retribution to at least some degree. In tit-for-tat, for example, an individual must be willing to "punish" a player who defected on the last round by defecting on the current round. As mentioned above, the grim trigger is a strategy that cooperates with others until someone defects, and then defects the rest of the rounds (Fudenberg and Maskin 1986). In repeated games where substantial joint benefits are to be gained from mutual cooperation, the threat of the grim trigger is posited to encourage everyone to cooperate. A small error on the part of one player or exogenous noise in the payoff function, however, makes this strategy a very dangerous one to use in large environments where the cooperators' dividend is substantial.

The Core Relationships: Reputation, Trust, and Reciprocity as They Affect Cooperation

In situations where individuals can acquire a reputation for using positive and negative reciprocity and being trustworthy, others can learn to trust those with such a reputation and begin to cooperate—as long as others also cooperate (Fukuyama 1995). Thus, at the core of an evolving theoretical explanation of successful or unsuccessful collective action are the links between the trust that one participant (P_i) has in the others ($P_j \ldots P_n$) involved in a collective-action situation, the investment others make in trustworthy reputations, and the probability of all participants using reciprocity norms (see Figure 1). When some individuals initiate cooperation in a repeated situation, others learn to trust them and are more willing to adopt reciprocity themselves leading to higher levels of cooperation. And, when more individuals use reciprocity, gaining a reputation for being trustworthy is a good investment as well as an intrinsic value. Thus, reputations for being trustworthy, levels of trust, and reciprocity are positively reinforcing. This also means that a decrease in any one of these can generate a downward cascade leading to little or no cooperation.

[Figure 1 about here]

The Contemporary Agenda: Linking Structural Variables to the Core Relationships

Instead of explaining cooperation directly from the material incentives facing individuals in social dilemmas, the contemporary task we face is how to link external structural variables to an inner core of individual level variables—reputation, trust, and reciprocity—as these in turn affect levels of cooperation and net benefits achieved. We already understand some of the potential linkages. For example, one can confidently posit that in a small, homogeneous group interacting in a face-to-face meeting to discuss producing a public good, the costs of coming to an agreement will tend to be low and the probability that individuals keep their promises will be high. Previous gossip will have identified which members of the group could be trusted to keep agreements and efforts to exclude such untrustworthy participants would be undertaken. The combined effect of the structural variables in this example on reputation, trust, and reciprocity is likely to overcome short-term, material benefits that individual participants are tempted to pursue. In a different context—a large, heterogeneous group with no communication and no information about past trustworthiness who jointly use a common-pool resource—individuals will tend to pursue short-term material benefits and potentially destroy the resource.

Thus, using a broader theory of human behavior that includes the possibility that participants use reciprocity and cooperate in social dilemmas when they trust others to do the same, enables scholars to generate testable hypotheses based on combinations of structural variables as they interact to increase or decrease the likelihood of cooperation and net benefits occurring (see Weber, Kopelman, and Messick 2004 for a similar effort). It is not possible, however, to link all of the structural variables identified above in a one definitive causal model given the large number of variables and that many of them depend for their impact on the value of other variables. For now, it is possible to illustrate this general approach with the framework shown in Figure 2 where the structural variables discussed above are linked in a general way to the core relationships.

[Figure 2 about here]

One cannot assign a fixed direction of relationships in this approach, however, given that the sign depends on the configuration of other variables in a particular focal collective action problem. A small group with extreme heterogeneity in the benefits to be obtained from a collective action, for example, is an entirely different group than a small group of relatively homogeneous players. Further, in a small group with extreme heterogeneity, face-to-face communication may lead to exacerbated conflict rather than reduction in conflict and agreement on new sets of rules. Instead of one large, general causal model, one can develop specific scenarios of causal direction, such as those posited above, that can be tested (see Ostrom 1998). Thus, an important next step in the development of collective-action theory is more careful attention to how structural variables interact with one another. One cannot posit simple explanations based upon an assumption that size alone makes a difference, that heterogeneity alone makes a difference, that a step level production function alone makes a difference, or the capacity to exit alone makes a difference—all proposed by some scholars as the primary variable one needs to examine. It is the combination of these variables that evoke norms, help or hinder building reputations and trust, and enable effective or destructive interactions and learning to occur. What is important about this simple and general framework is recognition that at any one time multiple variables affect the core variables of reputation, trust, and reciprocity.

Further, the variables linked together on Figure 2 are not an exhaustive set of all structural variables posited to affect collective action—they are the set that appears to be most frequently mentioned in the general literature reviewed above. Many of these variables are posited to affect other intermediate variables—such as transaction costs and the development of shared norms—that in turn affect the probability of cooperation.

Still other variables are identified in more specialized work. Agrawal (2002) has, for example, identified more than 30 variables posited by scholars studying collective action related to organizing the governance of common-pool resources. Many of the variables he identifies have interactional effects. Agrawal (2002, 68–70) develops several causal chains to connect a subset of these variables together for testing in field and laboratory settings. Some of the variables identified by Agrawal relate to the likelihood of participants changing the rules that affect the structural variables that, in turn, affect the core relationships.

Conclusions

A key lesson of research on collective-action theory is recognizing the complex linkages among variables at multiple levels that together affect individual reputations, trust and reciprocity as these, in turn, affect levels of cooperation and joint benefits. Conducting empirical research on collective action is thus extremely challenging (see Poteete, Janssen, and Ostrom, forthcoming). There is no way that one can analyze the entire "spaghetti plate" of variables that have been identified and their interactions in a single empirical analysis. The reason that experimental research has become such an important method for testing theory is that it is a method for controlling the setting of many variables while changing only one or two variables at a time (Camerer 2003). In addition, one can self-consciously examine the interaction of several variables over a series of carefully designed experiments—something that is almost impossible to do in field research.

Conducting research in similar environments that differ in regard to one or two key variables is also an important strategy, but very difficult to find such settings. Large-N research on collective action is a challenge both in terms of obtaining accurate and consistent data, but also because of the large number of variables that potentially affect any one type of collective action (Poteete and Ostrom 2004). Instead of looking at all of the potential variables, one needs to focus in on a well defined but narrow chain of relationships—as recommended by Agrawal (2002). One can then conduct analysis of a limited set of variables that are posited to have a strong causal relationship (for examples, see Gibson, Williams, and Ostrom 2005; Hayes and Ostrom 2005). Thus, the theory of collective action is not only one of the most important subjects for economists and other social scientists, it is also one of the most challenging.

Notes

[1] If the linkage structure is that of a pure hierarchy, it is presumed by many theorists that the dilemma disappears through the exercise of command and control mechanisms.

[2] Whenever games are repeated, the discount rates used by individuals also affects the adoption of norms including that of reciprocity. In settings where individuals do not strongly discount outcomes that will occur in the distant future, they can realize the benefits of cooperation over a long series of plays—thus offsetting the initial material advantage of not cooperating. As the future is more strongly discounted, however, the calculation made by an individual focuses more on the immediate materials payoffs. Thus, a delicate relationship exists between the discount rates used by individuals, the size of the potential benefit to be achieved, and the willingness of individuals to accept the norm of reciprocity (Abreau 1988; Axelrod 1984; Curry, Price, and Price 2005).

References

Abreau, D. 1988. On the theory of infinitely repeated games with discounting. *Econometrica,* 80(4): 383-96.

Adolphs, R., Damasio, H., Tranel, D., and Damasio, A. R. 1996. Cortical systems for the recognition of emotion in facial expressions. *The Journal of Neuroscience,* 16(23): 7,678-87.

Agrawal, A. 2000. Small is beautiful, but is larger better? Forest-management institutions in the Kumaon Himalaya, India. Pp. 57-86 in *People and Forests: Communities, Institutions, and Governance,* eds. C. Gibson, M. McKean, and E. Ostrom. Cambridge, Mass.: MIT Press.

___. 2002. Common resources and institutional sustainability. Pp. 41-85 in *The Drama of the Commons,* National Research Council, Committee on the Human Dimensions of Global Change, eds. E. Ostrom, T. Dietz, N. Dolšak, P. C. Stern, S. Stonich, and E. Weber. Washington, D.C.: National Academy Press.

Alchian, A. A. 1950. Uncertainty, evolution, and economic theory. *Journal of Political Economy,* 58(3): 211-21.

Andreoni, J. 1989. Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy,* 97: 1447-58.

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

___ and Hamilton, W. D. 1981. The evolution of cooperation. *Science,* 211: 1390-96.

Bardhan, P. 1993. Analytics of the institutions of informal cooperation in rural development. *World Development,* 21(4): 633-39.

Bolton, G. E., and Ockenfels, A. 2000. ERC: A theory of equity, reciprocity, and competition. *American Economic Review,* 90: 166-93.

Boyd, R. and Richerson, P. J. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology,* 13: 171-95.

Brennan, G. and Pettit, P. 2004. *The Economy of Esteem*. Oxford: Oxford University Press.

Camerer, C. F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, N.J.: Princeton University Press.

Chamberlin, J. 1974. Provision of collective goods as a function of group size. *American Political Science Review,* 68: 707-16.

Cook, K. S. and Hardin, R. 2001. Norms of cooperativeness and networks of trust. Pp. 327-357 in *Social Norms*, ed. M. Hechter and K.-D. Opp. New York: Russell Sage Foundation.

Cox, J. C. 2004. How to identify trust and reciprocity. *Games and Economic Behavior,* 46: 260-81.

___, Friedman, D. and Gjerstad, S. 2004. A tractable model of reciprocity and fairness. Working Paper. Tucson: University of Arizona, Department of Economics

Crawford, S. E. S. and Ostrom, E. 2005. A grammar of institutions. Pp. 137-74 in *Understanding Institutional Diversity*, ed. E. Ostrom. Princeton, N.J.: Princeton University Press. Originally published in *American Political Science Review* 89(3) (1995): 582-600.

Curry, O. S., Price, M. E. and Price, J. G. 2005. Is patience a virtue? Individual differences in discount rates and cooperativeness. Paper presented at the 17th meeting of the Human Behavior and Evolution Society, Austin, Texas.

Dawes, R. M., Orbell, J. M. and van de Kragt, A. 1986. Organizing groups for collective action. *American Political Science Review,* 80(4): 1171-85.

Dietz, T., Ostrom, E. and Stern, P. 2003. The struggle to govern the commons. *Science,* 302(5652) (December 12): 1907-12.

Falk, A., Fehr, E. and Fischbacher, U. 2002. Appropriating the commons: A theoretical explanation. Pp. 157-92 in *The Drama of the Commons,* National Research Council, Committee on the Human Dimensions of Global Change, ed. E. Ostrom, T. Dietz, N. Dolšak, P. C. Stern, S. Stonich, and E. Weber. Washington, D.C.: National Academy Press.

Fehr, E., and Gächter, S. 2000. Fairness and retaliation: the economics of reciprocity. *Journal of Economic Perspectives,* 14(3): 159-81.

Fiske, S. T., Harris, L. T. and Cuddy, A. J. C. 2004. Why ordinary people torture enemy prisoners. *Science*, 306: 1482-83.

Frohlich, N. and Oppenheimer, J. 1970. I get by with a little help from my friends. *World Politics,* 23: 104-20.

___. 1998. Some consequences of e-mail vs. face-to-face communication in experiment. *Journal of Economic Behavior and Organization*, 35: 389-403.

Fudenberg, D., and Maskin, E. 1986. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3): 533-54.

Fukuyama, F. 1995. *Trust: The Social Virtues and the Creation of Prosperity.* New York: Free Press.

Gellar, S. 2005. *Democracy in Senegal: Tocquevillian Analytics in Africa*. New York: Palgrave Macmillan.

Gibson, C. C., Williams, J. T. and Ostrom, E. 2005. Local enforcement and better forests. *World Development* 33(2) (February): 273-84.

Gigerenzer, G., and Selten, R., eds. 2001. *Bounded Rationality: The Adaptive Toolbox*. Cambridge, Mass.: MIT Press.

Granovetter, M. 1973. The strength of weak ties. *American Journal of Sociology*, 78: 1360-80.

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

Hauert, C., De Monte, S., Hofbauer, J. and Sigmund, K. 2002. Volunteering as red queen mechanisms for cooperation in public good games. *Science*, 296 (May 10): 1129-32.

Hauk, E. and Nagel, R. 2001. Choice of partners in multiple two-person prisoner's dilemma games: An experimental study. *Journal of Conflict Resolution,* 45: 770-93.

Hayes, T. M. and Ostrom, E. 2005. Conserving the world's forests: Are protected areas the only way? *Indiana Law Review* 37(3): 595-617.

Hirshleifer, J. and Rasmusen, E. 1989. Cooperation in a repeated prisoner's dilemma with ostracism. *Journal of Economic Behavior and Organization,* 12: 87-106.

Isaac, R. M., Mathieu, D. and Zajac, E. E. 1991. Institutional framing and perceptions of fairness. *Constitutional Political Economy,* 2(3): 329-70.

Isaac, R. M. and Walker, J. M. 1988. Communication and free-riding behavior: The voluntary contribution mechanism. *Economic Inquiry*, 26(4): 585-608.

Janssen, M. A. 2004. Evolution of cooperation when feedback to reputation scores is voluntary. Working Paper. Bloomington: Indiana University, Center for the Study of Institutions, Population, and Environmental Change (CIPEC).

___. 2005. Evolution of cooperation in a one-shot prisoner's dilemma based on recognition of trustworthy and untrustworthy agents. Working Paper. Bloomington: Indiana University, Center for the Study of Institutions, Population, and Environmental Change (CIPEC).

Johnson, R., and Libecap, G. 1982. Contracting problems and regulation: The case of the fishery. *American Economic Review*, 72 (Dec.): 1005-23.

Jones, B. D. 2001. *Politics and the Architecture of Choice: Bounded Rationality and Governance*. Chicago: University of Chicago Press.

Kagel, J. H. and Roth, A. E. 1995. *The Handbook of Experimental Economics*. Princeton, N.J.: Princeton University Press.

Kanbur, R. 1992. Heterogeneity, distribution and cooperation in common property resource management. Background Paper for the World Development Report. Mimeograph.

Kerr, N. L., and Kaufman-Gilliland, C. M. 1994. Communication, commitment and cooperation in social dilemmas. *Journal of Personality and Social Psychology,* 66: 513-29.

Knack, S. 1992. Civic norms, social sanctions, and voter turnout. *Rationality and Society,* 4 (April): 133-56.

Kreps, D., Milgrom, P., Roberts, J. and Wilson, R. 1982. Rational cooperation in finitely repeated prisoners' dilemma. *Quarterly Journal of Economics*, 103: 179-200.

Libecap, G., and Wiggins, S. 1984. Contractual responses to the common pool: Prorationing of crude oil production. *American Economic Review,* 74: 87-98.

Lichbach, M. I. 1996. *The Cooperator's Dilemma.* Ann Arbor: University of Michigan Press.
Marwell, G. and Oliver, P. 1993. *The Critical Mass in Collective Action: A Micro-Social Theory*. New York: Cambridge University Press.
Olson, M. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups.* Cambridge, Mass.: Harvard University Press.
Orbell, J. M. and Dawes, R. M. 1991. A 'cognitive miser' theory of cooperators' advantage. *American Political Science Review,* 85 (2): 515-28.
Orbell, J. M., Morikawa, T., Hartwig, J., Hanley, J. and Allen, N. 2004. 'Machiavellian' Intelligence as a Basis for the Evolution of Cooperative Dispositions. *American Political Science Review,* 98(1): 1-15.
Ostrom, E. 1998. A behavioral approach to the rational choice theory of collective action. *American Political Science Review,* 92(1) (March): 1-22.
___. 2001. Social dilemmas and human behavior. Pp. 21-41 in *Economics in Nature: Social Dilemmas, Mate Choice and Biological Markets,* ed. R. Noë, J. Van Hooff, and P. Hammerstein. Cambridge: Cambridge University Press.
___. 2007. Collective action theory. Pp. 186–208 in *The Oxford Handbook of Comparative Politics*, ed. Carles Boix and Susan Stokes. Oxford, UK: Oxford University Press.
___, Gardner R. and Walker, J. 1994. *Rules, Games, and Common-Pool Resources*. Ann Arbor: University of Michigan Press.
___, Walker, J. and Gardner, R. 1992. Covenants with and without a sword: Self-governance is possible. *American Political Science Review,* 86(2): 404-17.
Ostrom V. and Ostrom, E. 1999. Public goods and public choices. Pp. 107-18 in *Polycentricity and Local Public Economies: Readings from the Workshop in Political Theory and Policy Analysis,* ed. M. McGinnis. Ann Arbor: University of Michigan Press. Originally published in E. S. Savas, ed. *Alternatives for Delivering Public Services: Toward Improved Performance,* 7-49 (Boulder, CO: Westview Press, 1977).
Panchanathan, K. and Boyd, R. 2004. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature,* 432 (November 25): 499-502.
Parks, C. D., Henager, R. F. and Scamahorn, S. D. 1996. Trust and reactions to messages of intent in social dilemmas. *Journal of Conflict Resolution,* 40: 134-51.
Pecorino, P. 1999. The effect of group size on public good provision in a repeated game setting. *Journal of Public Economics,* 72: 121-34.
Poteete, A., Janssen, M., and Ostrom, E. Forthcoming. *Multiple Methods in Practice: Collective Action and the Commons*. Draft chapters online at: http://commons.asu.edu/content/papers
Poteete, A. and Ostrom, E. 2004. Heterogeneity, group size and collective action: the role of institutions in forest management. *Development and Change,* 35(3): 437-61.
Rawls, J. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
Richerson, P. J. and Boyd, R. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution.* Chicago: University of Chicago Press.
Sally, D. 1995. Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958-1992. *Rationality and Society,* 7: 58-92.
Sandler, T. 1992. *Collective Action: Theory and Applications.* Ann Arbor: University of Michigan Press.
Satz, D. and Ferejohn, J. 1994. Rational choice and social theory. *Journal of Philosophy,* 91(2): 71-87.
Seabright, P. 1993. Managing local commons: Theoretical issues in incentive design. *Journal of Economic Perspectives,* 7(4): 113-34.
Sen, Amartya K. 1977. Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs,* 6: 317-44.
Shivakumar, S. J. 2005. *The Constitution of Development: Crafting Capabilities for Self-Governance*. New York: Palgrave Macmillan.
Simon, H. A. 1955. A behavioural model of rational choice. *Quarterly Journal of Economics,* 69: 99-188.
___. 1957. *Models of Man.* New York: Wiley.

___. 1999. The potlatch between political science and economics. Pp. 112-19 in *Competition and Cooperation: Conversations with Nobelists about Economics and Political Science*, eds. J. Alt, M. Levi, and E. Ostrom. New York: Russell Sage Foundation.
Valley, K. L., Moag, J. and Bazerman, M. H. 1998. A matter of trust: effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior and Organization,* 34(2): 211-38.
Weber, J. M., Kopelman, S. and Messick, D. M. 2004. A conceptual review of decision making in social dilemmas: Applying a logic of appropriateness. *Personality and Social Psychology Review*, 8: 281-307.
Yamagishi, T. and Cook, K. S. 1993. Generalized exchange and social dilemmas. *Social Psychology Quarterly,* 56(4): 235-48.
___ and Takahashi, N. 1994. Evolution of norms without metanorms. Pp. 311-26 in *Social Dilemmas and Cooperation,* eds. U. Schulz, W. Albers, and U. Mueller. Berlin: Springer-Verlag.

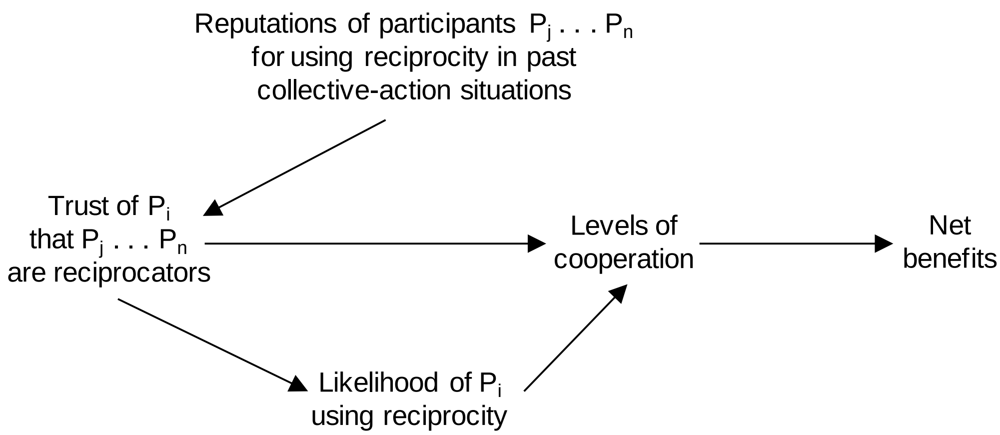

Figure 1. The core relationships at the individual level affecting levels of cooperation in a social dilemma.

Source: Ostrom (2007: 200).

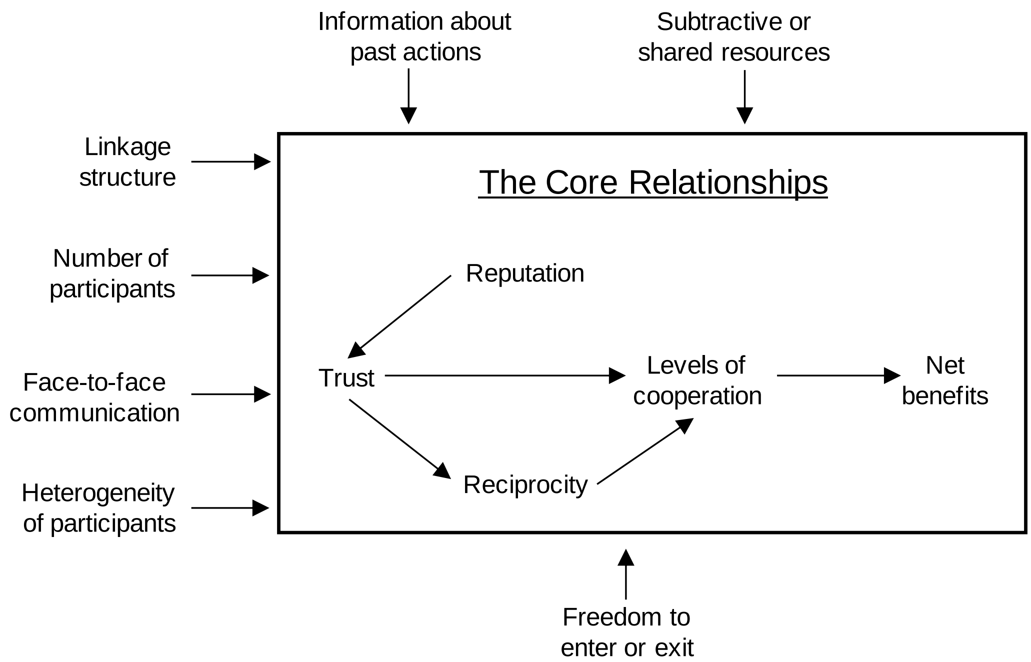

Figure 2. A framework linking structural variables to the core relationships in a focal dilemma arena.

Source: Ostrom (2007: 202).