Annu. Rev. Sociol. 1998. 24:183–214

SOCIAL DILEMMAS: The Anatomy of Cooperation

Peter Kollock
Department of Sociology, University of California at Los Angeles, Los Angeles, California 90095-1551; e-mail: KOLLOCK@UCLA.EDU

KEY WORDS: prisoner's dilemma, public good, commons, collective action

ABSTRACT

The study of social dilemmas is the study of the tension between individual and collective rationality. In a social dilemma, individually reasonable behavior leads to a situation in which everyone is worse off. The first part of this review is a discussion of categories of social dilemmas and how they are modeled. The key two-person social dilemmas (Prisoner's Dilemma, Assurance, Chicken) and multiple-person social dilemmas (public goods dilemmas and commons dilemmas) are examined. The second part is an extended treatment of possible solutions for social dilemmas. These solutions are organized into three broad categories based on whether the solutions assume egoistic actors and whether the structure of the situation can be changed: Motivational solutions assume actors are not completely egoistic and so give some weight to the outcomes of their partners. Strategic solutions assume egoistic actors, and neither of these categories of solutions involve changing the fundamental structure of the situation. Solutions that do involve changing the rules of the game are considered in the section on structural solutions. I conclude the review with a discussion of current research and directions for future work.

THE QUESTION OF COOPERATION

Social dilemmas are situations in which individual rationality leads to collective irrationality. That is, individually reasonable behavior leads to a situation in which everyone is worse off than they might have been otherwise. Many of the most challenging problems we face, from the interpersonal to the international, are at their core social dilemmas.

0360-0572/98/0815-0215/$0.800

As individuals we are each better off when we make use of a public resource, such as public television, without making any contribution, but if everyone acted on this conclusion, the public resource would not be provided and we would all be hurt. Each farmer does best by taking as much irrigation water as possible, and each fisher benefits from catching as many fish as possible, but the aggregate outcome of these individually reasonable decisions can be disaster—groundwater exhausted and fish species depleted to the point of extinction.

This review of the literature on social dilemmas is divided into two major sections. The first is a discussion of categories of social dilemmas and how they are modeled. The second is an extended treatment of possible solutions for social dilemmas. I conclude with a discussion of current research and directions for future work.

Any review of this length is necessarily selective. I have focused mainly but not exclusively on research since 1980 and on behavioral studies that use either experimental methods or field research. Even with these filters, the work I discuss is still a small sample; my goal is to provide a structure for understanding this area of research and a set of pointers to useful resources. For further information I would suggest other reviews by Komorita & Parks (1995, 1996), Ledyard (1995), Yamagishi (1995), van Lange et al (1992), Messick & Brewer (1983), Stroebe & Frey (1982), Orbell & Dawes (1981), Dawes (1980), Edney & Harper (1978a). A number of edited volumes are useful general resources, in particular the volumes that have come out of the biannual International Conference on Social Dilemmas: Liebrand & Messick (1996), Schulz et al (1994), Liebrand et al (1992), Wilke et al (1986). Other useful edited volumes include Schroeder (1995) and Hinde & Groebel (1991). An important set of field studies on social dilemmas can be found in Ostrom et al (1994), Bromley et al (1992), Ostrom (1990), McCay & Acheson (1987), Hardin & Baden (1977). There is even a popular press account of these issues in Poundstone (1992). Finally, a variety of resources are now available on the World Wide Web. I have collected a number of these sources at a Web page devoted to this review: www.sscnet.ucla.edu/soc/faculty/kollock/dilemmas (1998b).

MODELING SOCIAL DILEMMAS

All social dilemmas are marked by at least one *deficient equilibrium*. It is deficient in that there is at least one other outcome in which everyone is better off. It is an equilibrium in that no one has an incentive to change their behavior. Thus, at their worst, social dilemmas exemplify the true meaning of tragedy: "The essence of dramatic tragedy," wrote Whitehead, "is not unhappiness. It resides in the solemnity of the remorseless working of things" (quoted in Stroebe & Frey 1982). A group of people facing a social dilemma may com-

pletely understand the situation, may appreciate how each of their actions contribute to a disastrous outcome, and still be unable to do anything about it.

The most severe social dilemmas are also characterized by a *dominating strategy* that leads to a deficient equilibrium. A dominating strategy is a strategy that yields the best outcome for an individual regardless of what anyone else does. The compelling, and perverse, feature of these dilemmas is that there is no ambiguity about what one should do to benefit oneself, yet all are hurt if all follow this "rational" decision. However, not all social dilemmas involve dominating strategies, as we see below.

Necessary, Dangerous Metaphors

The literature in social dilemmas has revolved around three metaphorical stories that have assumed mythic proportions. These stories—the Prisoner's Dilemma, the problem of providing Public Goods, and the Tragedy of the Commons—have served as catalysts facilitating and structuring research. They have also served as blinders. The hegemony of these models has at times led researchers—or worse, policy makers—to believe mistakenly that these metaphors capture the whole range of social dilemmas or accurately model all empirical social dilemmas.

I deal with each of these models and their limitations below as well as with other models of social dilemmas that have traditionally received less attention. In categorizing social dilemmas, I make the first cut in distinguishing dilemmas that involve only two actors (known as dyadic or two-person dilemmas) from social dilemmas involving multiple actors (known as *N*-person dilemmas, where *N* is some number greater than two).

Two-Person Dilemmas

In 1950 Merrill Flood and Melvin Dresher—scientists at RAND Corporation in Santa Monica, California—carried out an informal experiment using a new game they had developed. The game was the simplest possible example of a social dilemma in that it involved only two people, each of whom faced a single choice between two options (termed cooperation and defection). Albert Tucker, a mathematician who was a colleague of theirs, created a story to go along with the game that involved two prisoners, which subsequently became known as the *Prisoner's Dilemma*—the game that launched a thousand studies (actually, several thousand).

The original story involves two prisoners who are separately given the choice between testifying against the other or keeping silent (see e.g. Luce & Raifa 1957). In my classroom I offer a simple example of the game: Two students are asked to take $1 out of their wallets. Each, in secret, decides whether to place the money in an envelope (cooperate) or to keep the money in one's

pocket (defect). Each envelope is then given to the other person, and I double whatever money has been given. The possible outcomes (in dollars) are seen in Figure 1*a*. The game is marked by the fact that whatever choice one's partner makes, one is better off defecting (i.e. defecting is a dominating strategy): If Player II cooperates, Player I's defection brings a payoff of $3 for Player I, and $0 for Player II. If Player II defects, Player I is still better off defecting, which yields a payoff of $1 for both. Since the payoff structure is identical for both actors, they converge on mutual defection even though both would be better off if they had cooperated, a move that gives both actors a payoff of $2. In other words, it is a deficient equilibrium. This dilemma is at the heart of unsecured transactions. For example, when I buy something through the mail, I may be tempted to not send a check and the other person may be tempted to not send the goods, but if we both defect, we are each worse off than if we had consummated the exchange.

What defines the Prisoner's Dilemma is the relative value of the four outcomes. The best possible outcome is defecting while one's partner cooperates (designated DC). The next best outcome is mutual cooperation (CC) followed by mutual defection (DD), with the worst outcome being the case in which one cooperates while one's partner defects (CD). Thus, in a Prisoner's Dilemma, DC > CC > DD > CD.[1]

Two other important games can be created by switching the relative value of the outcomes. If mutual cooperation leads to a better outcome than unilateral defection (CC > DC > DD > CD), the situation is known as an *Assurance Game*; an example of this game is shown in Figure 1*b*. The name comes from the fact that a person would be willing to cooperate as long as that person were assured that the partner would cooperate as well. A common misunderstanding is that an Assurance Game presents no dilemma and leads inevitably to mutual cooperation. In fact, cooperation is not a dominating strategy, and if the person believes the partner will defect, the best the person can do is to defect as well. In other words, the Assurance Game has two equilibria: mutual cooperation, which is an optimal equilibrium, and mutual defection, which is a deficient equilibrium.[2] I may be happy to work with you on preparing a joint report, and a report to which we have both contributed may be the best possible outcome for me, but if I cannot prepare the report myself and I do not believe you will cooperate, I am best off defecting as well. The key issue in the Assurance

[1]There is a second inequality that is also often included as part of the definition: CC > (CD + DC)/2. This inequality states that mutual cooperation is more profitable that alternating exploitation of self with exploitation of other.

[2]Technically, the equilibria I discuss in this section are known as Nash equilibria. A Nash equilibrium is "any pair of strategies with the property that each player maximizes his or her payoff given what the other player does" (Ostrom et al 1994, p. 54).

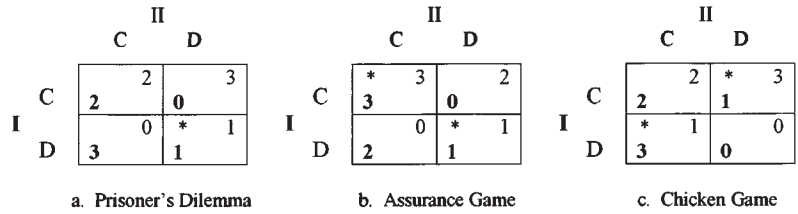

Figure 1 Three two-person games in their ordinal form: I and II designate Players I and II; C and D designate cooperation and defection. Player I's outcomes are shown in bold. Nash equilibria are designated with asterisks.

Game is whether we can trust each other. This game has received much less attention than the Prisoner's Dilemma Game, although I argue below that it is a more accurate model than the Prisoner's Dilemma Game of many social dilemma situations.

The third game discussed here is created by switching a different pair of outcomes in the Prisoner's Dilemma Game. If mutual defection yields a worse outcome than unilateral cooperation (DC > CC > CD > DD), we create the game of *Chicken*, which can be seen in Figure 1*c*. It is named after a game of dare that was made famous in the 1955 film *Rebel Without a Cause*. Two youths drive their cars toward each other (or in the case of the film, toward a cliff). The first youth to turn away is "chicken" and loses face, while the other youth basks in the glory of his courage. However, if neither youth turns away, they both end by dying—the worst outcome. If both turn away, the sting of being chicken is not as great since both drivers lost their nerve. There are two equilibria in the Chicken Game—unilateral defection and unilateral cooperation. If driving toward each other, you are sure the other person will lose their nerve and swerve, you are best off driving straight ahead, but if you believe the other person will not swerve, you are better off swerving and losing face rather than your life. In this sense, you have an advantage in this game if you can convince the other person that you are crazy, irrational, suicidal, or otherwise incapable or unwilling to change course. In such a setting the other driver will swerve and you will obtain the best possible outcome.

An alternate interpretation of the Chicken Game is a situation in which each person individually has the ability to produce an outcome that will benefit both parties, although providing the benefit involves some cost. Whereas mutual cooperation is the unambiguous goal for the Prisoner's Dilemma Game and the Assurance Game, that is not necessarily the case for the Chicken Game. If one person can provide a joint benefit, then it may make no sense for the second person to duplicate the effort. The problem comes when each person attempts

to "stare the other down," each refusing to budge and hoping the other will give in and cooperate. The key problem then is avoiding a stalemate that results in the worst possible outcome.[3]

Note that unlike the Prisoner's Dilemma Game, neither the Assurance Game nor the Chicken Game has a dominating strategy. In the latter two games the partner's choice is crucial in determining one's best outcome—one wants to match the partner's choice in the Assurance Game and to make the opposite choice in the Chicken Game.

Multiple-Person Dilemmas

The first cut in categorizing social dilemmas was distinguishing between two-person and *N*-person dilemmas. Within *N*-person dilemmas we make the next cut, distinguishing between two broad types of multiple-person dilemmas in terms of how the costs and benefits are arranged for each individual (Cross & Guyer 1980). In the first type, known as a social fence, the individual is faced with an immediate cost that generates a benefit that is shared by all. The individual has an incentive to avoid the cost, but if all do so each is worse off than if they had managed to "scale the fence." In the second type, termed a social trap, the individual is tempted with an immediate benefit that produces a cost shared by all. If all succumb to the temptation, the outcome is a collective disaster.[4] Within each of these broad categories lies a richly developed metaphor that has driven research in the area: (*a*) the provision of public goods (a social fence) and (*b*) the tragedy of the commons (a social trap).[5] The potentially noxious outcomes of both types of social dilemmas stem from what economists refer to as externalities, which are present "whenever the behavior of a person affects the situation of other persons without the explicit agreement of that person or persons" (Buchanan 1971, p. 7). Broadly speaking, externalities are uncompensated interdependencies (Cornes & Sandler 1996).

PUBLIC GOODS DILEMMAS A public good is a resource from which all may benefit, regardless of whether they have helped provide the good—I can enjoy public television whether or not I contribute any money, and I can enjoy the parks in my city even if I do not pay municipal taxes. This is to say that public

[3]Poundstone (1992, p. 203) reports an anecdote by Anatol Rapoport that "the Fuegian language of the natives of Tierra del Fuego contains the word *mamihlapinatapai*, meaning, 'looking at each other hoping that either will offer to do something that both parties desire but are unwilling to do.'"

[4]Note that this general formulation of social dilemmas rests on learning models rather than on any assumption of more complex strategic decision-making, as is common in the economic models to be discussed.

[5]Other important dimensions along which social dilemmas can be categorized include the temporal lag between the original action and the eventual disaster (e.g. Messick & Brewer 1983) and whether the actors in a social dilemma each face an identical incentive structure or not (e.g. Marwell & Oliver 1993).

goods are *non-excludable* and as a result there is the temptation to enjoy the good without contributing to its creation or maintenance. Those who do so are termed free-riders, and while it is individually rational to free-ride, if all do so the public good is not provided and all are worse off. This decision is based on greed, i.e., the simple desire to obtain the best possible outcome for oneself. There is also a second reason that can lead to defection—a person may be willing to cooperate but fear that not enough others will do so to actually provide a public good. Rather than greed, the concern here is the fear of being a sucker, i.e., throwing away one's efforts on a lost cause.

Public goods are also distinguished by the fact that they are *nonrival* (Cornes & Sandler 1996) in that one person's use of the good does not diminish its availability to another person—my enjoyment of public television does not make less of it available to anyone else. A pure public good is completely nonexcludable and nonrival, but many public goods exhibit these two qualities only to a varying degree. The basic problem was described at least as early as 1739 by Hume, articulated by Samuelson in 1954, and made famous by Olson in 1965 with the publication of *The Logic of Collective Action*.

A key characteristic of public goods dilemmas is the relationship between the level of resources contributed toward the production of a public good and the level of the public good that is provided. This relationship is known as the production function (Marwell & Oliver 1993, Heckathorn 1996). Production functions can take on any number of forms, but the four basic production functions shown in Figure 2 can be used to model many of the most important dynamics in public goods dilemmas.[6]

With a decelerating production function (Figure 2*a*), initial contributions have the greatest effect, with additional contributions generating increasingly diminishing returns. With a linear production function (Figure 2*b*), each unit of resource contributed produces the same return. An accelerating production function (Figure 2*c*) produces few returns for the initial contributions but brings increasing returns as the contributions increase. Finally, discontinuities in the production function, such as the step-level function in Figure 2*d*, create thresholds (also known as provision points). In these cases little or no amount of the public good is produced until a certain level is reached, at which point a small increase in the level of contributions returns a large and discontinuous amount of the public good. In the analysis of public goods, one of the most important distinctions is whether there are threshold points in the production function (Ledyard 1995).

A common misunderstanding is the assumption that all *N*-person dilemmas have the structure of an *N*-person Prisoner's Dilemma Game. That is, that there

[6]Some production functions are combinations of these basic forms. A third-order S-shaped curve, for example, is composed of accelerating, linear, and decelerating components.

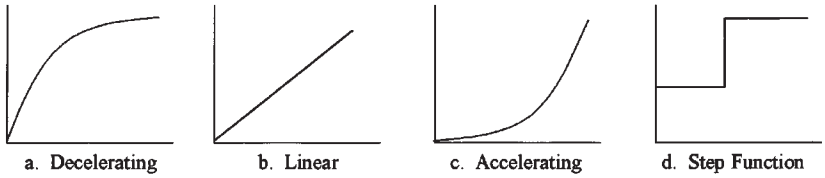

Figure 2 Four production functions (*a*) Decelerating, (*b*) Linear, (*c*) Accelerating, (*d*) Step Function

is a dominating strategy that leads to a deficient equilibrium in which no one cooperates.[7] In fact, various production functions can yield *N*-person versions of the Assurance Game and Chicken Game as well as the Prisoner's Dilemma Game.[8] Note that if a situation has the structure of an Assurance Game, there is no temptation to free-ride—the only concern is whether one will be a "sucker." In a Chicken Game, however, the incentive to free-ride can be even more severe than in the Prisoner's Dilemma Game (Yamagishi 1995).

COMMONS DILEMMAS The second mythic story commonly used in describing *N*-person dilemmas is the tragedy of the commons. Early statements of the basic problem can be found in Aristotle (*Politics*, Book II, Chapter 3).[9] The dilemma in its modern form was carefully described by Lloyd in 1832 and made famous by Hardin in 1968 when he published his article in *Science* on the topic. Hardin described a group of herders having open access to a common parcel of land on which they could let their cows graze. It is in each herder's interest to put as many cows as possible onto the land, even if the commons is damaged as a result. The herder receives all the benefits from the additional cows, and the damage to the commons is shared by the entire group. Yet if all herders make this individually reasonable decision, the commons is destroyed and all will suffer. When timber is harvested faster than it can grow or when fish are caught in greater numbers than their reproductive capacity, we face tragedies of the commons. Here again the problem is the non-excludability of a joint resource, but unlike public goods, a key feature of commons dilemmas is

[7]This belief has led to some misguided critiques of the game theoretic models underlying work on social dilemmas. For a rejoinder to these flawed criticisms as well as an honest discussion of the limits of game theoretic models, see Lohmann (1995).

[8]Heckathorn (1996) has written a very useful general analysis of different types of games and how they are the result of various production functions in combination with the relative value of the public good (note that this analysis concerns dyadic games). See also Schelling (1978) for an influential discussion and method of modeling different social dilemmas.

[9]And later in traditional rhymes: "They hang the man and flog the woman, That steal the goose from off the common, But let the greater villain loose, That steals the common from the goose" (quoted in Fairlie et al 1994).

the *subtractability* of the benefits (the opposite of being nonrival): The tree I cut, the fish I catch, and the water I use are not available for others.

For commons dilemmas, the issue is not the production function but the carrying capacity of the commons, which is a function of its replenishment rate. Different resource pools will be renewed at different rates—the reproduction rate of a species of fish, the yearly rainfall that adds to groundwater reserves, the rate at which pollutants dissipate in the air; this will determine the rate at which the subtractable joint resource can be appropriated without exhausting the commons.

In sum, public goods dilemmas concern the production of, and commons dilemmas involve the use of, a joint good from which it is difficult to exclude others. There are many reasons why excluding others might be costly (Ostrom et al 1994): the physical nature of the resource (it is difficult to fence in ocean fish or exclude tax scofflaws from the benefits of secure national borders); the available technology (enclosing huge range lands used to be prohibitively expensive—until the introduction of barbed wire); or existing laws and traditional norms (which might prohibit anyone being excluded from a commons or public good). These two basic forms of *N*-person dilemmas are further distinguished by the fact that public goods are significantly nonrival, whereas commons dilemmas involve a subtractable resource.[10]

The move from two-person to *N*-person dilemmas involves a number of profound shifts that affect the dynamics of the game. Dawes (1980) described three important ways in which the two-person Prisoner's Dilemma Game differed from the *N*-person version: First, in an *N*-person dilemma, one's actions are not necessarily revealed to others—anonymity becomes possible and an individual can free-ride without others noticing her or his actions. In the two-person case, each player "knows with certainty how the other has behaved" (Dawes 1980, p. 51). Second, the cost one imposes on others from defecting is focused completely on one's partner in the case of a two-person dilemma, whereas it is diffused throughout the group in an *N*-person dilemma. Finally, in a two-person dilemma, each person has significant control over one's partner's outcomes and so can shape the partner's behavior in important ways. In contrast, in an *N*-person dilemma, one may have little or no direct control over the outcomes others receive.

This list of distinguishing features is a useful starting point, but as we think about the whole range of social dilemmas, it is important to keep in mind two significant qualifications. First, anonymity, the diffusion of the harm of defection, and the inability to significantly impact others' outcomes are possible

[10]A nonrival good from which it is easy to exclude people is known as a club good (see Cornes & Sandler 1996 for an extended discussion of club goods). Goods that exhibit both subtractability and excludability are termed private goods.

though not inevitable features of *N*-person dilemmas. Kollock & Smith (1996), for example, discuss large-scale dilemmas that do not exhibit all of these features. Second, some of the features identified as characteristic of *N*-person dilemmas can be found in two-person dilemmas. For example, in the Prisoner's Dilemma Game as it is traditionally played in the experimental lab, each player knows with certainty how the partner has acted, but this is not the case for many of the two-person dilemmas in our empirical lives. I may promise my partner that I will run an errand for her while I am out, decide to take care only of my needs, and then claim upon returning that heavy traffic prevented me from fulfilling her request.

SOLVING SOCIAL DILEMMAS

In this section possible solutions to social dilemmas are considered. These solutions are divided into three broad categories based on whether the solutions assume egoistic actors and whether the structure of the situation ("the rules of the game") can be changed. Motivational solutions assume actors are not completely egoistic and so give some weight to the outcomes of their partners. Strategic solutions assume egoistic actors, and neither of these categories of solutions involves changing the fundamental structure of the situation. Solutions that do involve changing the rules of the game are considered in the third section on structural solutions. I have grouped together both structural solutions that assume egoistic actors and structural solutions that assume some weight is given to what others receive.

Motivational Solutions

Do individuals take their partners' outcomes into account when making a decision? Many of the models in the literature on social dilemmas assume actors who are focused only on their own outcomes, but it seems clear that many of us do give some weight to what our partner receives. As Dawes (1980, p. 176) commented, "Few of us would accept $500 with nothing for our friend in lieu of $495 for each of us."

SOCIAL VALUE ORIENTATIONS Research on social value orientations (e.g. Kuhlman & Marshello 1975, McClintock & Liebrand 1988) has sought to determine if there are stable individual differences in "preferences for particular distributions of outcomes to oneself and others" (van Lange et al 1992, p. 17). Many different social value orientations are theoretically possible, but most work has concentrated on various linear combinations of individuals' concern for the outcomes for themselves and their partners. One possibility is that an individual might behave so as to maximize joint outcomes (this is described as a cooperative orientation in this literature). An individual might also desire to

maximize the relative difference between self and partner (a competitive orientation). Other orientations include maximizing the partner's outcome without regard for own outcome (altruism) or maximizing own outcome without any concern for the partner's outcome (individualism).[11] Research in numerous countries has found that most individuals can be classified as either cooperators, competitors, or individualists.

This work has shown that individuals with different social value orientations behave differently when faced with the same objective game (McClintock & Liebrand 1988, Liebrand et al 1986, Kramer et al 1986, Liebrand 1984). The researchers in this area have also shown that these orientations are relatively stable over time (Kuhlman et al 1986).[12] These studies make use of work by Kelley & Thibaut (1978) and Kelley (1979) on matrix transformations. Kelly & Thibaut argue that individuals often subjectively transform a given game and play it as if it were another game. There is, after all, no guarantee that subjects play an experimental game as intended by the researcher—for any of a variety of different reasons people might value particular outcomes more or less than the immediate objective payoff they receive.

Work on social value orientations has concentrated on assessing transformations that are the results of personality traits. Understanding that some individuals routinely give different weights to their own and partner's outcomes is an important piece of information in explaining the observed rates of cooperation in social dilemma situations. However, this knowledge is not very useful as a solution to dilemmas—this research does not tell us how to increase the level of cooperation.

One possibility would be to study how social value orientations are formed. A group of researchers (McClintock & Keil 1983, Toda et al 1978, McClintock 1974) have studied the development of cooperative and competitive orientations in children in several countries. Among other results, they have found that competition seems to be learned significantly earlier than cooperation and that overall levels of competitiveness can vary from country to country. But until the actual mechanisms are identified by which social orientations are learned, these insights still do not provide the basis for intervention in a social dilemma. Along these lines, a few studies (Frank et al 1993, Marwell & Ames 1981) have examined the effects of education later in life, examining whether

[11]Other more exotic orientations are also possible, such as martyrdom and sadism. Note that not all orientations can be modeled within this framework—another important social value orientation is minimizing the difference between own and partner's outcomes; this is related to issues such as equality and justice.

[12]Research has also found that stable individual differences exist in how trusting a person is and that this affects cooperation levels in social dilemma situations (e.g. Yamagishi 1986). Another individual difference that is noteworthy is the finding that people who cooperate are more likely to expect others to cooperate as well (Orbell & Dawes 1993).

students in different majors are more or less likely to cooperate. The amusing outcome is that some evidence suggests that people who study economics are more likely to free-ride. However, there are questions (mostly by economists) about how this result should be interpreted and how robust the finding is.

COMMUNICATION Another approach is to ask what features in the situation or environment (rather than in an individual's personality) affect the weight individuals give others' outcomes. One of the most robust findings in the literature is the positive effects of communication on rates of cooperation. Across a wide variety of studies, when individuals are given the chance to talk with each other, cooperation increases significantly (e.g. Orbell et al 1990, Orbell et al 1988, Liebrand 1984, Edney & Harper 1978b, Dawes et al 1977, Jerdee & Rosen 1974). While the effect is readily observed, explaining it has been more of a challenge. Messick & Brewer (1983) suggested four reasons (which touch on both motivational and strategic factors) why communication might increase cooperation. First, individuals may be able to gather information about the choices others are likely to make. This information, however, can have ambiguous effects. If I believe that most other people will cooperate in an *N*-person dilemma, does that give me a reason to cooperate or a greater temptation to defect? In part the decision will depend on the structure of the dilemma (in an Assurance Game, I will be happy to cooperate if others do) and on one's social value orientation. Second, communication gives group members the chance to make explicit commitments and promises about what they will do. However, research has been inconclusive about whether such commitments have an effect on cooperation rates (Orbell et al 1990, Dawes et al 1977). Third, communication offers an opportunity for moral suasion, i.e., appeals to what is the "right" or "proper" thing to do. The effects of moralizing have been the subject of very little research, although there are at least some indications that it can have a salutary effect on cooperation (Orbell & Dawes 1981). Finally, communication may create or reinforce a sense of group identity. This last point seems especially important, and the opinion of one of the key researchers in this area (Dawes 1991) is that the key effects of communication come from eliciting group identity.

GROUP IDENTITY The impact of group identity is manifold and profound, having effects across all three categories of solutions: motivational, strategic, and structural. Indeed, group identity can have such a powerful effect that it can influence rates of cooperation even in the absence of communication. Kramer & Brewer (1984, 1986; Brewer & Kramer 1986) have demonstrated that subjects are more willing to exhibit personal restraint in a commons dilemma simply as a result of being identified as members of a common group. Intergroup competition can have even more striking effects. In a study involving

naturally occurring groups, Kollock (1998a,b) uncovered evidence of consistent transformations of a social dilemma situation, such that a Prisoner's Dilemma was treated as an Assurance Game when the partner was an in-group member and as a Prisoner's Dilemma Game when the partner was an out-group member. Experimental work by Bornstein and Rapoport (Bornstein et al 1990; Rapoport et al 1989; Rapoport & Bornstein 1987, 1989; Bornstein & Rapoport 1982) and the classic field experiments of Sherif et al (1961) have shown the powerful effect of intergroup competition in promoting cooperation within groups. However, this solution can be double-edged. Encouraging or creating group competition can serve the needs of group members (and leaders and politicians), but the social costs of the conflicts that result between groups can be severe.

Why are individuals more willing to cooperate if they feel part of a group? One possibility is that a collective social identity increases the altruism of the members. This is certainly a possibility, but something more strategic may be happening as well. Indeed, it can be difficult to distinguish apparent altruism from subtle long-term strategic considerations. This issue is taken up in the next section.[13]

Strategic Solutions

Strategic solutions assume egoistic actors and no changes to the structure of the game. These approaches rely on the ability of actors to shape the outcomes and hence behavior of other actors. For this reason, many of these strategic solutions are limited to repeated two-person dilemmas.

RECIPROCITY Far and away the most influential study on strategic solutions to social dilemmas is Axelrod's *The Evolution of Cooperation* (1984), in which he reports the results of a series of computer tournaments investigating the two-person Prisoner's Dilemma Game. While research on the Prisoner's Dilemma Game had gone on for many years prior to Axelrod's book, it was distinguished by its intriguing method, a provocative set of conclusions and recommendations, and arresting examples taken from such diverse areas as biology and the history of trench warfare. The study centered on two tournaments in which prominent game theorists (and in the second tournament, computer hobbyists) were invited to submit strategies for playing the Prisoner's Dilemma in a round-robin contest.

Axelrod identified three requirements in this environment for there to be even the possibility of the emergence of cooperation. First, it was essential that

[13]Note also that Kuhlman et al (1986) have argued that social value orientations themselves might be explained as different strategic responses to the problem of how to maximize one's own outcomes.

individuals be involved in an ongoing relationship. If individuals met only once, or equivalently, if this was the last time they would meet, the dominating strategy to defect in the Prisoner's Dilemma Game would make the pursuit of cooperation hopeless.[14] If the partners would meet again in the future, cooperation at least has a chance. The second condition is that individuals must be able to identify each other. The third condition is that individuals must have information about how the other person has behaved in the past. If identity is unknown or unstable and if there is no recollection or record of past interactions, individuals will be motivated to behave selfishly because they will not be accountable for their actions.

It was surprising to many at the time that the winner of Axelrod's two tournaments was the simplest strategy that had been submitted. This strategy, named Tit-for-Tat, cooperates on the first interaction and thereafter simply does whatever its partner did on the previous round. It has proven to be an effective strategy in many different environments, and it has the effect, in essence, of transforming a repeated Prisoner's Dilemma Game into a repeated Assurance Game (Yamagishi 1995; cf. Rapoport 1967). Playing against an individual using Tit-for-Tat means that the only long-term possibilities are mutual cooperation and mutual defection—there is no hope of exploiting this strategy in any kind of sustained way. In this sense it can provide a route to sustained mutual cooperation in a two-person Prisoner's Dilemma Game.

After studying the most successful strategies in the tournaments, Axelrod (1984, p. 110) distilled four pieces of advice that he would offer an individual playing an iterated Prisoner's Dilemma Game: (*a*) Don't be envious; (*b*) don't be the first to defect; (*c*) reciprocate both cooperation and defection; and (*d*) don't be too clever. The key point in his fourth piece of advice was that it was important for one's partner to clearly understand what strategy one was using. His first piece of advice is essentially an admonition against playing the Prisoner's Dilemma Game as if it were a zero-sum game, that is, a game in which one's interests were completely opposed to one's partner's (e.g. chess, competitive sports, mortal combat). In a zero-sum game, using one's partner as a standard of comparison is useful, as anything that works against one's partner necessarily helps oneself. However, trying to beat one's partner or being envious of their success[15] can lead to trouble in a mixed-motive situation such as the Prisoner's Dilemma. Trying to beat your partner can be self-defeating if it results in mutual defection.

[14]Axelrod's analysis is based on the logic of game theory. Empirically, one does sometimes observe cooperation in a one-shot Prisoner's Dilemma Game (e.g. Orbell & Dawes 1993, Hayashi et al 1997).

[15]This can be thought of as negatively weighting one's partner's outcomes, as is the case with the competitive social orientation discussed above.

A very important lesson from the tournaments was that Tit-for-Tat won not by beating its partners (indeed, it can only tie or do slightly worse than its partners), but by doing well on average, encouraging mutual cooperation with many of its partners. This seems to be one of the hardest lessons for individuals to learn, perhaps because of the dominance of the competitive game as a model in many cultures—if the only metaphor you have is the zero-sum game, you tend to treat everything as if it were a war. The book spawned a cottage industry of hundreds of studies that supported, extended, or critiqued the original work. Two very useful reviews of research that make use of Axelrod's studies are Axelrod & Dion (1988) and Axelrod & D'Ambrosio (1994).

The success of Tit-for-Tat led some commentators to suggest that this strategy be used as the basis of everything from childhood education to international relations. Here again we see the dangers of taking a useful metaphor too literally, assuming it accurately modeled any situation that even vaguely resembled a Prisoner's Dilemma. As Axelrod himself appreciated, the results of his tournaments depended on both the particular sample of strategies that were submitted and the assumptions underlying his study. One of the most important scope conditions of Axelrod's simulations was the assumption of perfect information. In a world in which mistakes, misperceptions, and accidents can occur, Tit-for-Tat can turn out to be an unsuccessful strategy because it retaliates immediately (Kollock 1993). Strategies that are more generous or forgiving than Tit-for-Tat can have important advantages in such settings because they avoid the danger of cycles of recrimination that can occur with Tit-for-Tat.

CHOICE OF PARTNERS Another key assumption of Axelrod's model was the network structure of the interacting strategies. In a sense it represented a very unusual social structure in which each actor was forced to interact each round and to interact with every other possible partner (as this was a round-robin tournament). Not playing the game or choosing only some partners with whom to interact were not options. One of the most important recent developments has been studies that permit players to exit a current relationship and/or choose alternative partners. Computer simulations by Schuessler (1989), Vanberg & Congleton (1992), and Hayashi and associates (Hayashi 1993, Yamagishi et al 1994) all found that a very successful strategy in these situations was to cooperate on the first interaction and continue cooperating until the first defection from one's partner, at which point the strategy exited the relationship. Hayashi (1993) also discovered that a version of this strategy (called Out-for-Tat), which incorporated some degree of forgiveness (i.e. a willingness to give a partner who had defected before a second chance), was even more successful. The conclusion of this work is that the strategy used in selecting one's partner can be more important than the strategy that is used in actually playing the

Prisoner's Dilemma Game. Experimental work by the same researchers (Yamagishi et al 1994) suggests that subjects follow something like an Out-for-Tat strategy in which the response to defection is not defection, but desertion.

GRIM TRIGGERS All of these studies involved two-person social dilemmas. Strategic solutions for *N*-person dilemmas are much more of a challenge because one's own actions may have little or no influence on what others do. One possibility that has been explored is the adoption of a "grim trigger" strategy, in which each individual agrees to cooperate only on the condition that all others in the group cooperate. In theory, if all adopt this strategy then each person's decision is decisive and free-riding is impossible. However, experimental work by Watabe (1992; Watabe & Yamagishi 1994) found subjects were leery of adopting such a risky strategy, and field studies by Ostrom and her colleagues (Ostrom et al 1994) uncovered no instances of groups actually using a trigger strategy in their community.

SOCIAL LEARNING A different approach to solving *N*-person dilemmas has been investigated by Macy (e.g. 1993, 1991). His model of decision-making does not assume that actors calculate marginal rates of return or work out dominating strategies. Basing his work on the principles of social learning theory, he assumes reward-seeking, penalty-aversive actors and asks under what conditions such cognitively modest actors might escape social dilemmas. In a series of computer simulations, he isolates a number of factors that can promote cooperation, including the presence of thresholds and the tendency for actors to imitate those around them.

GROUP RECIPROCITY Finally, we return to the issue of group identity and its effects. Making group identity salient has been shown to increase cooperation. While work in social identity theory (Tajfel 1981) argues that simply categorizing individuals into a common group is enough to increase their altruism toward the group, research by Karp et al (1993; see also Jin et al 1996) contests this conclusion. The effects of group identity stem, they argue, not from an altruism born of categorization, but from a belief in the interdependencies of group members and expectations of reciprocity among the members. In a series of studies they carefully removed any possibility or connotation of interdependency and found that simple categorization was not enough to create in-group favoritism.

It is the belief in future reciprocal exchanges between members, they argue, that moderates the temptation to defect and encourages cooperation. The expectation of in-group reciprocity seems to serve as a very deep heuristic that shapes our strategic decisions (Jin & Yamagishi 1997, Brewer 1981). The ex-

pectation of reciprocity appears to be so great that it sometimes manifests itself even in situations in which reciprocity is not logically possible (Watabe et al 1996, Hayashi et al 1997, Karp et al 1993). Further, this heuristic means that many Prisoner's Dilemma situations will be transformed into Assurance Games. Evidence for this transformation can be found in Watabe et al (1996) and Hayashi et al (1997).

Structural Solutions

In this section I relax the assumption that the rules of the game cannot be changed. Here I examine structural changes to social dilemmas that either modify the dilemma or eliminate it entirely. An important issue discussed below is how these structural changes are provided.

ITERATION AND IDENTIFIABILITY One approach to structural solutions is to create or reinforce those features of the environment that are prerequisites for strategic solutions. Returning to Axelrod (1984), this approach suggests three changes: (*a*) Make interaction more durable or frequent; (*b*) increase identifiability; and (*c*) increase information about individuals' actions. If individuals will not interact in the future, if identity is unknown or unstable, and if there is no recollection or record of past interactions, individuals will be motivated to behave selfishly because they will not be accountable for their actions. Knowing the identity and history of a person allows one to respond in an appropriate manner. If information about individuals and their actions is shared among the group, this also encourages the development of reputations, which can be a vital source of social information and control. These features will be important not just for facilitating strategic solutions but also as prerequisites for some of the other structural solutions discussed below, notably the use of monitoring and sanctioning systems. Along these lines, several studies have found that anonymity (the absence of identifiability) lowers rates of cooperation (Fox & Guyer 1978, Jerdee & Rosen 1974, Kahan 1973).[16]

However, it is important to note that ongoing interaction may not always have a salutary effect on social dilemmas. Axelrod was concerned with two-person Prisoner's Dilemmas, and in *N*-person Prisoner's Dilemmas (with no thresholds), there is a stronger temptation to move toward the equilibrium of zero cooperation, as defection has a smaller effect and one may not be able to impact others' outcomes and so encourage cooperation. Ledyard (1995) discusses this general issue and points to a number of studies that have found significant declines in cooperation over time in *N*-person dilemmas with no

[16]Note that anonymity in and of itself may not always have an effect (Kerr 1997). For example, if individuals do not understand the situation or do not care about the sanctions others impose, whether one is identifiable or not may not matter.

thresholds (e.g. Andreoni 1988, Banks et al 1988, Isaac et al 1985, Isaac et al 1984, Kim & Walker 1984).

PAYOFF STRUCTURE As one would expect, numerous studies have demonstrated that the greater the personal return from cooperation and the lower the return from defecting, the higher the levels of cooperation (Isaac & Walker 1988, Issac et al 1984, Komorita et al 1980, Bonacich et al 1976, Kelley & Grzelak 1972). Perhaps more surprising is the finding that cooperation rates increase significantly as the benefits to *others* from one's cooperation increase (Bonacich et al 1976, Kelley & Grzelak 1972, Komorita et al 1980). This argues that many people are positively weighting the outcomes of others.

The nature of the public good and how it is distributed can also have an effect. Alfano & Marwell (1980) found that cooperation levels were much greater when group members were asked to contribute to a public good that was nondivisible. That is, rather than each person getting an individual return, the group would receive a lump sum that had to be spent on a group activity. The very fact that the public good was indivisible may have helped reinforce a sense of group identity and interdependence among the subjects.

EFFICACY Many researchers have argued that one of the key reasons people do not cooperate in an *N*-person dilemma is the fact that a single person's actions may have no discernable effect on the situation. No one will be fired and no program will go off the air if I do not send in a $30 contribution to public television, and even if I do conserve water in a drought, it will have no measurable impact on the overall situation.

If a dilemma is structured in such a way that individuals can have a noticeable effect on the outcome—that is, they can make an efficacious contribution—cooperation rates can be increased. One way in which this can occur is if a public good has a step-level production function. If an individual believes the group is close to the threshold, then adding one's own contribution can be enough to put the group "over the top" and provide the good. One study (van de Kragt et al 1983) found that groups who were attempting to provide a public good with a threshold designated a subgroup of contributors (via lottery or volunteering) who would be just enough to provide the good. In this situation each person within the minimally contributing set knew that the provision of the public good was critically dependent on each of their actions. Free-riding was impossible, and each knew that their actions were necessary for the success of the group.[17]

[17]Other work has shown that increasing the level of the threshold can increase the amount of cooperation, although this also decreases the probability that the threshold will be met (Isaac et al 1988, Suleiman & Rapoport 1992).

Another experimental study by Bornstein et al (1990) demonstrates the joint effects of a step-level production function and group identity. The key innovation in these studies (see also Rapoport et al 1989; Rapoport & Bornstein 1987, 1989; Bornstein & Rapoport 1982) is that two groups are set up in competition against each other, with a prize going to the group that demonstrates the higher level of cooperation; the prize is then distributed equally to the winning group's members. This changes the structure of what was originally a Prisoner's Dilemma into a step-level public goods problem in which defection is no longer a dominating strategy. The creation of a step-level function (which is to say, a threshold point) shifts the structure of the game into an *N*-person version of the Chicken Game.

Work by Kerr (e.g. 1992) has also shown that cooperation in a public goods dilemma is more likely the larger the impact of a person's contribution. A similar strategy is used by public television and charities when they create "matching grants" in which someone agrees to double the contributions that others make.

The perception of efficacy can be enough to affect cooperation. Kerr (1989) and Rapoport et al (1989) found a significant relationship between perceived efficacy and contributions to a public good. Survey and field research have also found that most individuals involved in collective action believed that their actions had a significant effect on the provision of the public good, even if the size of the group was very large (Klandermans 1986, Mueller & Opp 1986, Moe 1980).

The creation of efficacy, real or perceived, can be an art. Consider the difficulties of a charity trying to raise money to feed poor children. A potential contributor may fear that her or his contribution will be wasted or wonder what good one person can do for an organization that raises millions of dollars. The response of at least one charity has been to assign each contributor a specific child. The contributor receives a photo and personal information about the child they are sponsoring and even an occasional letter from the child or one of the parents. The sense of personal responsibility it creates ("what happens if I stop contributing?") profoundly changes the decision of whether to cooperate.

GROUP SIZE Numerous studies have found that cooperation declines as group size increases (e.g. Komorita & Lapworth 1982, Fox & Guyer 1977, Bonacich et al 1976, Hamburger et al 1975). The possible reasons for this effect are many. Returning to Dawes's (1980) points about differences between two-person and *N*-person games, increasing group size may spread the harm caused by defection, make it harder to shape others' behavior, and make it easier to defect anonymously. The costs of organizing can also increase as group size grows (Olson 1965)—groups can find it harder to communicate and coordinate their actions. The efficacy and visibility of one's actions can also be di-

luted, and monitoring and sanctioning the behavior of others (see below) can become more of a challenge. This general effect has led some commentators to argue in favor of anarchistic social systems in which communities are organized as networks of small groups (Fox 1985).

However, none of these effects are inevitable as groups grow in size (Udehn 1993, Kollock & Smith 1996). Interestingly, some experimental work has found that the decrease in cooperation as group size increases tapers off quickly (Fox & Guyer 1977, Liebrand 1984), and other work has actually shown an increase in cooperation with larger groups (Yamagishi & Cook 1993, Isaac et al 1990). Part of the problem in reaching any precise conclusion about the effects of group size is that so many elements can vary as group size increases. It is in the end impossible to control for all possible parameters in order to study a "pure" group size effect (Orbell & Dawes 1981, Ledyard 1995). Researchers must decide which parameters are most important and carefully control them. Another difficulty with researching this effect thoroughly is that one must examine groups of varying sizes, including large groups, and running experiments with large groups creates extraordinary logistical difficulties and costs.

One explanation for why larger groups may be more likely to solve social dilemmas comes from Marwell & Oliver (1993). They argue that if a public good is highly nonrival, a large group is more likely to contain a critical mass of individuals whose interests are served by providing the good. One feature of a group that encourages the formation of a critical mass is the heterogeneity of the group in terms of the diversity of group members' interests and resources. The importance of group heterogeneity in solving social dilemmas is also explored by Glance & Huberman (1994).

BOUNDARIES This set of structural solutions deals with a core characteristic of social dilemmas—the nonexcludability of a joint good. Each of these solutions attempts to draw some kind of boundary around the collective good.

One of the first solutions proposed for commons dilemmas is the establishment of an external authority to regulate who had access to the commons or how people were to withdraw resources from the commons. This is, in a broad sense, Hobbes' classic solution of Leviathan: People give up some part of their personal freedom to an authority in return for some measure of social order. This is also the solution Hardin proposed in his famous article (1968) when he concluded that "freedom in a commons brings ruin to all." Hardin fully acknowledged that the outcome might be grossly unfair to some people, but given the global tragedy he felt was inevitable, he declared that "injustice is preferable to total ruin" (1968). This echoed Lloyd's (1832) grim conclusion that "To a plank in the sea, which cannot support all, all have not an equal right." A direct example of this strategy can be seen in the establishment of fish

and game authorities that set strict limits on what can be caught and the length of the season. A similar approach can be taken in public goods dilemmas, where an external authority compels individuals to contribute money (as when a government collects taxes to provide public services) or labor, such as in the case of military conscription.

The willingness to hand over personal choice to a leader has been shown in some experimental studies. Messick et al (1983) and Samuelson & Messick (1986a) found that a group that was overusing a commons was willing to change the structure of the situation by electing a leader who would manage the harvesting of the commons for the group. Interestingly, subjects did not usually vote for themselves, instead electing a person who counteracted the group's performance to that point in time: someone who harvested few resources from the commons if the group had overharvested, and someone who harvested substantial resources if the group had underharvested to that point (Messick et al 1983, Samuelson et al 1984). However, Samuelson & Messick (1986b) and Rutte & Wilke (1985) found that subjects preferred not to create a leader if other structural changes were possible.

Severe problems can arise in establishing such an authority, as Crowe (1969) commented soon after the publication of Hardin's article. According to Crowe, Hardin assumes (*a*) that the global community can come to an agreement about what to value and how to rank those values, (*b*) that authorities will have sufficient coercive force to compel people to obey, and (*c*) that authorities can be trusted to remain free of corruption and to resist the influence of special interest groups. Crowe vigorously questions each of these assumptions and argues that even if Hardin's basic presumptions are correct, his solution is unworkable on a broad scale.

Another commonly suggested solution to the tragedy of the commons is to privatize the commons, that is, to break the commons up into private parcels on the assumption that individuals will take better care of their own property than common property. Two experimental studies (Cass & Edney 1978, Messick & McClelland 1983) indicate that individuals did better at managing their own "private commons" than they did harvesting as a group. However, there are a number of difficulties with this solution. First, not all goods can be privatized—it may be easy to divide up an actual meadow,[18] but how does one divide up schools of fish in the ocean, clean air, or many public goods such as national defense? Second, even if it is possible to divide up the common good, doing so raises grave questions about social justice: Who gets the newly privatized commons, and how are the parcels allocated? To the highest bidder? In a lottery? Third, while it may be reasonable to expect people to take good care

[18]Assuming the meadow is homogenous; see Ostrom 1990, p. 13.

of their own property, empirically there is no question that individuals routinely destroy their own property. Some have argued that there are "tragedies of enclosure" (Bromely 1991) just as there are tragedies of the commons. Finally, private property rights require a great deal of institutional support so that these rights can be enforced.

Some of the other assumptions in Hardin's original analysis have also been criticized. Notably, he assumes that commons are always open-access, that is, that there are no restrictions as to who may use the commons. However, this assumption is neither necessary nor historically accurate (Fairlie et al 1994, McCay & Acheson 1987). In fact, commons are often surrounded by local rules of access and enforcement mechanisms. One of the key findings of field research done on how communities manage common property is that groups often do find ways to regulate their own actions, and some of these arrangements have proven to be remarkably robust, lasting across several generations (McCay & Acheson 1987, Ostrom 1990, 1992, Ostrom et al 1994).

Thus, Ostrom (1990) proposes a third route away from the tragedy of the commons: the local regulation of access to and use of common property by those who actually use and have local knowledge of the resource. Ostrom isolated a number of design characteristics that were shared by communities that had a long history of successfully managing common resources. The first characteristic she discusses deals explicitly with the issue of excludability: Successful communities are marked by clearly defined boundaries—"Individuals or households who have rights to withdraw resource units from the [commons] must be clearly defined, as must the boundaries of the [commons] itself" (1990, p. 91).[19]

This is not to say that local communities inevitably solve their social dilemmas—there is no shortage of true tragedies as well as victories—but it does make the essential point that it is inappropriate to conclude that the only way out of a commons dilemma is through the use of some form of Leviathan or privatization. This has been the conclusion of a number of commentators who took Hardin's parable too literally. It is also the case that misguided intervention by an outside authority can take a bad situation and make it much worse (McCay & Acheson 1987).

SANCTIONS As Dawes (1980) pointed out, one of the great challenges of *N*-person dilemmas is that it is often not possible to directly affect others' outcomes and so shape their behavior. If the cooperators could be rewarded for their action and defectors punished, even large-scale dilemmas might be solved.

[19]Other design features identified by Ostrom are discussed in subsequent sections. Many of these characteristics are applicable to public goods dilemmas as well.

Indeed, one of Olson's (1965) key conclusions was the necessity of using selective incentives in encouraging cooperation. A selective incentive is a private good given as an inducement to contributing toward a public good. Anyone can watch public television, but only subscribers receive program guides, discount cards, and other rewards for subscribing. We may all have access to a common pool of blood at the blood bank, but at the University of California at Los Angeles, only those who do contribute blood receive a half day off with pay, food and drink, and occasionally even the chance to win more substantial prizes in a lottery. Field research on conservation behavior (Maki et al 1978, Winett et al 1978) has shown that selective incentives in the form of monetary rewards are effective in decreasing the consumption of water and electricity.

If carrots work, so do sticks. Experimental studies have shown that cooperation is more likely if individuals have the ability to punish defectors (Caldwell 1976, Komorita 1987). Such negative sanctions are the complement of the positive sanctions used in selective incentive systems—the target in this case is the defector rather than the cooperator.

However, implementing sanctioning systems raises two important problems. First, there are often significant costs to providing these systems. In order to reward or punish individuals, one must first be able to monitor their behavior (Hechter 1984). This may be trivially easy if we are working next to each other building an irrigation system, or essentially impossible, as when individuals in a large city decide to leave the water running in the privacy of their home. Even if one is able to keep track of individuals' actions, there are still costs in administering rewards or punishments. The rewards themselves can be costly, and administering negative sanctions can require the support of large and expensive institutions (e.g. a police force, an internal revenue service). Thus, it will sometimes be the case that the costs to monitor and sanction individuals will be greater than the benefits that come in terms of higher cooperation.

While monitoring and sanctioning costs can be very great, some situations exist in which the costs can be made very small through the right institutional arrangements (Ostrom 1990). In general, these costs can also be quite modest in small groups.[20] Ostrom (1990) in particular documents the many ways face-to-face communities create local monitoring and sanctioning systems. The presence of a monitoring and sanctioning system run by the community members themselves (as opposed to an external authority) was one of the design features Ostrom found in each of the successful communities she studied. Another common element Ostrom identified was that cooperative communities

[20]While this is true in many cases, secret defection is possible in even the smallest group, and there are some situations where monitoring is easily accomplished even in very large groups (e.g. Kollock & Smith 1996).

employed a graduated system of sanctions. While sanctions could be as severe as banishment from the group, the initial sanction for breaking a rule was often very low. Community members realized that even a well-intentioned person might break the rules when facing an unusual situation or extreme hardship. Severely punishing such a person might alienate him or her from the community, causing greater problems. Ostrom also found that even with a well-designed internal monitoring and sanctioning system, some conflict was inevitable. Thus, it was important that community members had access to low-cost conflict resolution mechanisms.

The second key problem in implementing sanctioning systems is that these systems are themselves public goods because one can enjoy the benefits of a sanctioning system without contributing to its provision or maintenance. Whether the sanctions are provided by an external authority or locally, there is the temptation to free ride. The police and judicial system continue to work even if I avoid paying taxes, and if everyone else in my community takes on the task of informally admonishing and criticizing defectors, I can avoid the costs of such actions and still enjoy the benefits they bring. This raises the question of when people will cooperate in providing this second-order public good (i.e. a public good designed to be a solution to an underlying social dilemma).

The most extensive set of experimental studies on the provision and use of sanctioning systems has been by Yamagishi (1992, 1988, 1986). While one might expect that people's decisions when faced with a second-order dilemma mirror their actions when faced with a first-order dilemma, this turns out not to be the case. Yamagishi found that trusting individuals (as measured via a scale administered prior to the experiment) were likely to cooperate in a first order-dilemma, but when they were given the opportunity to contribute toward the provision of a sanctioning system, relatively few did so. In contrast, a group of distrustful individuals exhibited low levels of cooperation in the first-order dilemma but were more willing to cooperate in the creation of a sanctioning system. The existence of a sanctioning system led to cooperation rates in the end that were similar to those of the trusting individuals.

Why people might be willing to cooperate in a second-order dilemma is an area ripe for research. Studies by Heckathorn (1996, 1989, 1988), Axelrod (1986), and Yamagishi & Takahashi (1994) provide some initial investigations based on computer simulations that examine the effect of collective sanctioning, "hypocritical" sanctioning (i.e. sanctions by actors who defect in the first-order dilemma), and the evolution of traits that encourage sanctioning.

CONCLUSIONS

The study of social dilemmas is the study of the tension between individual and collective rationality. It is the study of tragic (deficient) equilibria caused

by externalities, that is, uncompensated interdependencies. Social dilemmas are also a sensitive research domain, in that a great many variables can affect cooperation rates, and small changes in these variables can sometimes have large effects (Ledyard 1995).

Studying Social Dilemmas

One of the great advantages of doing research in social dilemmas is that a well-specified set of models exists that allows one to capture the key dynamics in a simple and tractable way in the laboratory. This is also one of the downfalls of research in the area—it is perhaps too easy to set up a social dilemma and vary any one of an infinite number of variables rather than thinking strategically about which situations and which parameters are most important. The Prisoner's Dilemma Game in particular has served as a kind of readily available thermometer of cooperation that can be stuck into any situation. There is nothing to keep a researcher from examining the effects of eye color or pounds of meat consumed on cooperation rates, and some studies have come close to such esoterica. As Messick & Brewer (1983, p. 40) warned us at the end of their influential review, "There are more experiments that can be done than are worth doing and it is as important as it is tricky to determine which are which."

Current experimental work has introduced a number of important innovations in the design of studies. In particular, a number of researchers have increasingly relaxed the constrained designs of early work. One can now find research in which actors have the option of leaving the interaction, of choosing new partners, and even of choosing the game structures (Kakiuchi & Yamagishi 1997). Other researchers have developed designs that highlight the importance of the group, examining the effects of inter-group competition or the use of exclusion from the group (ostracism) as a sanction (Kerr 1997).

There are some chronic problems, however, in the manner in which experimental research is being conducted. Many experiments in this area have used trivially small incentives. "It makes no sense," said Orbell & Dawes (1981), "to spend large amounts of money for summer salaries, secretaries, computer terminals, and research assistants, and then motivate the subjects with microscopic amounts of money or course credits." The generalizability of our results is limited to the extent we use small incentives, and it is entirely possible that many of the inconclusive or contradictory results that are reported in the literature are due in part to subjects being faced with outcomes that are trivial. It is also the case that with few exceptions, most of these studies have involved very small groups. Our results are thus limited again to the extent that a 10-person group is defined as a "large" group.

Turning briefly to other methodologies, experimental work has been bolstered by studies based on computer simulations that allow one to investigate

models involving very large groups, explore the logical terrain of one's theories, study problems that do not lend themselves to analytical solutions, and develop tenable models that can guide behavioral experiments (for recent collections, see Liebrand & Messick 1996, Axelrod 1997). Also important in recent years has been the emergence of a great many field studies based on the logic of social dilemmas. These include many studies on resource management as well as fascinating accounts on such topics as the Sicilian Mafia (Gambetta 1993) and a study of trespass disputes among ranchers in northern California (Ellickson 1991). While one gives up the careful control of experimental work by moving into the field, one can examine situations involving truly large groups and significant (even life-threatening) outcomes. I believe the strongest work combines multiple methodologies. Yamagishi and his colleagues (Yamagishi et al 1994, Yamagishi & Takahashi 1994, Yamagishi & Yamagishi 1994), for example, have combined simulations, survey research, and experimental studies in their research on trust and social dilemmas. Ostrom and her colleagues (Ostrom et al 1994) also have a long history of using multiple methods, combining field studies with experimental work.

Future Directions

In addition to the advances in research design and issues mentioned above, I believe especially promising directions for research in the future include work on the expectations and effects of generalized reciprocity within groups, the transformation of incentive structures, and a greater focus on the Assurance Game as a core model in understanding social dilemmas.

A great deal of attention has focused on how dilemmas might be structurally changed to reduce or eliminate the temptations to defect. But additional studies on how the incentive structure in dilemmas can be transformed via motivational or strategic means are also crucial. We have seen that there appear to be stable personality traits that result in distinct transformations of objective payoffs, and work on group identity provides evidence of transformation of payoffs depending on group membership. The robust effects of group identity and the expectation of reciprocity imply that such issues as the construction of group boundaries and the signaling of group membership will be of fundamental importance to the study of social dilemmas. Incentive structures can also be transformed via strategic mechanisms such as the adoption of a Tit-for-Tat strategy in a repeated Prisoner's Dilemma Game.

It is noteworthy that the result of many of these transformations is the framing of the social dilemma as an Assurance Game. Working within an Assurance Game does not eliminate the challenge of cooperation, but it does change our focus in many ways. Trustworthiness, trustfulness, and all those factors that influence these concepts become even more important (cf Yamagishi &

Yamagishi 1994). Further, because the key issue in these dilemmas is the assurance that others will cooperate, attempts to signal and advertise one's commitment to cooperate will be critical. This might be as simple as a public pledge to cooperate or an act that is more symbolic.[21] In this sense, signs that one is committed to a group or to a particular goal would be important in encouraging others to cooperate (e.g. wearing a crucifix, a lapel pin from a fraternal organization, gang colors, a union pin). More broadly, using the Assurance Game as one's model makes signaling and signal detection (or in the language of social psychology, dramaturgy and attribution) centrally relevant to a study of human cooperation.[22]

Transformations can also be important because they provide another potential path to solving social dilemmas. Rather than trying to solve the dilemma as it exists, it may be easier to work to transform the dilemma to, e.g., an Assurance Game and then use another set of more viable strategies to encourage cooperation (Kollock 1998a, Yamagishi 1995). Bornstein et al's (1990) research in which an *N*-person Prisoner's Dilemma Game is transformed into a Chicken Game via inter-group competition is an example of such an approach.

Ideally, we should pursue experimental designs that permit large groups, sizable incentives, and diverse populations. One possible approach to these challenges is to move away from physical group laboratories to experimental systems that are designed to make use of the many advantages of the Internet (Macy et al 1997). An experimental system based on the World Wide Web would make it much easier to run studies involving very large groups that are composed of more than just college undergraduates. Such a Web-based lab would also permit cross-national experiments, which would enable studies of cross-cultural interactions. One could even make use of monetary exchange rates in order to run experiments in countries where the incentives offered would be truly significant.

Uniting all of these studies is a core set of social dilemmas that can be explicitly and precisely modeled. This core set of models can serve as a kind of lingua franca for communication between disciplines. Researchers with very different goals and methodologies can map the results of each other's work onto their own through the use of these models. However, this potential is often not realized, and it is often the case that scholars in different disciplines remain unaware of each other's work. One of the most important goals for future work is for researchers to become more aware of related literatures in neigh-

[21]An excellent example of these processes is given by Fantasia (1988) in his case study of a wildcat strike.

[22]Note that dramaturgical issues are very important in a Chicken Game as well as in an Assurance Game, but for different reasons. In a Chicken Game one can try to convince the other that one will certainly not cooperate in hopes of forcing the partner to do so.

boring disciplines. In particular, I would encourage closer ties with experimental economists, who have produced a very useful body of work on social dilemmas (for a starting point, see Kagel & Roth 1995). We should work toward the integration of these various research traditions and the future collaboration of experimental social scientists across the disciplines of sociology, psychology, economics, political science, and anthropology.

ACKNOWLEDGMENTS

I owe a particular debt to Toshio Yamagishi, Marilynn Brewer, David Messick, Robyn Dawes, and John Orbell for their work and for the conversations I have had with them. I also thank Motoki Watabe for comments on an earlier draft.

Literature Cited

Alfano G, Marwell G. 1980. Experiments on the provisions of public goods by groups. III. Nondivisibility and free riding in real groups. *Soc. Psychol. Q.* 43(3):300–9

Andreoni J. 1988. Why free ride? Strategies and learning in public goods experiments. *J. Public Econ.* 37:291–304

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

Axelrod R. 1986. An evolutionary approach to norms. *Am. Polit. Sci. Rev.* 80:1095–1111

Axelrod R. 1997. *The Complexity of Cooperation.* Princeton, NJ: Princeton Univ. Press

Axelrod R, D'Ambrosio L. 1994. Annotated bibliography of the evolution of cooperation. Online document: http://ippsweb.ipps.lsa.umich.edu:70/1/ipps/papers/coop

Axelrod R, Dion D. 1988. The further evolution of cooperation. *Science* 242:1385–90

Banks J, Plott C, Porter D. 1988. An experimental analysis of unanimity in public goods provision mechanisms. *Rev. Econ. Stud.* 55:301–22

Bonacich P, Shure G, Kahan J, Meeker R. 1976. Cooperation and group size in the n-person prisoner's dilemma. *J. Conflict Resolution* 20:687–706

Bornstein G, Erev I, Rosen O. 1990. Intergroup competition as a structural solution to social dilemmas. *Soc. Behav.* 5(4): 247–60

Bornstein G, Rapoport A. 1982. Intergroup competition for the provision of step-level public goods: effects of preplay communication. *Eur. J. Soc. Psychol.* 18:125–42

Brewer MB. 1981. Ethnocentrism and its role in interpersonal trust. In *Scientific Inquiry and the Social Sciences*, ed. M Brewer, B Collins, pp. 345–60. San Francisco: Jossey-Bass

Brewer MB, Kramer RM. 1986. Choice behavior in social dilemmas: effects of social identity, group size, and decision framing. *J. Pers. Soc. Psychol.* 50:543–49

Bromley DW. 1991. *Environment and Economy: Property Rights and Public Policy*. Oxford, UK: Blackwell

Bromley DW, Feeny D, McKean M, Peters P, Gilles J, et al, eds. 1992. *Making the Commons Work: Theory, Practice, and Policy.* San Francisco: Inst. Contemp. Stud.

Buchanan JM. 1971. *The Bases for Collective Action*. New York: General Learning

Caldwell MD. 1976. Communication and sex effects in a five-person prisoner's dilemma game. *J. Pers. Soc. Psychol.* 33:273–80

Cass RC, Edney JJ. 1978. The commons dilemma: a simulation testing the effects of resource visibility and territorial division. *Hum. Ecol.* 6(4):371–86

Cornes R, Sandler T. 1996. *The Theory of Externalities, Public Goods, and Club Goods*. Cambridge: Cambridge Univ. Press. 2nd ed.

Cross J, Guyer M. 1980. *Social Traps*. Ann Arbor: Univ. Mich. Press
Crowe BL. 1969. The tragedy of the commons revisited. *Science* 166:1103–7
Dawes R. 1980. Social dilemmas. *Annu. Rev. Psychol.* 31:169–93
Dawes R. 1991. Social dilemmas, economic self-interest, and evolutionary theory. In *Frontiers of Mathematical Psychology: Essays in Honor of Clyde Coombs*, ed. DR Brown, JEK Smith, pp. 53–79. New York: Springer-Verlag
Dawes R, McTavish J, Shaklee H. 1977. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *J. Pers. Soc. Psychol.* 35:1–11
Edney JJ, Harper CS. 1978a. The commons dilemma: a review. *Contrib. Psychol. Environ. Manage.* 2:491–507
Edney JJ, Harper CS. 1978b. The effects of information in a resource management problem: a social trap analog. *Hum. Ecol.* 6: 387–95
Ellickson RC. 1991. *Order Without Law: How Neighbors Settle Disputes*. Cambridge, MA: Harvard Univ. Press
Fairlie S, Hildyard N, Lohmann N, Sexton S, et al. 1994. Whose common future: reclaiming the commons. *Environ. Urbanization* 6(1):106–30
Fantasia R. 1988. *Cultures of Solidarity: Consciousness, Action, and Contemporary American Workers*. Berkeley: Univ. Calif. Press
Fox DR. 1985. Psychology, ideology, utopia, and the commons. *Am. Psychol.* 40(1): 48–58
Fox J, Guyer M. 1977. Group size and others' strategy in an n-person game. *J. Conflict Resolut.* 21:323–38
Fox J, Guyer M. 1978. 'Public' choice and cooperation in n-person prisoner's dilemma. *J. Conflict Resolut.* 22:469–81
Frank R, Gilovich T, Regan D. 1993. Does studying economics inhibit cooperation? *J. Econ. Perpsect.* 7(2):159–71
Gambetta D. 1993. *The Sicilian Mafia: The Business of Private Protection*. Cambridge, MA: Harvard Univ. Press
Glance NS, Huberman BA. 1994. The dynamics of social dilemmas. *Sci. Am.* (Mar.): 76–81
Hamburger H, Guyer M, Fox J. 1975. Group size and cooperation. *J. Conflict Resolut.* 19:503–31
Hardin G. 1968. The tragedy of the commons. *Science* 162:1243–48
Hardin G, Baden J. 1977. *Managing the Commons*. San Francisco: Freeman
Hayashi N. 1993. From tit-for-tat to out-for-tat: the dilemma of the prisoner's network. *Sociol. Theory Methods* 8(1)(13):19–32 (in Japanese)
Hayashi N, Ostrom E, Walker J, Yamagishi T. 1997. *Reciprocity, trust and the illusion of control: a cross-societal study*. Work. Pap. Indiana Univ.
Hechter M. 1984. When actors comply: monitoring costs and the production of social order. *Acta Sociol.* 27:161–83
Heckathorn DD. 1988. Collective sanctions and the creation of prisoner's dilemma norms. *Am. J. Sociol.* 94(3):535–62
Heckathorn DD. 1989. Collective action and the second-order free-rider problem. *Rationality Soc.* 1(1):78–100
Heckathorn DD. 1996. The dynamics and dilemmas of collective action. *Am. Sociol. Rev.* 61:250–77
Hinde RA, Groebel J. 1991. *Cooperation and Prosocial Behaviour*. Cambridge, UK: Cambridge Univ. Press
Hume D. 1739/1976. *A Treatise of Human Nature*. Oxford, UK: Oxford Univ. Press
Isaac RM, McCue K, Plott C. 1985. Public goods provision in an experimental environment. *J. Public Econ.* 26:51–74
Isaac RM, Schmidtz D, Walker J. 1988. The assurance problem in a laboratory market. *Public Choice* 62(3):217–36
Isaac RM, Walker J. 1988. Group size effects in public goods provision: the voluntary contribution mechanism. *Q. J. Econ.* 103: 179–99
Isaac RM, Walker J, Thomas S. 1984. Divergent evidence on free riding: an experimental examination of possible explanations. *Public Choice* 43:113–49
Isaac RM, Walker J, Williams A. 1990. *Group size and the voluntary provision of public goods: experimental evidence utilizing very large groups.* Work. Pap. Indiana Univ.
Jerdee TH, Rosen B. 1974. Effects of opportunity to communicate and visibility of individual decisions on behavior in the common interest. *J. Appl. Soc. Psychol.* 59: 712–16
Jin N, Yamagishi T. 1997. Group heuristics in social dilemmas. *Jpn. J. Soc. Psychol.* 12: 190–98 (in Japanese with English summary)
Jin N, Yamagishi T, Kiyonari T. 1996. Bilateral dependency and the minimal group paradigm. *Jpn. J. Psychol.* 67:77–85 (in Japanese with English summary)
Kagel J, Roth A. 1995. *The Handbook of Experimental Economics.* Princeton, NJ: Princeton Univ. Press
Kahan JP. 1973. Noninteraction in an anonymous three-person prisoner's dilemma game. *Behav. Sci.* 18(2):124–27

Kakiuchi R, Yamagishi T. 1997. *The dilemma of trust*. Seventh Int. Conf. on Soc. Dilemmas, Cairns, Australia

Karp D, Jin N, Yamagishi T, Shinotsuka H. 1993. Raising the minimum in the minimal group paradigm. *Jpn. J. Exp. Soc. Psychol.* 32:231–40

Kelley HH. 1979. *Personal Relationships: Their Structures and Processes*. Hillsdale, NJ: Lawrence Erlbaum

Kelley HH, Grzelak J. 1972. Conflict between individual and common interest in an N-person relationship. *J. Pers. Soc. Psychol.* 21:190–97

Kelley HH, Thibaut JW. 1978. *Interpersonal Relations: A Theory of Interdependence*. New York: Wiley & Sons

Kerr N. 1989. Illusions of efficacy: the effects of group size on perceived efficacy in social dilemmas. *J. Exp. Soc. Psychol.* 25: 287–313

Kerr N. 1992. Efficacy as a causal and moderating variable in social dilemmas. In *Social Dilemmas: Theoretical Issues and Research Findings*, ed. WBG Liebrand, DM Messick, HAM Wilke, pp. 59–80. Oxford: Pergamon

Kerr N. 1997. *Anonymity and social control in social dilemmas*. Seventh Int. Conf. on Soc. Dilemmas, Cairns, Australia

Kim O, Walker M. 1984. The free rider problem: experimental evidence. *Public Choice* 43:3–24

Klandermans B. 1986. Individual behavior in real life social dilemmas: a theory and some research results. In *Experimental Social Dilemmas,* ed. HAM Wilke, DM Messick, C Rutte, pp. 87–111. Frankfurt: Verlag Peter Lang

Kollock P. 1993. 'An eye for an eye leaves everyone blind': cooperation and accounting systems. *Am. Sociol. Rev.* 58(6):768–86

Kollock P. 1998a. Transforming social dilemmas: group identity and cooperation. In *Modeling Rational and Moral Agents,* ed. P Danielson, pp. 186–210. Oxford, UK: Oxford Univ. Press

Kollock P. 1988b. Social Dilemmas: Online Resources. www.sscnet.ucla.edu/ soc/faculty/kollock/dilemmas

Kollock P, Smith M. 1996. Managing the virtual commons: cooperation and conflict in computer communities. In *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*, ed. S Herring, pp. 109–28. Amsterdam: John Benjamins

Komorita SS. 1987. Cooperative choice in decomposed social dilemmas. *Pers. Soc. Psychol. Bull.* 13:53–63

Komorita SS, Lapworth CW. 1982. Cooperative choice among individuals versus groups in an n-person dilemma situation. *J. Pers. Soc. Psychol.* 42:487–96

Komorita SS, Parks CD. 1995. Interpersonal relations: mixed-motive interaction. *Annu. Rev. Psychol.* 46:183–207

Komorita SS, Parks CD. 1996. *Social Dilemmas*. Boulder, CO: Westview

Komorita SS, Sweeney J, Kravitz DA. 1980. Cooperative choice in the n-person dilemma situation. *J. Pers. Soc. Psychol.* 38: 504–16

Kramer RM, Brewer MB. 1984. Effects of group identity on resource use in a simulated commons dilemma. *J. Pers. Soc. Psychol.* 46:1044–56

Kramer RM, Brewer MB. 1986. Social group identity and the emergence of cooperation in resource conservation dilemmas. In *Experimental Social Dilemmas,* ed. HAM Wilke, DM Messick, C Rutte, pp. 205–34. Frankfurt: Verlag Peter Lang

Kramer RM, McClintock CG, Messick DM. 1986. Social values and cooperative response to a simulated resource conservation crisis. *J. Pers.* 54:576–91

Kuhlman DM, Camac CR, Cunha DA. 1986. Individual differences in social orientation. In *Experimental Social Dilemmas,* ed. HAM Wilke, DM Messick, C Rutte, pp. 151–74. Frankfurt: Verlag Peter Lang

Kuhlman DM, Marshello AFJ. 1975. Individual differences in game motivation as moderators of preprogrammed strategy effects in prisoner's dilemma. *J. Pers. Soc. Psychol.* 32(5):922–31

Ledyard JO. 1995. Public goods: a survey of experimental research. In *The Handbook of Experimental Economics,* ed. JH Kagel, AE Roth, pp. 111–94. Princeton, NJ: Princeton Univ. Press

Liebrand WBG. 1984. The effect of social motives, communication and group size on behavior in an n-person multi-stage mixed-motive game. *Eur. J. Soc. Psychol.* 14:239–64

Liebrand WBG, Messick DM, eds. 1996. *Frontiers in Social Dilemmas Research*. Berlin: Springer Verlag

Liebrand WBG, Messick DM, Wilke HAM, eds. 1992. *Social Dilemmas: Theoretical Issues and Research Findings*. New York: Pergamon

Liebrand WBG, Wilke HAM, Vogel R, Wolters FJM. 1986. Value orientation and conformity in three types of social dilemma games. *J. Conflict Resolut.* 30:77–97

Lloyd WF. [1832] 1977. On the checks to population. Reprinted in *Managing the Commons,* ed. G Hardin, J Baden. San Francisco: Freeman

Lohmann S. 1995. The poverty of Green and Shapiro. *Crit. Rev.* 9(1–2):127–300
Luce D, Raifa H. 1957. *Games and Decisions*. New York: Wiley
Macy MW. 1991. Chains of cooperation: threshold effects in collective action. *Am. Sociol. Rev.* 56(12):730–47
Macy MW. 1993. Social learning and the structure of collective action. In *Advances in Group Processes,* ed. E Lawler et al, 10:1–36. Greenwich, CT: JAI
Macy MW, Kollock P, Yamagishi T. 1997. *The 'Tiger's Cave' experiments: a proposal for a web-based laboratory for cross-cultural research on trust in social dilemmas.* Work. Pap. Cornell Univ.
Maki JE, Hoffman DM, Berk RA. 1978. A time series analysis of the impact of a water conservation campaign. *Eval. Q.* 2: 107–18
Marwell G, Ames R. 1981. Economists free ride, does anyone else? Experiments on the provision of public goods, IV. *J. Public Econ.* 15:295–310
Marwel G, Oliver P. 1993. *The Critical Mass in Collective Action: A Micro-Social Theory.* Cambridge, MA: Cambridge Univ. Press
McCay BJ, Acheson JM, eds. 1987. *The Question of the Commons: The Culture and Ecology of Communal Resources*. Tucson: Univ. Ariz. Press
McClintock CG. 1974. Development of social motives in Anglo-American and Mexican-American children. *J. Pers. Soc. Psychol.* 29:348–54
McClintock CG, Keil LJ. 1983. Social values: their definition, their development, and their impact upon human decision making in settings of outcome interdependence. In *Small Groups and Social Interaction*, ed. H Blumberg, A Hare, V Kent, M Davies, 2:123–43. London: Wiley
McClintock CG, Liebrand WBG. 1988. Role of interdependence structure, individual value orientation, and another's strategy in social decision making: a transformational analysis. *J. Pers. Soc. Psychol.* 55(3): 396–409
Messick DM, Brewer MB. 1983. Solving social dilemmas. In *Review of Personality and Social Psychology,* ed. L Wheeler, P Shaver, 4:11–44. Beverly Hills, CA: Sage
Messick DM, McClelland CL. 1983. Social traps and temporal traps. *Soc. Psychol. Bull.* 9:105–10
Messick DM, Wilke HAM, Brewer MB, Kramer RM, Zemke P, Lui L. 1983. Individual adaptations and structural change as solutions to social dilemmas. *J. Pers. Soc. Psychol.* 44:294–309
Moe TM. 1980. *The Organization of Interests: Incentives and Internal Dynamics of Political Interest Groups.* Chicago, IL: Univ. Chicago Press
Mueller EN, Opp KD. 1986. Rational choice and rebellious collective action. *Am. Polit. Sci. Rev.* 80:471–87
Olson M. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard Univ. Press
Orbell J, Dawes R. 1981. Social dilemmas. In *Progress in Applied Social Psychology*, ed. GM Stephenson, JM Davis, 1:37–65. New York: Wiley
Orbell J, Dawes R. 1993. Social welfare, cooperators' advantage, and the option of not playing the game. *Am. Sociol. Rev.* 58(6): 787–800
Orbell J, Dawes R, van de Kragt A. 1990. The limits of multilateral promising. *Ethics* 100:616–27
Orbell J, van de Kragt A, Dawes R. 1988. Explaining discussion-induced cooperation. *J. Pers. Soc. Psychol.* 54:811–19
Ostrom E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge Univ. Press
Ostrom E. 1992. *Crafting Institutions for Self-Governing Irrigation Systems*. San Francisco: ICS
Ostrom E, Gardner R, Walker J. 1994. *Rules, Games, and Common-Pool Resources*. Ann Arbor: Univ. Mich. Press
Poundstone W. 1992. *Prisoner's Dilemma.* New York: Doubleday
Rapoport A. 1967. Optimal policies for the prisoner's dilemma. *Psychol. Rev.* 75: 136–48
Rapoport A, Bornstein G. 1987. Intergroup competition for the provision of binary public goods. *Psychol. Rev.* 94:291–99
Rapoport A, Bornstein G. 1989. Solving public good problems in competition between equal and unequal size groups. *J. Conflict Resolut.* 33:460–79
Rapoport A, Bornstein G, Erev I. 1989. Intergroup competition for public goods: effects of unequal resource and relative group size. *J. Pers. Soc. Psychol.* 56: 748–56
Rutte CG, Wilke HAM. 1985. Preferences for decision structures in a social dilemma situation. *Eur. J. Soc. Psychol.* 15:367–70
Samuelson CD, Messick DM. 1986a. Alternative structural solutions to resource dilemmas. *Organ. Behav. Hum. Decision Processes* 37:139–55
Samuelson CD, Messick DM. 1986b. Inequities in access to and use of shared re-

sources in social dilemmas. *J. Pers. Soc. Psychol.* 51:960–67
Samuelson CD, Messick DM, Rutte CG, Wilke HAM. 1984. Individual and structural solutions to resource dilemmas in two cultures. *J. Pers. Soc. Psychol.* 47:94–104
Samuelson P. 1954. The pure theory of public expenditure. *Rev. Econ. Statist.* 36:387–89
Schelling TC. 1978. *Micromotives and Macrobehavior*. New York: Norton
Schroeder DA, ed. 1995. *Social Dilemmas: Perspectives on Individuals and Groups*. Westport, CT: Praeger
Schuessler R. 1989. Exit threats and cooperation under anonymity. *J. Conflict Resolut.* 33:728–49
Schulz U, Albers W, Mueller U, eds. 1994. *Social Dilemmas and Cooperation*. New York: Springer-Verlag
Sherif M, Harvey O, White B, Hood W, Sherif C. 1961. *Intergroup Conflict and Cooperation: The Robber's Cave Experiment*. Norman, OK: Inst. Group Relat., Univ. Oklahoma
Stroebe W, Frey BS. 1982. Self-interest and collective action: the economics and psychology of public goods. *Br. J. Soc. Psychol.* 21:121–37
Suleiman R, Rapoport A. 1992. Provision of step-level public goods with continuous contribution. *J. Behav. Decision Making* 5(2):133–53
Tajfel H. 1981. *Human Groups and Social Categories*. Cambridge, UK: Cambridge Univ. Press
Toda M, Shinotsuka H, McClintock CG, Stech F. 1978. Development of competitive behavior as a function of culture, age, and social comparison. *J. Pers. Soc. Psychol.* 36: 835–39
Udehn L. 1993. Twenty-five years with *The Logic of Collective Action*. *Acta Sociol.* 36:239–61
van de Kragt A, Orbell J, Dawes R. 1983. The minimal contributing set as a solution to public goods problems. *Am. Polit. Sci. Rev.* 77:112–22
van Lange PAM, Liebrand WBG, Messick DM, Wilke HAM. 1992. Social dilemmas: the state of the art—introduction and literature review. In *Social Dilemmas: Theoretical Issues and Research Findings*, ed. WBG Liebrand, DM Messick, HAM Wilke, pp. 3–28. Oxford: Pergamon
Vanberg VJ, Congleton RD. 1992. Rationality, morality, and exit. *Am. Polit. Sci. Rev.* 86:418–31
Watabe M. 1992. Choice of strategies in a social dilemma. *Jpn. J. Exp. Soc. Psychol.* 32:171–82 (in Japanese with English summary)
Watabe M, Terai S, Hayashi N, Yamagishi T. 1996. Cooperation in the one-shot Prisoner's Dilemma based on expectations of reciprocity. *Jpn. J. Exp. Soc. Psychol.* 36: 183–96 (in Japanese with English summary)
Watabe M, Yamagishi T. 1994. Choice of strategies in social dilemma supergames. In *Social Dilemmas and Cooperation,* ed. U Schulz, W Albers, U Mueller, pp. 251–67. New York: Springer-Verlag
Wilke HAM, Messick DM, Rutte CG, eds. 1986. *Experimental Social Dilemmas*. Frankfurt: Verlag Peter Lang
Winett RA, Kagel JH, Battalio RC, Winkler RC. 1978. The effects of monetary rebates, feedback, and information on residential energy conservation. *J. Appl. Psychol.* 63: 73–80
Yamagishi T. 1986. The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* 51:110–16
Yamagishi T. 1988. Seriousness of social dilemmas and the provision of a sanctioning system. *Soc. Psychol. Q.* 51(1):32–42
Yamagishi T. 1992. Group size and the provision of a sanctioning system in a social dilemma. In *Social Dilemmas: Theoretical Issues and Research Findings,* ed. WBG Liebrand, DM Messick, HAM Wilke, pp. 267–87. Oxford: Pergamon
Yamagishi T. 1995. Social dilemmas. In *Sociological Perspectives on Social Psychology*, ed. K Cook, GA Fine, JS House, pp. 311–35. Boston: Allyn & Bacon
Yamagishi T, Cook K. 1993. Generalized exchange and social dilemmas. *Soc. Psychol. Q.* 56(4):235–48
Yamagishi T, Hayashi N, Jin N. 1994. Prisoner's dilemma networks: selection strategy versus action strategy. In *Social Dilemmas and Cooperation,* ed. U Schulz, W Albers, U Mueller, pp. 233–50. New York: Springer-Verlag
Yamagishi T, Takahashi N. 1994. Evolution of norms without metanorms. In *Social Dilemmas and Cooperation,* ed. U Schulz, W Albers, U Mueller, pp. 311–26. New York: Springer-Verlag
Yamagishi T, Yamagishi M. 1994. Trust and commitment in the United States and Japan. *Motivation Emotion* 18(2):129–66