THE DYNAMICS AND DILEMMAS OF COLLECTIVE ACTION*

Douglas D. Heckathorn
University of Connecticut

Theoretical accounts of participation in collective action have become more divergent. Some analysts employ the Prisoner's Dilemma paradigm, other analysts suggest that different social dilemmas underlie collective action, and still others deny that social dilemmas play any significant role in collective action. I propose a theoretically exhaustive inventory of the dilemmas arising in collective action systems and show that five games, including the Prisoner's Dilemma, can underlie collective action. To analyze action within each game I use a dynamic selectionist model based on three modes of organization—voluntary cooperation, strategic interaction, and selective incentives. Social dilemmas exist in four of the five games, and conflicting accounts of collective action have focused on different games and modes of organization. As collective action proceeds from initiation to rapid expansion to stability, its game type varies in a way that can be precisely characterized as movement through a two-dimensional game-space. Finally, I distinguish between two ways of promoting collective action: One way focuses on resolving the dilemma within a particular game; the other focuses on changing the game so the dilemma is more easily resolved or eliminated.

Collective interests do not necessarily produce collective action. For example, not all oppressed groups revolt, even when, in combination, their power vastly exceeds that of their controllers. Similarly, in contemporary political systems, large but poorly organized constituencies are often controlled by smaller but better organized groups. These examples illustrate failures of collective action. On other occasions, collective action can succeed under exceedingly adverse conditions. For example, in trench warfare during World War I, informal truces provided a daily respite from the fighting. Truces emerged even though soldiers on opposing sides could not communicate directly, and these truces remained stable despite active opposition by officers on both sides (Axelrod 1984). Despite an impressive growth of theory and empirical research during the last several decades, the origins and dynamics of collective action remain disputed. Indeed, explanations of the forces leading to participation in collective action have grown more divergent.

Olson's (1965) analysis frames much of the debate. In his view, collective action contains a social dilemma, that is, a situation in which actions that are individually rational can lead to outcomes that are collectively irrational. Game theorists have identified many such dilemmas (Rasmusen 1989). In Olson's analysis, everyone seeks to reap the benefits of others' participation while evading the cost of participation. When everyone

* Direct all correspondence to Douglas Heckathorn, Department of Sociology, University of Connecticut, Storrs, CT 06269–2068 (heckath@uconnvm.uconn.edu). Earlier versions of this paper were presented at meetings of the Public Choice Society and at colloquia series sponsored by the University of Connecticut Department of Economics, the University of Utrecht, the University of Gröningen, and the University of Leipzig. The research was funded in part by grants from the National Science Foundation (SES–9022926) and the National Institute on Drug Abuse (DA RO1 DA05517). I thank Denise Anthony, Robert Broadhead, Jack Hirshleifer, Siegwart Lindenberg, Susan LoBello, Karl-Dieter Opp, Werner Raub, David Weakliem, Rolf Ziegler, and anonymous *ASR* reviewers for helpful comments and advice. [Reviewers acknowledged by the author include Satoshi Kanazawa and Gerald Marwell. —*Ed.*]

acts in this manner, collective action fails. Thus, according to Olson, free-riding is a problem in all but the smallest groups. In small groups, each individual's participation affects the group's prospects for success, and it is noticeable to all members. However, in larger groups, the impact of any individual's participation is negligible, so self-interested, rational individuals will choose to free-ride unless they are restrained. The result is a free-rider problem that can be formally represented as an N-person Prisoner's Dilemma (Hardin 1971). This dilemma can be resolved in two ways. In moderate-sized groups, it can be resolved through strategic interaction, that is, reciprocity that says "if you cooperate, then I will too." In larger groups, collective action requires selective incentives, such as laws or social norms that punish defectors or reward cooperators. This analysis has been immensely influential (Oberschall 1973; Opp 1989).

Marwell and Oliver (1993) take issue with many of Olson's conclusions. They see collective goods, like other goods, as characterized by third-order production functions (Figure 1A). At the beginning of production, the curve *accelerates*, reflecting increasing marginal returns to initial contributions as start-up costs are progressively absorbed. Here, collective action faces a start-up problem because the return to the initial contributors is minimal. Unless a "critical mass" of strongly motivated individuals is willing to absorb these costs, collective action never begins. As Marwell and Oliver (1993) state, ". . . free riding is *not* the problem in the accelerative case, unless all public goods dilemmas are said to be free riding by definitional fiat" (p. 182). Here the threat to collective action is not that some members free-ride on the contributions of others, but that no one will see any gain from contribution. This poses a start-up dilemma.

Next, the function becomes linear in the range where the marginal gain becomes identical for early or late contributors. Here, collective action has an all-or-nothing character—it is rational for everyone to contribute or for no one to contribute (Oliver, Marwell, and Teixeira 1985:533–34), so no subset of actors can free-ride on the contributions of others. Hence, again there is no free-rider problem.

Finally, the curve *decelerates*, reflecting the decreasing marginal returns that occur when output limits are approached. Here free-riding can arise from two similar mechanisms, "order effects" and a "surplus." In the order effect, "[t]he less interested members free-ride on the initial contributions of the most interested, and total group contributions are suboptimal" (Marwell and Oliver 1993: 82–83). This is Marwell and Oliver's theoretical explication of Olson's (1965) well-known principle of "the exploitation of the great by the small" (p. 29). In the surplus mechanism, free-riding can arise when, consistent with a decelerating production function, the number of individuals willing to contribute to the collective good decreases as the number of contributors increases. This produces a surplus of contributors in the sense that some of those who were initially willing to contribute will refuse if others have contributed first. In this case, ". . . the first ones who happen to be faced with the decision [to contribute] are 'stuck.' They will contribute because they find it profitable to do so while those whose turn to decide comes later will free-ride" (Marwell and Oliver 1993:85). This free-rider problem differs from Olson's. Whereas for Marwell and Oliver there is no initial temptation to free-ride, this temptation is ever-present for Olson. Thus, Marwell and Oliver's free-rider problem cannot be represented as an N-person Prisoner's Dilemma. Thus Marwell, Oliver, and Olson identify three distinct social dilemmas that can underlie collective action.

A third group of analysts denies that a free-rider problem significantly affects collective action (Fireman and Gamson 1979; Klandermans 1988; Ferree 1992; Oegema and Klandermans 1994). For example, Fireman and Gamson (1979) argue that when riots and demonstrations are successful it is because the promoters of collective action "build solidarity, raise consciousness of common interests, and create opportunities for collective action" (pp. 8–9). These increase the expected value of the movement's goal and reduce the cost of participation. They conclude that Olson's account of collective action as requiring selective incentives is applicable only in "special circumstances." Klandermans's (1988) conclusions after analyzing participation in unions are similar:

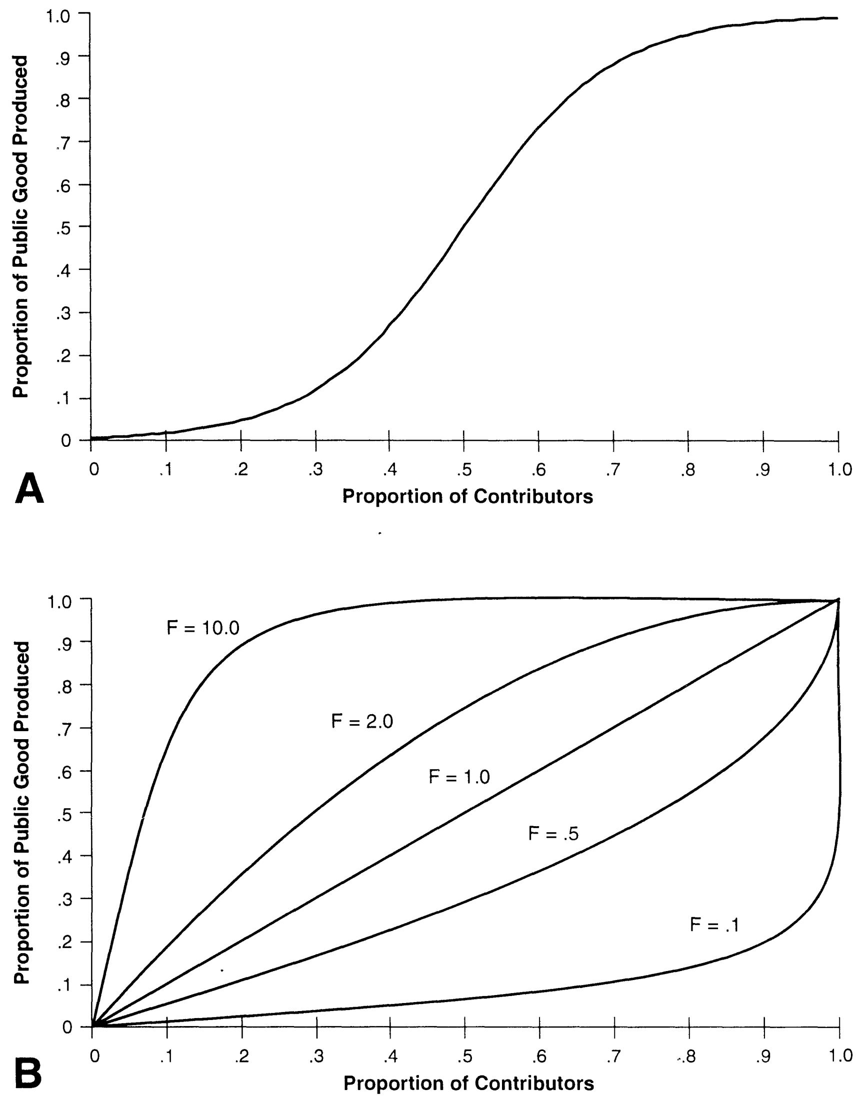


Figure 1. Production Functions Showing the Relationship between Proportion of Contributors and Proportion of the Public Good Produced

> . . . [T]he possibility that free-rider conditions are satisfied is highest in companies with the highest degree of unionization, but such companies also provide the most opportunities for promoting participation by means of selective incentives, . . . and so it does not automatically follow that the number of free riders will be highest in these companies. (P. 89)

Klandermans (1988:89) attributes most nonparticipation in collective action, not to free-riding, but to an expectation that collective action will fail. He says that participants in collective action tend to believe that they can succeed and that their own participation will make a difference.

Thus, views regarding whether or how social dilemmas are related to collective action are highly divergent. My first aim is to derive a theoretically exhaustive inventory of the dilemmas arising in collective action systems. This part of the analysis draws on game theory because each game corresponds to a unique structure of individual and collective interests. I prove that exactly five games characterize the structure of opposed and complementary interests within which collective action occurs. In one game, consistent with some critiques of the free-rider model, no dilemma arises. In each of the other four games, a distinctive dilemma arises, including Marwell and Oliver's start-up and free-rider problems, Olson's free-rider problem, and a fourth type. The analysis further shows that the nature of the strategic problem is dynamic: It changes during the course of collective action because a system can move through all five game types—from the initiation of collective action to the eventual stabilization of a mature production system. Therefore, differing depictions of collective action validly apply to distinct phases of collective action.

Second, I analyze collective action based on each of the three mechanisms that underlie collective action—voluntary cooperation, strategic interaction, and selective incentives. Each mechanism plays a distinctive role in each game type, and hence in each phase of collective action. Yet many analysts have focused on only one or two of these mechanisms, and this has contributed to the increasingly diverse views of collective action.

Finally, instead of employing the rational choice and experiential learning models of previous analyses of collective action, I show that an observational learning model can be used to analyze collective action.

THE ANATOMY OF COLLECTIVE ACTION

To identify the full range of factors that can render collective action problematic, I adopt a two-step approach. First, I specify the range of situations that can arise during collective action. Second, I analyze these situations to see what types of games they fit. This serves to specify when cooperation by rational actors is problematic and, if so, whether the dilemma lies in issues of coordination, in bargaining over how gains are to be divided, or in lack of trust that the other will cooperate.

Any collective action system involves some form of public or semipublic good (Mueller 1989:11) and hence has two characteristics. First, excluding anyone from consumption of the good is impractical. For example, scabs benefit from higher wages won through strikes. A second characteristic is jointness of supply. In the case of pure jointness, production costs are fixed. For example, a public monument can be enjoyed by any number of individuals, so the marginal cost of each additional viewer is 0. Some measure of jointness exists if average production costs decline as the number of consumers of the good increases. In addition, I assume that there is a production function linking contributions by group members to the ultimate level of public good produced, and that contributions entail a cost, even if it is only the opportunity cost of alternative investments of time forgone. These assumptions can be formalized as follows: Cooperation involves production of a collective good with value V at level L, where L is the proportion of the collective good produced relative to the maximum amount that could be produced if all group members contributed to its production. As defined, L can vary from 0 (no production) to 1 (full production). The utility of the good produced is the product of its value and production level, $V * L$.

Collective action involves two distinct levels. Level one refers to personal contributions to produce the collective good. In the simplest case, individuals have the choice of contributing (cooperating at the first level), or not contributing (defecting at the first level). For example, one may decide whether to participate in a demonstration or help build a new canoe for one's village. Level two refers to what are termed "second-order collective goods" (Oliver 1980), such as selective incentives to reward first-level cooperators or punish first-level defectors. For example, one may decide whether to join in praising those who attended the demonstration or helped build the canoe, and whether to scorn those who failed to help.

In the proposed model, individuals are assumed to make dichotomous choices at the first level, and trichotomous choices at the second level. At the first level, they make a choice whether to contribute to collective goods production at cost K_{c1}, the cost (K) of cooperation (c) at level one. The amounts each actor can potentially contribute are assumed to be equal. (For analyses of the effects of variations of individual resources on collective action, see Heckathorn 1993:344–47 and Marwell and Oliver 1993:116). At the second level, actors are assumed to have three choices: They can contribute to a selective incentive system at cost K_{c2}, the cost of cooperation at level two; they can oppose the system of selective incentives at cost K_{o2}, the cost of opposition (*o*) at level two; or they can defect at this level, incurring no costs. Finally, if their choices are conditional on those of others, as when they engage in strategic interaction, they incur information costs, K_{oc}, termed the "cost of complexity" (Hirshleifer and Martinez Coll 1988:382). Under the above assumptions, utility functions have the form,

$$U = VL - K_1 - K_2 - K_{IN}, \tag{1}$$

where U is the actor's utility, V is the value of full production of the collective good, L is the proportion of the collective good that was produced, K_1 is first-level cost of cooperation (i.e., K_{c1} or 0, depending on whether the actor contributes to collective goods production or defects), K_2 is second-level cost of cooperation (i.e., K_{c2}, K_{o2}, or 0, depending on whether the actor exercises compliant control, oppositional control, or defects at the second [normative] level), and K_{IN} is information costs (i.e., K_{oc} or 0 depending on whether the actor adopts a conditional or nonconditional strategy).

Voluntary cooperation, strategic interaction, and selective incentives can all be conceived as entailing a choice among alternative strategies, where a strategy is defined as a rule specifying what choices to make in each possible situation. In a system based on *voluntary cooperation*, actors choose between two alternative strategies. If they cooperate, their strategy is called *private cooperation* (CD), because they contribute at the first level by contributing to the collective good but defect at the second level by forgoing any attempts to influence others. Alternatively, if they defect, their strategy is called *full defection* (DD), because they defect at both levels by refusing to contribute and permitting others to do as they wish.

When collective action is organized through strategic interaction, some actors make their choices conditional on others' choices. The standard strategy that embodies this approach is *Tit-for-Tat* (TFT), in which an actor begins by cooperating and subsequently behaves according to principles of reciprocity by answering cooperation with cooperation and defection with defection. Here, the threat to withhold cooperation serves as an incentive to motivate others to cooperate.

When collective action is organized through *selective incentives*, four additional strategies become possible. Two of these strategies use selective incentives to mandate participation in collective action. *Full cooperation* (CC) involves contributing to collective goods production (first-level cooperation) and sanctioning those who fail to contribute (second-level cooperation). In the terms of evolutionary game theory, full cooperation combines two traits: It is "moralistic" (Boyd and Richerson 1992:173), in that it acts to control other players, and it is "nice" (Axelrod 1984), in that it initially cooperates. Yet many moralists are not nice, as evidenced by the frequency of complaints about hypocrisy among public officials. It is easier to tell others how to behave properly than to do so oneself. Thus, *hypocritical cooperation* (DC) means an actor defects at the first level but cooperates at the second level, failing to contribute to the collective good while acting to compel others to contribute. Two final strategies embody countermobilization, that is, opposition to the creation of a system of selective incentives. *Compliant opposition* (CO) means cooperating at the first level but exercising oppositional control at the second level—the actor contributes to the collective good but defends the rights of others to refuse to contribute. Finally, *full opposition* (DO) means refusing to contribute and opposing norms that would compel compliance.

The Setting of Collective Action: The First-Level Game

The structure of the setting in which collective action occurs depends on the first-level game (i.e., the level of collective goods production), because this defines the link between individuals' choices of whether to contribute to collective goods production and the ultimate level of collective good produced. This link is defined by the collective good's production function. Consistent with other models (Marwell and Oliver 1993:77), I assume that the level of collective goods production in cooperative systems is a monotonically increasing function of the proportion of actors who contribute to production (i.e., each contributing actor makes some nonzero contribution to production).

The standard production function is an S-shaped curve. Because of this complexity, I follow Oliver et al. (1985) and Heckathorn (1989) in analyzing the accelerating, linear, and decelerating portions of the production functions.

To represent segments of the production function, I employ a function, previously proposed for this purpose (Heckathorn 1989), in which D is the number of actors in the group who defect (i.e., who fail to contribute to production of collective goods), N is the number of actors in the group, F is an exponent controlling the shape of the production function, and L is the level of collective goods produced:

$$L = 1 - (D/N)^F. \quad (2)$$

As defined, the level of collective goods produced can vary from $L = 0$, or no production, to $L = 1$, indicating full production. When all actors defect (i.e., $D = N$), no collective good is produced, and $L = 0$. When all actors contribute (i.e., $D = 0$), the collective good is fully produced and $L = 1$. When intermediate numbers of actors contribute (i.e., $0 < D < N$), the link between the proportion contributing and the level of collective goods produced depends on the value of the exponent, F.

Figure 1B presents the relationship between the proportion of the group that contributes $(N - D)/N$) and the level of production of the collective good (L) for various values of F, the term that controls the shape of the production function. When $F = 1$, the production function is linear—contributions from any given proportion of the group produce that proportion of the collective good. When $F > 1$, the production function is decelerating (i.e., the slope of the production function is a decreasing function of the number of contributors, $N - D$). Thus, the first contributors are the most productive, while subsequent contributors face decreasing marginal returns. For example, when $F = 2$ and 10 percent of the group contributes, they produce at the 19 percent level. In contrast, the last 10 percent contributing add only 1 percent to the level of collective goods production, so their incentive to contribute is weaker, as is the incentive of others to compel them to cooperate. Consequently, the incentive to initiate collective action is comparatively strong, while the incentive to compel 100 percent cooperation is rather weak. This pattern is more pronounced in more sharply decelerating production functions (e.g., when F=10, the first 10 percent produce 65 percent of the collective good, whereas the last 10 percent produce only .00000001 percent).

When $F < 1$, the production function is accelerating. When $F = .5$, the first 10 percent contributing produce only 5.1 percent of the collective good, while the last 10 percent contributing produce 32 percent. In a more sharply accelerating function, $F = .1$, the first 10 percent contributing produce only 1 percent of the collective good, while the final 10 percent produce 79 percent. Thus, a sharply accelerating production function requires near universal contribution to produce meaningful amounts of the collective good.

As is customary in much of the evolutionary game literature (Axelrod 1984; Hirshleifer and Martinez Coll 1988), actors are assumed to interact sequentially in pair-wise fashion.[1] Therefore, only three levels of collective good production are possible. If neither actor defects, the production level is $L = 1 - (0/2)^F = 1$; if only one defects, the production level is $L = 1 - (1/2)^F = 1 - .5^F$; if both defect the level of production is $L = 1 - (2/2)^F = 0$.

Based on the above production function, Table 1 shows the structure of the first-level collective-action game. Each player chooses between the two first-level strategies, private cooperation (CD), or full defection (DD). In

[1] That actors interact in pair-wise fashion is a restrictive assumption that will be removed in a subsequent paper. This extension of the model is straightforward because the seven strategies analyzed here can be validly extended to an N-person setting. For example, Boyd and Richerson (1985) use an N-person generalization of the tit-for-tat strategy in which the actor cooperates initially, and then cooperates if a threshold proportion of actors have cooperated during the previous period. Similarly, the other six strategies have been analyzed in N-person settings by Heckathorn (1993) using a forward-looking decision model.

the upper left cell, where both players contribute, the collective good is fully produced ($L = 1$). Each player receives the collective good's full value, V, but the payoff is reduced by the cost of that contribution K_{c1}, so the net payoff is

$$R = V - K_{c1}. \tag{3}$$

In deference to the Prisoner's Dilemma, this outcome is termed R for reward, although this term is not descriptive of all collective-good systems. In the upper right cell, row contributes while column defects, so row's "sucker" payoff, S, is

$$S = V(1 - .5^F) - K_{c1.} \tag{4}$$

This reflects the lower level of collective goods production resulting from one-half of the group's defection ($L = [1 - .5^F]$), and row's absorption of the contribution cost, K_{c1}. In the lower left cell, row defects while column contributes, so row's "temptation" payoff, T, is

$$T = V(1 - .5^F). \tag{5}$$

This reflects the lower level of collective goods production when only one player contributes, and row incurs no contribution cost. Finally, in the lower right cell, both players defect, so no collective good is produced ($V * 0 = 0$) and no contribution costs are incurred. Thus, the net payoff from the "punishment" outcome, P, is

$$P = 0. \tag{6}$$

The first-level game defines a set of games that depend on the value of two parameters (Figure 2). The horizontal axis depicts production functions that range from a sharply accelerating function ($F = .1$) through a linear function ($F = 1$) to a sharply decelerating function ($F = 10$). The vertical axis depicts valuations of the collective good ranging from the case in which the collective good has little value (i.e., $V/K_{c1} = .1$, so the value of full production of the collective good is only one-tenth the cost of contributing to its production), to cases in which the collective good is highly valued (i.e., $V/K_{c1} = 100$). Each point in the game-space depicted in Figure 2 corresponds to a unique value of the core game's payoffs T, R, P, and S. The solid lines show where two payoffs are equal. Therefore, *within* each of the five areas bordered by the lines, the order of the core game's payoffs is different. Because this order determines the structure of individual and collective interests, each order corresponds to a different type of game. *These five games are theoretically exhaustive of those that can arise in collective-action systems as defined in equations 1 through 6.* This is a small subset of all possible 2 × 2 games. There are 732 strategically distinct 2 × 2 games (Rapoport, Guyer, and Gordon 1976:31), of which 654 are "degenerate" in that some payoffs are equal. A taxonomy exists only for the 78 ordinally distinct games, that is, games in which all payoffs differ. In Rapoport et al.'s taxonomy, the five games depicted in Figure 2 correspond to numbers 6, 9, 12, 61, and 66. Table 2 presents the payoff matrices for the five games.

Table 1. Payoff Matrix for the First-Level Game

	CD	DD
CD	$R = V - K_{c1}$	$S = V(1 - .5^F) - K_{c1}$
DD	$T = V(1 - .5^F)$	$P = 0$

Note: The term in each cell is the *row* strategy's payoff. CD indicates cooperate at the first level and defect at the second level (i.e., contribute to the public good but exercise no normative control); DD indicates defect at first level and second level (i.e., no contribution to public good and exercise no normative control). R is the "reward" payoff, S is the "sucker" payoff, T is the temptation payoff, and P is the "punishment" payoff.

The Prisoner's Dilemma (#12). Consistent with its special place in the analysis of collective action, the Prisoner's Dilemma occupies the central region of the game-space diagram. It is named for a vignette in which two criminal suspects are questioned separately about a crime. Their interests derive from the *preference order* of the core game's payoffs. The preferred outcome is unilateral defection ("temptation," T), in which one benefits from confessing when the other remains quiet; then comes universal cooperation (the "reward," R) in which both remain quiet and receive light sentences; next comes universal defection ("punishment," P) in which both confess and are severely punished; and the worst is unilateral cooperation ("sucker," S), in which only the other confesses so one's own penalty is most harsh. The essential

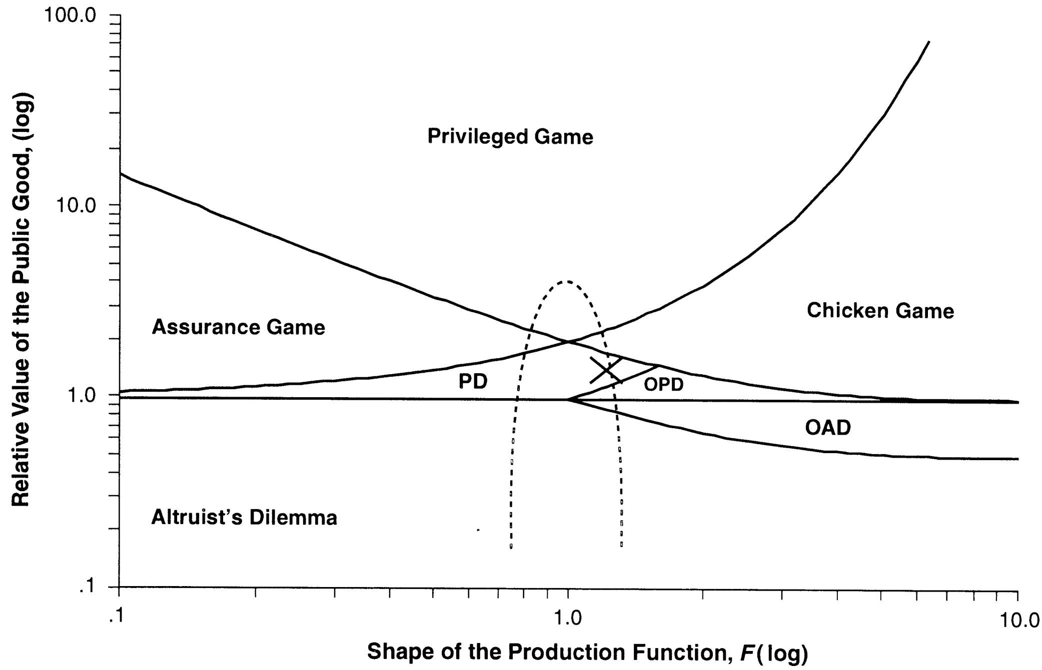


Figure 2. Game-Space Diagram Showing the Family of Games Generated by the Relationship between the Shape of the Production Function (F) and the Relative Value of the Public Good (V/K_{c1}).

problem is one of *trust*. If the prisoners can trust one another to act on their common interest in remaining quiet, they can escape with a light sentence. This game has become the paradigm for cases in which individually rational actions lead to a collectively irrational outcome. A vast amount of research on the Prisoner's Dilemma has focused on a single game described by the parameters $T = 5$, $R = 3$, $P = 1$, $S = 0$ (Axelrod 1984; Hirshleifer and Martinez Coll 1988). This corresponds to a game with a modestly valued collective good ($V/K_{c1} = 1.4$), and a modestly decelerating production function ($F = \ln(3/7)/\ln(1/2) = 1.222$). The location of this game in Figure 2 is indicated by an "x" in the Prisoner's Dilemma region.

A collective-action system corresponds to a PD game if universal cooperation is preferred to universal defection (e.g., a united labor union would succeed in gaining valuable concessions from management), yet the preferred outcome is unilateral defection (e.g., to strike is costly and bargaining remains effective with one less striker). Whereas in Olson's (1965) analysis the PD fits most collective-action systems, Klandermans (1988:85), in a study of union action, found that only 8 percent of respondents believed that a strike would be effective and considered their own participation to be superfluous. Most of these workers said they would not take part in a strike.

The PD region has two parts, labeled PD for the (true) "Prisoner's Dilemma" and OPD for the "Ordinal Prisoner's Dilemma." In the OPD, a defining characteristic of the PD is lacking. Though the order of payoffs is correct, players prefer alternating between unilateral defection and unilateral cooperation (i.e., a mixture of T and S) to mutual cooperation (R), so $(T + S)/2 > R$.

Figure 2 indicates that there is an upper limit of F beyond which a collective goods problem is no longer a true PD. This can be stated as follows:[2]

[2] The maximum value for F in a true PD occurs at the point where the line separating the PD region from the chicken region (mathematically, this is the line defined by $S = P$) intersects the line separating the PD region from the OPD region (mathematically, this is the line defined by $R = [T + S]/2$). Therefore, the maximum value of F can be found by solving for F at this intersec-

Theorem 1: *If any collective good's production function in a dyadic game is more than moderately decelerating (i.e., if $F > ln(1/3)/ln(1/2) = 1.58$), the system is not a true PD.*

Another restriction is that if the relative value of the collective good exceeds or falls below a critical amount, the game cannot be a PD. This leads to a second theorem:[3]

Theorem 2: *If the relative value of the collective good exceeds 2 (i.e., if $V/K_{c1} > 2$), or is less than 1 (i.e., if $V/K_{c1} < 1$), the system is not a PD.*

The collective good must be more valuable than its production cost, but its value cannot be more than double the production cost. These theorems place stringent bounds on the range of collective goods problems that correspond to PDs.

The chicken game (#66). A PD game is transformed into a game of chicken when the production function becomes more sharply decelerating (i.e., when F increases). In this game, the order of the two least valued payoffs is reversed from that in the PD (i.e., the new order is $T > R > S > P$; see Table 2). This reversal of preferences occurs because as the production function becomes more sharply decelerating, an additional cooperator produces an ever larger proportion of the collective good. The effect is to increase the relative value of the outcomes in which cooperation is partial, T and S. When the value of S increases until it exceeds the value of P, a PD is transformed into a chicken game.

Table 2. Payoff Matrices for Five Ordinally Distinct Games

Game and Strategy	Strategy: CD	DD
Prisoner's Dilemma		
CD	3, 3	0, 5
DD	5, 0	1, 1
Chicken Game		
CD	3, 3	1, 5
DD	5, 1	0, 0
Assurance Game		
CD	5, 5	0, 3
DD	3, 0	1, 1
Privileged Game		
CD	5, 5	1, 3
DD	3, 1	0, 0
Altruist's Dilemma		
CD	1, 1	0, 5
DD	5, 0	3, 3

This game is named for a contest in which teenage drivers test their courage by driving straight at one another. Each player chooses between two strategies: *chicken* (swerve to avoid a collision) or *daredevil* (do not swerve). Thus, the order of preferences is temptation (T), the other swerves; then reward (R), both swerve; then sucker (S), ego swerves; and the worst of all, punishment (P), a head-on collision. The essential problem in a chicken game is *bargaining*. Players have a common interest in avoiding conflict but have opposed interests regarding the terms of agreement, such as the allocation of courage, honor, or profit. This game fits systems in which a common interest in collective action coexists with opposed preferences regarding the precise direction that action

tion, that is, by solving for F on the assumption that $S = P$ and $R = (T + S)/2$. First, substitute the former expression into the latter, yielding $R = (T + P)/2$. Using the payoff equations in Table 1, this expands to $V - K_{c1} = (V * (1 - .5^F) + 0)/2$. With a bit of manipulation, this simplifies to $1/(1-.5^F) = 1/(2\,.5^F)$. Finally, solving for F yields $F = \ln(1/3)/\ln(1/2) = 1.58496250$. This conclusion is derived for a dyadic group. In larger groups, more sharply decelerating production curves are compatible with the PD. For example, in groups larger than 15, F can exceed 10.

[3] The maximum net value of the collective good (V/K_{c1}) in a true PD can be computed by solving for the net value at the intersection of the line separating the PD region from the assurance game region (the $R = T$ line), and the line separating the PD region from the chicken game region (the $S = P$ line), that is, solving for V/K_{c1} on the assumption that $S = P$ and $R = T$. One way to do this is to solve first for V/K_{c1} assuming that $S = P$. This yields $V/K_{c1} = 1/(1-.5^F)$. Second, solve for V/K_{c1} assuming $R = T$. This yields $V/K_{c1} = 1/.5^F$. Third, solve for F using the above two equations, $1/(1-.5^F) = 1/.5^F$ so, $F = .5$. Finally, substitute $F = .5$ into the initial equation to yield $V/K_{c1} = 2$.

As was the case for Theorem 1, constraints on the PD are less stringent in larger groups. The lower bound is defined by the line separating the PD region from the altruist's dilemma region (the $V/K_{c1} = 1$ line).

should take. Examples include the hawk/dove split that arises in many social movements, as when purists claim that pragmatists are selling out by forsaking the movement's essential goals, and pragmatists claim that purists' unwillingness to compromise will lead the movement into ruin. Resolutions of the chicken game bargaining problem that are grounded in game theory include resistance theory (Heckathorn 1980) and Rubinstein's (1982) model.

The assurance game (#61). An exactly opposite dynamic transforms a PD into an assurance game. When the production function for the collective good becomes more sharply accelerated (i.e., F becomes smaller), the order of the two most highly valued outcomes is the reverse from their positions in the PD, so that the new order is $R > T > P > S$ (Table 2). The effect is to reduce the relative value of the partial-cooperation outcomes, so T and S decline in value. When T falls below R, the PD becomes an assurance game.

The assurance game derives its name from the fact that each player can be motivated to cooperate by the mere assurance that the other will do the same. A collective-action system is an assurance game if participation with others is highly valued, there is consensus on the direction of collective action, and the only uncertainty is that individuals do not want to participate unless others will do the same. This fits Klandermans's (1988) analysis of union participants: "If most members believe that only few people will participate, it becomes a self-fulfilling prophecy" (p. 90) that produces a downward spiral of participation. Alternatively, in these systems greater participation is self-reinforcing. Hence, the essential problem is one of *coordination*.

The privileged game (#6). PDs can be transformed into other types of games by changing the relative valuation of the collective good, V/K_{c1} (Figure 2). If the value of the collective good is increased sufficiently, the result is a game sometimes labeled "spite" because it lacks conflict unless players are so competitive that they seek to minimize the other's payoff. When compared to the PD, the two most valued payoffs are reversed, as in the assurance game (Table 2). The two least valued payoffs are also reversed, as in the chicken game, so the order of payoffs is $R > T > S > P$. As the collective good increases in value (or, equivalently, when contribution becomes less costly), a point occurs where the incentive to defect disappears because the net loss in the value of the collective good produced after defecting would exceed the costs of contribution. This is the point at which Olson (1965) described a group as "privileged." Therefore, I call this the *privileged game*. Because individual and collective rationality coincide perfectly, there is no social dilemma. Hence a definitional issue arises: If collective goods provision is defined as inherently problematic, no collective good exists in this system. However, by the standard definition of collective goods as entailing both jointness and nonexcludability, the conclusion is that dilemmas are not inherent in all collective goods provision.

The altruist's dilemma game (#9). If the relative value of the collective good in a PD game is reduced sufficiently, it is eventually transformed into an altruist's dilemma game (Heckathorn 1991) (Figure 2). In this game, the order of the two middle payoffs is reversed from their positions in the PD so that $T > P > R > S$ (Table 2). This reversal occurs because the incentive to defect increases as the collective good loses value. When the collective good's value is less than the contribution cost (i.e., when $V/K_{c1} < 1$), universal defection (P) becomes preferable to universal cooperation (R). At that point, the PD becomes an altruist's dilemma.

The altruist's dilemma game has a unique characteristic. Played egoistically, there is no dilemma because everyone defects, which is both individually and collectively rational. However, if players are altruistic, they cooperate because that is what the other prefers they do. Thus, altruistic players fare poorly when compared to defectors, hence the name "altruist's dilemma." An example arose this winter in the holiday gift exchange system in one branch of my extended family. Senior relatives concluded that too much money was being spent on gifts among adults. Therefore, names were thrown in a hat and each adult drew the name of the recipient of the gift he or she would buy at year's end. The idea was that less would be spent on single gifts than on a myriad of gifts. More generally, the altruist's dilemma fits cases in which the cost of providing a public good exceeds its benefits, so its production is collectively irrational. Examples include spending "too much" on highways, environmental protection, crime control, or other public goods. Here the problem is exactly opposite that of the PD. In an altruist's dilemma, it is not too little social cooperation that creates a problem, but rather too much. This problem can arise not only among altruists, but also when selective incentives are employed to organize collective action, because actors can then compel one another to act altruistically (Heckathorn 1991).

These five games are theoretically exhaustive. *Theorem 3* states:[4]

Theorem 3: *Only five ordinally distinct games arise in collective-action systems: the Prisoner's Dilemma, the chicken game, the assurance game, the altruist's dilemma, and the privileged game.*

This proof is highly general, in that the assumptions on which it is based, as formalized in equations 1 through 6, are intended to capture the essential features of collective action.

The analysis of collective action should be free from the Prisoner's Dilemma paradigm. Analysis of the PD persists despite criticism (Mueller 1989:15–17) because the other types of games that arise in collective goods contexts have not been exhaustively specified. In addition to the trust problem arising in the PD, collective action also confronts the bargaining problem of the chicken game, the coordination problem of the assurance game, the overcooperation problem of the altruist's dilemma, and the absence of a problem in the privileged game. Hence, studies of collective action should explore the full range of possible games.

A DYNAMIC FORMAL MODEL

Most formal analyses of collective action have employed rational-choice models in which decision-making is forward-looking (Olson 1965; Marwell and Oliver 1993). That is, actors formulate expectations about the future payoffs associated with each alternative course of action. In simple rational-maximizing models, actors choose the alternative promising the highest payoff. In more complex models, satisficing or framing may

[4] I assume that parameter values are constrained as follows: $V > 0$, so the collective good is valued; $K_{c1} > 0$, so contribution entails a cost; and $F > 0$, so the level of collective good produced is a monotonically increasing function of the number of contributors. These assumptions and equations 3 through 6 suffice to generate three inequalities. First, $T > S$, because $T - S$ equals K_{c1} when $T - S$ is expanded using equations 4 and 5 and then simplified; K_{c1} is positive by assumption. Second, $T > P$, because $T - P$ reduces to $V\,(1 - .5^F)$; V is positive by assumption, and $(1 - .5^F)$ is positive because F is positive by assumption. Because the product of two positive numbers is itself positive, $V\,(1 - .5^F)$ is necessarily positive. Third, $R > S$, because $R - S$ reduces to $V * (.5^F)$, and both V and $.5^F$ are positive. Based on the first two inequalities ($T > S$ and $T > P$) and the assumption that all outcomes are differentially valued, it must be the case that either $T > S > P$, or $T > P > S$.

Assume now that the first case is true, so $T > S > P$. If so, only two values of the remaining term, R, are consistent with the third inequality, $R > S$. These are $R > T > S > P$ (the privileged game) and $T > R > S > P$ (the chicken game). Now assume that the second case is true, so $T > P > S$. If so, only three values of the R term are consistent with the third inequality, $R > S$. These are $R > T > P > S$ (the assurance game), $T > R > P > S$ (the Prisoner's Dilemma), and $T > P > R > S$ (the altruist's dilemma). This completes the demonstration of Theorem 3.

Theorem 3 is unaffected by group size because the three inequalities on which it is based are unaffected by group size. Furthermore, even if some other limiting assumptions are modified, the theorem is essentially unchanged. For example, if the assumption that contributing to collective action is costly is removed to permit the case in which it is costly *not* to contribute ($K_{c1} < 0$), a sixth type of game becomes possible: $R > S > T > P$. However, from a strategic standpoint, this game is essentially identical to the privileged game. (The case in which $K_{c1} = 0$ is excluded, because if that were the case T would equal S, and the game would not be ordinally distinct.)

Similarly, if $V < 0$ is permitted so that the public "good" is negatively valued, a seventh game is added: $P > T > S > R$. This is a conflict-free game with universal defection as the strongly stable equilibrium and remains so even when it is played altruistically. Hence, this game provides a plausible lower bound for the altruist's dilemma zone, for even altruists will not incur costs to produce a product that they and others do not value.

affect decisions, but behavior remains forward-looking because expectations about future events govern choices in the present.

In contrast, experiential learning models are backward-looking (Cross 1983; Macy 1990). They place weaker demands on the calculating capacities of actors and rely instead on memory. For example, in Bush-Mosteller (1964) learning models, actors begin by behaving randomly. They then adjust the relative probabilities of choosing each strategy based on the rewards each strategy has produced in the past. Thus, the actor's behavior is purposive and adaptive in that actions that have led to rewards in the past are more likely to be repeated, while actions that have led to losses are progressively abandoned. The resulting behavior may be difficult to distinguish from rational optimizing. For example, Cross (1983) showed that if the rate of learning is greater than the rate of environmental change, adaptive learning produces patterns of behavior that can always be described as though actors were maximizing expected utility. Similarly, Macy (1990, 1993) used a backward-looking decision model to replicate and extend results from forward-looking models of collective action.

Selectionist or, equivalently, evolutionary game models provide a third type of decision model in which actors adopting a range of strategies interact with one another (Hirshleifer 1982; Maynard Smith 1982; Axelrod 1984). Based on the resulting payoffs, the actors with the most successful strategies proliferate at the expense of the less successful. This process is then repeated, generation after generation, until the system either approaches stable equilibrium or cyclical variation. Biologists employ these approaches to model evolutionary selection. However, the selection process has also been interpreted as reflecting a process of *observational learning* in which actors compare their own outcomes to those of their peers, imitating peers who do best (Brown, Sanderson, and Michod 1982; Boyd and Richerson 1985). In essence, actors look around to see who is doing well and take as role models those who appear most successful. When interpreted in this manner, these models can be termed *sideways-looking models* of behavior.

Backward-looking and sideways-looking models are similar structurally in that both assume an adaptive learning process by which more successful strategies proliferate at the expense of less successful strategies. The difference is that in backward-looking models, competition among strategies occurs within each actor, while in sideways-looking models, competition occurs among actors. Sideways-looking models have a potential advantage over forward-looking and backward-looking models in that actors can benefit from any single actor's discovery.

Adaptive learning models may seem more realistic than forward-looking models because they do not require that actors have highly developed calculating capacities. However, learning models also have stringent requirements. For example, a sideways-looking model requires that actors compare their own level of success to those of other actors. Therefore, they must observe the actions of all others and the rewards they all receive—a difficult task in all but the smallest systems. Similarly, a backward-looking model requires that actors compare the success of their most recent actions to that of previous actions, so they must retain a record of innumerable past actions. Therefore, demands on memory are extensive. Thus, all three models place demands that can appear implausible in some situations. This occurs because purposive action requires information, and information is not always easily acquired.

A realistic model of social actors would no doubt synthesize forward-looking, backward-looking, and sideways-looking models of decision-making because everyone formulates expectations about the future, learns from his or her own experiences, and benefits from the experiences of others. Such a model might be based on the assumption that individuals choose among alternative mechanisms depending on the relative credibility and cost of the information required by each mechanism. However, considering such a model would exceed the scope of this paper. Instead, I employ a sideways-looking model for several reasons. First, it provides the opportunity to assess the robustness of conclusions about collective action that were derived using forward-looking and backward-looking models. Second, a sideways-looking

Table 3. Payoff Matrix for Seven Strategies

Strategy	Private Cooperation (CD)	Full Defection (DD)	Tit-for-Tat (TFT)	Full Cooperation (CC)	Hypocritical Cooperation (DC)	Compliant Opposition (CO)	Full Opposition (DO)
CD	R	S	R	R	S	R	S
DD	T	P	P	$E_{c2} R + (1-E_{c2}) T$	$E_{c2} S + (1-E_{c2}) P$	T	P
TFT	$R-K_{oc}$	$P-K_{oc}$	$R-K_{oc}$	$R-K_{oc}$	$E_{c2} S + (1-E_{c2}) P - K_{oc}$	$R-K_{oc}$	$P-K_{oc}$
CC	$R - K_{c2}$	$E_{c2} R + (1-E_{c2}) S - K_{c2}$	$R-K_{c2}$	$R-K_{c2}$	$E_{c2} R + (1-E_{c2}) S - K_{c2}$	$R-K_{c2}$	$E_{c2} (1-E_{o2}) R + [1-E_{c2} (1-E_{o2})] S - K_{c2}$
DC	$T-K_{c2}$	$E_{c2} T + (1-E_{c2}) P - K_{c2}$	$E_{c2} T + (1-E_{c2}) P - K_{c2}$	$E_{c2} R + (1-E_{c2}) T - K_{c2}$	$E_{c2}^2 R + (1-E_{c2})^2 P + E_{c2} (1-E_{c2}) S + E_{c2} (1-E_{c2}) T - K_{c2}$	$T-K_{c2}$	$E_{c2} (1-E_{o2}) T + [1-E_{c2} (1-E_{o2})] P - K_{c2}$
CO	$R-K_{o2}$	$S-K_{o2}$	$R-K_{o2}$	$R-K_{o2}$	$S-K_{o2}$	$R-K_{o2}$	$S-K_{o2}$
DO	$T-K_{o2}$	$P-K_{o2}$	$P-K_{o2}$	$E_{c2} (1-E_{o2}) R + [1-E_{c2} (1-E_{o2})] T - K_{o2}$	$E_{c2} (1-E_{o2}) S + [1-E_{c2} (1-E_{o2})] P - K_{o2}$	$T-K_{o2}$	$P-K_{o2}$

Note: The term in each cell is the *row* strategy's payoff. Shaded area indicates core game and outcomes.

model introduces a new and potentially powerful tool for the analysis of collective action. Finally, it provides the opportunity to link two previously isolated fields—evolutionary game theory and collective action.

I draw on Hirshleifer and Martinez Coll's (1988) procedure for modeling evolutionary games because it is mathematically more elegant and less demanding computationally than competing approaches, yet it produces generally equivalent results. This procedure assumes that players with conditional strategies respond *immediately* to the strategies of others. In essence, it avoids the unnecessary steps involved in recomputing the expected payoff each time two equivalent strategies interact. Therefore, using only a single play, this approach mimics the effects of many plays.

The first step in constructing such a model is to generate a matrix that indicates the payoff each strategy is awarded when it plays each strategy, including itself. Table 1 (p. 256) depicts a payoff matrix for the core game. This matrix must now be expanded to include all seven strategies to be modeled.

Expanding the Core Payoff Matrix: Tit-for-Tat and Second-Level Strategies

Table 3 presents an "expanded" payoff matrix. Like Table 1, each cell represents the *row* player's payoff when encountering the column player. The payoff matrix for the core game's two strategies (shaded) occupy the first two rows and columns. The five additional strategies produce a 7×7 matrix: tit-for-tat (TFT) and the four strategies that employ selective incentives. Consider TFT. The payoffs awarded to TFT from encounters with the core strategies are computed as follows: When TFT encounters CD, the private cooperation strategy, each contributes, so TFT receives a payoff of $R - K_{oc}$ (i.e., the reward payoff less TFT's information cost and CD's payoff is R). Similarly, when TFT encounters DD, the full defection strategy, the equilibrium outcome is for neither to contribute, so TFT's payoff is $P - K_{oc}$, and DD's payoff is P. Finally, when TFT encounters itself, each contributes and incurs costs of K_{oc}, so the payoff is $R - K_{oc}$. Up to this point, the construction of the expanded matrix has followed Hirshleifer and Martinez Coll (1988) except that they use numbers drawn from a particular numerical example instead of locating equations in the cells.

The CC and DC strategies in the matrix employ selective incentives to change the partner's behavior. In full cooperation, CC, the player cooperates at the first and second levels by contributing to the collective good and exercising compliant control; in hypocritical cooperation, DC, the player defects at the first level and cooperates at the second level by not contributing while exercising compliant control to compel others to contribute. Two parameters define compliant control. First, compliant control has a cost, K_{c2}, the cost of cooperation at the second level. This cost is deducted from the payoffs shown in rows for CC and DC of the matrix. I assume this cost can vary from 0 to any positive number. Second, compliant control is not "cheap talk." It has a direct effect on others' behavior to a degree determined by its efficacy, E_{c2}—the efficacy of cooperation at level two. In essence, this is the *power* of the actor to alter the behavior of the target. In this model, I assume that control is exercised through constraining the partner's opportunity to act. (For a discussion of the assumptions about interpersonal control underlying this model, see Heckathorn 1990:369.) The effect is to truncate the game matrix so that the target of control faces fewer choices. The efficacy of compliant control is defined formally as the proportional reduction in the opportunities of the target of control to defect, so E_{c2} can vary from 0 (no power) to 1 (complete power). For example, an actor's efficacy of compliant control of $E_{c2} = .75$ reduces others' opportunities to free-ride by 75 percent (e.g., an actor who had chosen to not contribute would now have only a $1 - .75 =$ 25 percent opportunity not to contribute). More generally, when the target of control has chosen not to contribute, the exercise of compliant control imposes a mixed strategy on the actor, so the target has an E_{c2} probability of contributing and a $1 - E_{c2}$ probability of not contributing.

The payoffs associated with the full cooperation strategy (CC) reflect these two parameters. When CC encounters CD, both cooperate and thereby earn R, but CC's payoff is reduced by K_{c2}. When CC encounters DD, the situation is more complex. Because DD chooses to not contribute, two outcomes are possible. There is an E_{c2} probability that CC's compliant control is successful, in which case CC is awarded a payoff of $R - K_{c2}$, and DD is awarded R. There is a $1 - E_{c2}$ probability that CC's compliant control fails, in which case CC is awarded a payoff of $S - K_{c2}$, and DD earns T. The payoffs shown in Table 3 reflect these two possible outcomes weighted by their probabilities (i.e., the expected utility of CC when encountering DD is $E_{c2}\ R + [1 - E_{c2}]\ S - K_{c2}$, and the expected utility of DD is $E_{c2}\ R +$ $[1 - E_{c2}]\ T$). The other payoffs to the CC strategy are computed in like fashion.

The hypocritical cooperation strategy (DC) is identical to the CC strategy except that the player does not contribute. A particularly complex outcome arises when two hypocritical cooperators meet because each chooses to not contribute, while seeking to compel the other to contribute. Therefore, four outcomes are possible. If the compliant control of both the actor and the target succeed, both contribute and the outcome is R. The probability of this occurring is E_{c2}^2, so the expected utility of this prospect is $R * (E_{c2}^2)$. Second, if the compliant control of both fails, neither contributes and the outcome is P. The probability of this occurring is $(1-E_{c2})^2$. Third, if the actor's control succeeds but the target's control fails, the outcome for the actor is T. The probability of this occurring is $E_{c2}\ (1 - E_{c2})$. Finally, if the actor's control fails while the target's control succeeds, the actor's outcome is S. This also has a probability of $E_{c2}\ (1 - E_{c2})$. When the expected utilities of each of the four possible outcomes are summed and the costs of exercising compliant control are deducted, the result is $E_{c2}^2\ R + (1 - E_{c2})^2\ P + E_{c2}\ (1 - E_{c2})\ S$ $+ E_{c2}\ (1 - E_{c2})\ T - K_{c2}$.

The payoffs associated with oppositional control (the CO and DO rows) reflect both the cost of exercising that control, K_{o2} (the cost of opposition at level two) and the efficacy of oppositional control, E_{o2}. The efficacy of oppositional control is defined as the proportional reduction in the efficacy of compliant control produced by the exercise of oppositional control (Heckathorn 1990). As defined, it can vary from 0 (no effect) to 1 (in which it neutralizes totally the effect of compliant control). For example, if the target actor is exercising compliant control with efficacy $E_{c2} = .8$, and the efficacy of oppositional control is $E_{o2} = .75$, the net efficacy of the target's compliant control is reduced from .8 to E_{c2} $(1 - E_{o2}) = .8 * (1 - .75) = .2$.

The compliant opposition strategy is less interesting because it is strictly dominated by the full opposition strategy in the class of games examined here. The compliant opposition strategy is included in this analysis only for purposes of logical completeness.

Modeling Evolutionary Change

Consider now how the population of strategies may change. Every evolutionary process begins with an initial population distribution of strategies. Each strategy i represents a specific proportion of the total population, p_i. Depending on the strategy's payoff relative to other strategies, its prevalence in the population changes by the amount Δp_i, so the new population proportion is $p_i + \Delta p_i$. Following Hirshleifer and Martinez Coll (1988), this term may be defined as

$$\Delta p_i = Z\, p_i\, (Y_i - M), \tag{7}$$

where Z is a sensitivity parameter (a constant representing the speed with which population distributions change in response to different payoffs), Y_i is strategy i's payoff from playing each strategy in the system, including itself, weighted by the prevalence of each strategy. Hence, the strategy's average payoff during the current generation is

$$Y = \sum_{j-1}^{N} p_j U_{ij}, \tag{8}$$

where p_j is the proportion strategy j in the population, U_{ij} is strategy i's utility from playing strategy j, and N is the number of strategies. Finally, the mean payoff for all strategies is

$$M = \sum_{j-1}^{N} p_j Y_j. \tag{9}$$

In this model, strategies change in prevalence over generations depending on their payoffs relative to other strategies in the populations. A strategy becomes more prevalent if it earns a payoff greater than the mean (i.e., if $Y_i > M$), and declines if it earns less than the average. The magnitude by which the strategy's prevalence expands or declines depends both on its current prevalence in the population, p_i, and on the "sensitivity" parameter Z.

DYNAMICS OF COLLECTIVE ACTION

I now apply the formal model to analyze the evolution of collective action in systems with production functions that range from sharply decelerating to sharply accelerating, and collective goods that vary from extraordinarily valuable to virtually worthless. Analysis begins with the theoretically simplest case in which collective action is organized only through voluntary cooperation.

Voluntary Cooperation

Marwell and Oliver (1993) assume that collective action is organized solely through voluntary cooperation. In terms of the above strategies, this means that individuals have only two options—private cooperation or full defection. Figure 3A shows what happens when these two strategies compete in the standard Prisoner's Dilemma. The horizontal axis shows the number of generations during which the system has evolved, and the vertical axis represents the proportional distribution of strategies. At the beginning, the two strategies are equally prevalent (.5). This produces slightly more than 50 percent production of the collective good (indicated by the bold line), because of the gently decelerating production function assumed in the standard PD game. As evolution proceeds, consistent with expectation full defection quickly drives private cooperation into extinction. Consequently, by the 500th generation, the level of collective goods production has declined to 0. This explains why Marwell and Oliver do not identify a free-rider problem in the PD game—everyone defects, so no one can free-ride on the contributions of others. Thus, exploitation is avoided, but at the cost of leaving the social dilemma unresolved.[5]

[5] For a copy of a computer program that implements the model presented here and graphs the results, send a blank disk along with a stamped, self-addressed diskette mailer to the author. The program requires an IBM-compatible computer with VGA graphics.

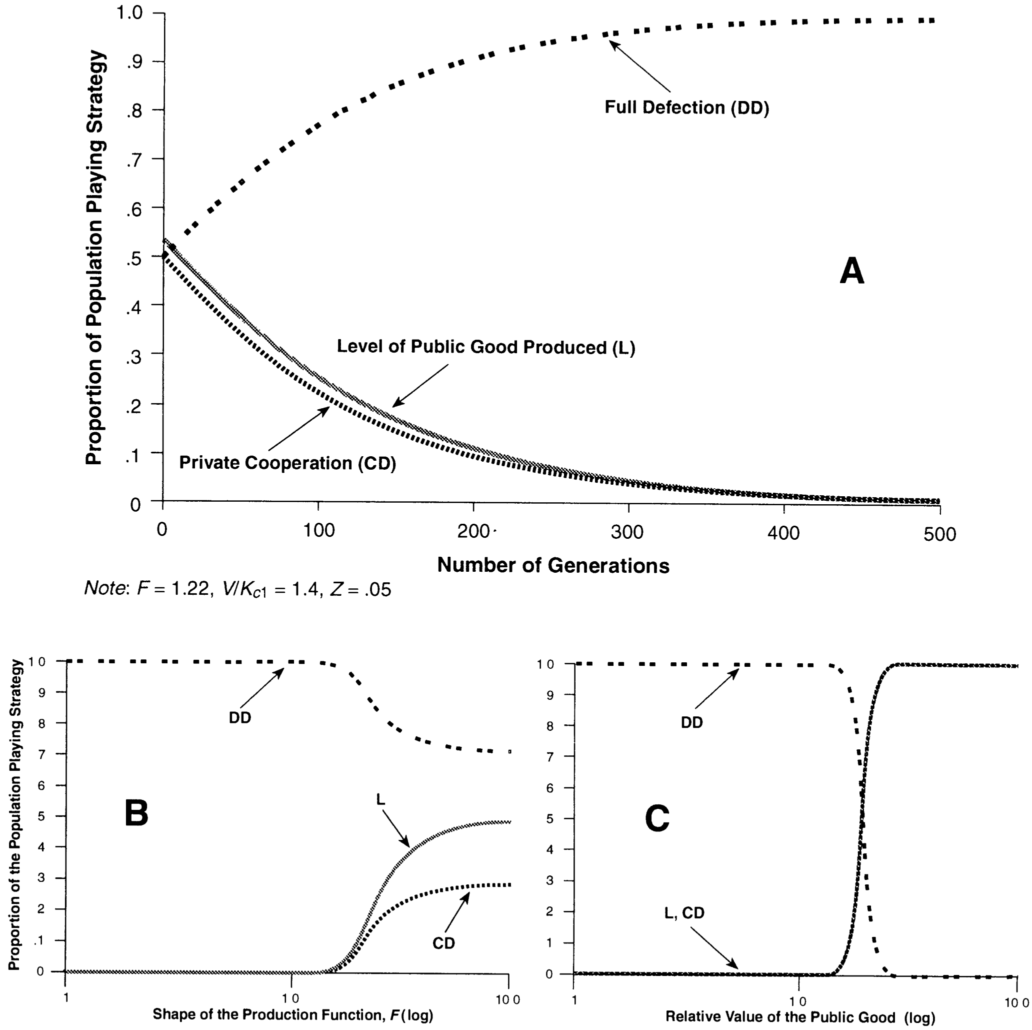


Figure 3. Proportion of the Population Playing Private Cooperation (CD) or Full Defection (DD), by Number o fgenreations in the Standard Prisoner's Dilemma, Shape of the Production function, and Relative Value of the Public Good (V/K_{c1})

Voluntary cooperation produces different results when the shape of the production function is changed (Figure 3B). Here the horizontal axis represents the shape of the production function, F; the vertical axis again represents the proportional distribution of strategies, but the lines now depict the distribution of strategies after 20,000 generations. In terms of Figure 2, Figure 3B portrays systems lying on a horizontal line intersecting the "x" that represents the standard PD. Figure 3B depicts several types of games. When F is large ($F \geq 1.9$), the system corresponds to a chicken game; when F is small ($F \leq .48$), the system corresponds to an assurance game; and when F is moderate ($.49 \leq F \leq 1.8$) the system corresponds to a PD like that portrayed in Figure 3A.

Private cooperation survives only in games with strongly decelerating production functions—the chicken games. Here the equilibrium outcome is a mixture of private cooperation and full defection strategies because instabilities arise when either private cooperation or full defection becomes too numerous. Consider again the vignette for which this game is named. When most players are "chickens" (private cooperators) who swerve to avoid a collision, the daring players who never swerve (full defectors) can thrive by intimidating the chickens. Therefore bold players proliferate. Yet this expansion is limited because bold players do poorly when confronting one another—they have head-on collisions. Hence, in a population of chickens, it is best to be daring; in a population of daring people, it is best to be a chicken. Thus, the equilibrium strategy in the chicken game is a mix of both types of players.

This dynamic corresponds to the free-rider problem described by Marwell and Oliver. Their "order effects" occur when one player makes a choice while observing another. For example, if two players are racing toward a confrontation and one sees the other surrender, there is no longer any pressure on the former to concede. Similarly, a chicken game is characterized by what they term a "surplus" because both players prefer swerving to a head-on collision, but only one of them needs to swerve to avoid the crash so there is a surplus of potential swervers. Thus, Marwell and Oliver are correct in associating surplus and order effects with decelerating production functions. However, the PD, the privileged game, and the altruist's dilemma can also have decelerating production functions, for these games also occur on the right half of Figure 2. Thus, a decelerating production function is a necessary but not a sufficient condition for surplus and order effects.

Consider now the left half of Figure 3B. Private cooperation fares poorly in these systems with accelerating production functions, which correspond to assurance games, but this is not invariably the case. Consistent with Marwell and Oliver's analysis of collective action in systems with accelerating production functions, collective action faces a start-up problem. For example, when $F = .1$, private cooperation can win if its initial prevalence is 75 percent or more, and when $F = .3$, private cooperation can win if its prevalence is 85 percent or more. Private cooperation usually loses in this system only because the initial prevalence of each strategy is assumed to be 50 percent, which is below these thresholds. Thus, Marwell and Oliver are correct in associating the start-up problem with an accelerating production function. However, not all systems with accelerating production functions are assurance games. There are also PDs, altruist's dilemmas, and privileged games on the left half of Figure 2.

Figure 3C shows the effect of changing the relative value of the collective good in the standard PD game portrayed in Figure 3A's. The horizontal axis represents the net value of the collective good, V/K_{c1}; the vertical axis again represents the proportional distribution of strategies, and the lines depict the equilibrium distribution of strategies attained after 20,000 generations. In terms of Figure 2, Figure 3C portrays the systems lying on a vertical line intersecting the "×" that represents the standard PD. These systems encompass several types of games. When the relative value of the collective good is low, the system corresponds to an altruist's dilemma ($V/K_{c1} < 1$). A slight increase produces an ordinal or true PD game ($1 < V/K_{c1} \leq 1.7$). A further increase yields a chicken game ($1.8 \leq V/K_{c1} \leq 2.3$). Subsequent increases produce a privileged game ($V/K_{c1} \geq 2.4$).

In Figure 3C's system, full defection wins in the altruist's dilemma because, just as in the PD, the payoffs from full defection strictly dominate those from private cooperation. Like Figure 3B, the outcome in the chicken game is a mixture of private cooperation and full defection. Finally, private cooperation wins in the privileged game because the payoffs from private cooperation strictly dominate those from full defection, so no social dilemma arises. Under this most favorable of circumstances, collective action emerges unburdened by free-riding. For example, Oegema and Klandermans (1994) analyzed a Dutch petition campaign against cruise missiles. Organizers sent signature cards to every postal address in the country and volunteers went door to door to collect them to ensure that signing the petition "required little or no effort." This reduced the cost of participation and thereby augmented the relative value of the movement's public good.

Strategic Interaction and Norms of Reciprocity

Early studies of the evolution of cooperation suggested that strategic interaction as embodied in the TFT strategy was extraordinarily robust (Axelrod 1984). Axelrod drew his conclusions from two celebrated tournaments in which TFT won against dozens of other strategies suggested by prominent game theorists. His analysis showed that these victories depended on two conditions: The prospects for future interaction must be substantial, and cooperative strategies must selectively interact with one another.

Figure 4A shows the effect of adding TFT to Figure 3A's two-strategy system. The results show that the TFT strategy can promote collective action. After about 700 generations, a mixed-strategy equilibrium is attained between TFT and private cooperation. TFT does not win out fully because the private cooperation strategy behaves, in essence, as a second-order free-rider (Martinez Coll and Hirshleifer 1991). That is, private cooperation reaps the benefit of TFT's suppression of the full defection strategy without bearing any of the potential costs.

TFT's ability to promote social cooperation is not limited to the PD. Figure 4B depicts the effect of altering the shape of the production function shown in Figure 4A. The effect can be seen by comparing Figure 3B, in which TFT is absent, with Figure 4B in which TFT is present. TFT has no effect in the chicken games ($F > 1.8$), in which the equilibrium mix of private cooperation and full defection remains unchanged. However, in the assurance games (in which F is small), and in the PD (in which F is moderate), a mixture of TFT and private cooperation produces full production of the collective good (i.e., $L = 1$).

Further effects of the introduction of TFT are apparent when variations in the value of the collective good are considered. Compare Figures 4C and 3C. The introduction of TFT has no effect in the altruist's dilemma—full defection dominates both TFT and private cooperation. However, TFT excludes all rivals in the PDs with the least-valued collective goods. In PDs with more highly valued collective goods, the payoffs to strategies that contribute to collective goods production increase relative to the payoff to full defection. That benefits both TFT and private cooperation. This trend continues as the collective good's value increases and the system is transformed first into the chicken game ($1.8 < V/K_{c1} < 2.3$), and finally into the privileged game ($V/K_{c1} > 2.4$).

Recent studies emphasize the fragility of TFT and associated reciprocity-based strategies (Martinez Coll and Hirshleifer 1991). One vulnerability concerns the effect of adding a cost of complexity (K_{oc}). Figure 4D shows the effect of adding that cost to Figure 4A's system—TFT dies out in about 2,000 generations. Consistent with Martinez Coll and Hirshleifer (1991), private cooperation serves as a second-order free-rider that dooms TFT to extinction. *Remarkably, this phenomenon occurs however small the cost of complexity, and not only in the PD but in all other games as well.*

Selective Incentives: Compliant Control

Recently, both evolutionary game models and collective action models have been expanded to include selective incentives. Evolutionary game theorists sought strategies more robust than TFT. For example, Boyd and Richerson (1992) explore punishment-based strategies. Consistent with the earlier analysis of Hirshleifer and Rasmusen (1989), they find that these strategies are so robust that they allow the evolution not only of cooperation, but of anything else—"moralistic strategies can cause any individually costly behavior to be evolutionarily stable, whether or not it creates a group benefit" (Boyd and Richerson 1992:173). Their analysis includes a moralistic strategy termed "cooperator-punishers" in which the actor cooperates and punishes all those who fail to cooperate. This coincides with the "full-cooperation" (CC) strategy of Heckathorn's (1989) collective action theory.

Figure 5A shows the effects of introducing the two strategies that embody compliant control, full cooperation and hypocritical cooperation, into the systems portrayed in Figure 4D. The effect is dramatic. Consistent with earlier research employing forward-looking models (Heckathorn 1990) and stochastic-learning models (Macy 1993), hypocritical cooperation is extraordinarily robust. In the standard game depicted in Figure 5A, hypocritical cooperation drives all competitors into extinction in about 1,000 generations. This outcome remains stable even if the efficacy of control is reduced rather dramatically (to $E_{c2} = .15$), or its cost is substantially increased (to $K_{c2} = .5$).

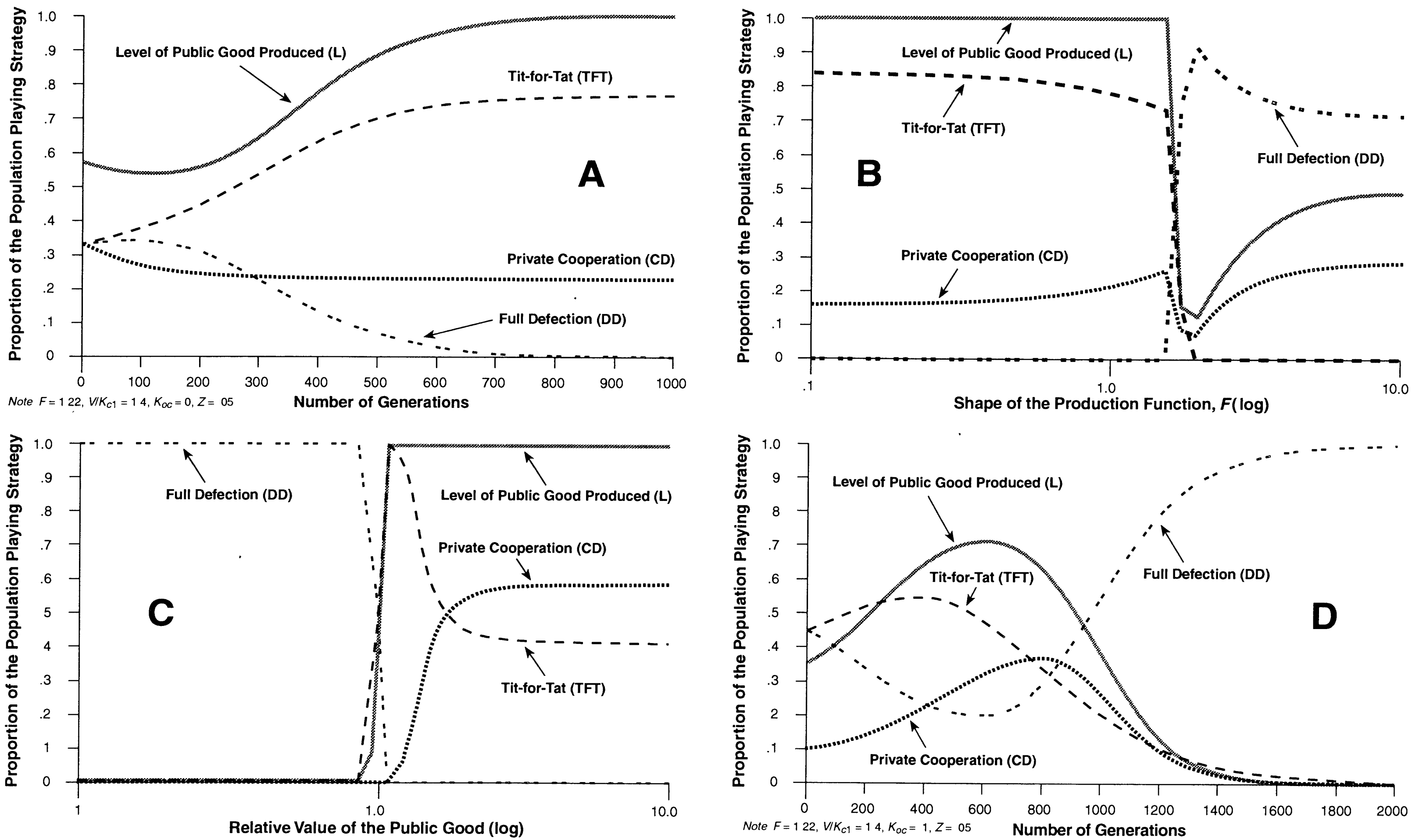


Figure 4. Proportion of the Population Playing Private Cooperation (CD), Full Defection (DD), and Tit-for-Tat (TFT), by Number of Generations, Shape of the Production Function, and Relative Value of the Public Good

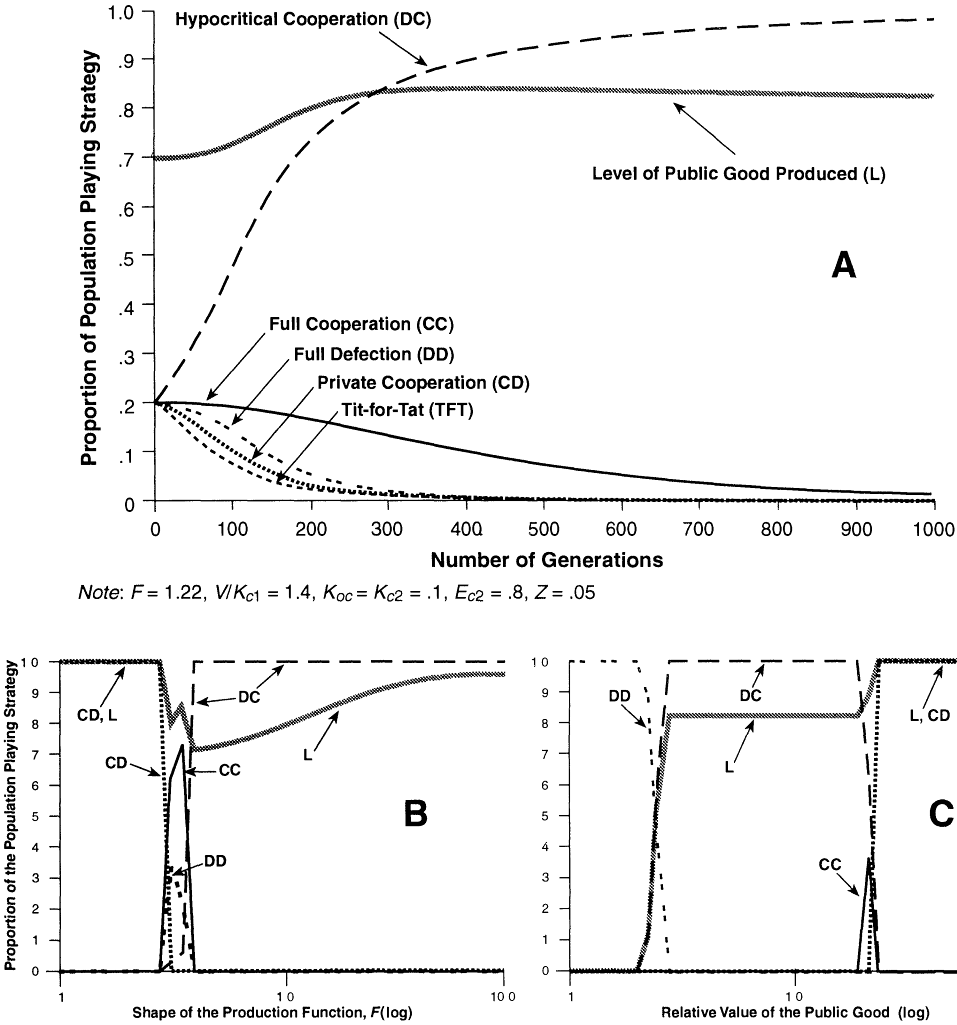


Figure 5. Proportion of the Population Playing Selected Strategies, by Number of Generations, Shape of the Production Function, and Relative Value of the Public Good: The Effect of Hypocritical Cooperation

The robustness of hypocritical cooperation becomes further apparent when variations in the shape of the production function are considered (Figure 5B). Hypocritical cooperation fails to survive only in the "strong" assurance games (i.e., the games in which $F \leq .28$), where it loses out to private cooperation. Moving from left to right in Figure 5B, there then occurs a brief transition zone of "moderate" assurance games ($.28 < F < .4$) characterized by a mix of hypocritical cooperation, full cooperation, and full defection. Finally, in the "weak" assurance games (F = .4 to .48) and in the subsequent PD and chicken games, hypocritical cooperation extinguishes all competitors.

Some of the strengths and limitations of hypocritical cooperation become apparent when variations in the value of the collective good are considered (Figure 5C). Hypocritical cooperation survives only when the collective good's value is neither too high nor too low. Consider first the altruist's dilemma games ($V/K_{c1} < 1$). Here it wins in all but the most extreme dilemmas ($V/K_{c1} < .31$). The result is an outcome termed "overcontrol" (Heckathorn 1990); in which a collective good is produced even though its value is less than its contribution cost. Thus, the group is worse off because of its production.

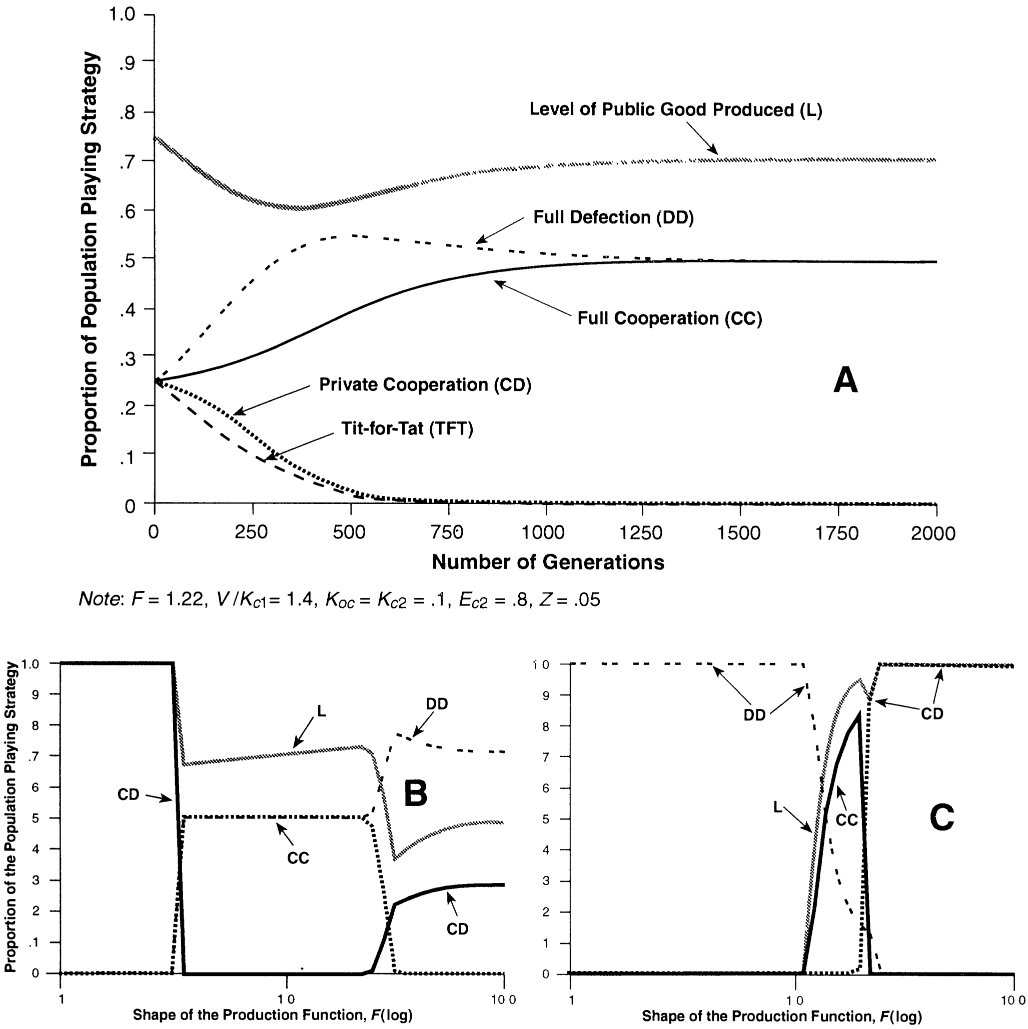


Figure 6. Proportion of the Population Playing Selected Strategies, by Number of Generations, Shape of the Production Function, and Relative Value of the Public Good: The Effect of Full Cooperation

In Figure 5C, hypocritical cooperation also dominates the PD games and the "weaker" chicken games. In the "stronger" chicken games, the outcome is a mix with full cooperation. Finally, in the privileged game, hypocritical cooperation loses out to private cooperation, as must any strategy that incurs second-level costs such as TFT's cost of complexity or a normative strategy's costs of compliant control.

Although full cooperation is usually overwhelmed by hypocritical cooperation, full cooperation is quite robust in its own right (Figure 6). It does well in competition with the voluntary cooperation and strategic interaction strategies. Full cooperation thrives under moderate circumstances, that is, when the production function is neither sharply accelerating nor decelerating and the value of the collective good is moderate. Thus, it functions like a more robust TFT. Moreover, even under circumstances in which full cooperation cannot survive, it sometimes leaves an enduring legacy. Specifically, although it dies out when $F < .3$, it first solves private cooperation's start-up problem, thereby allowing that strategy to win (compare Figures 3B and 6B when $F < .3$). This shows that relationships among strategies need not be competitive—they can also be cooperative.

A similarly cooperative relationship exists in strong assurance games between private cooperation and hypocritical cooperation (compare Figures 3B and 5B when $F < .3$); and between private cooperation and TFT when TFT is burdened by a cost of complexity.

Selective Incentives: Oppositional Strategies

When social movements are successful, they may overreach by overproducing their collective good. An example occurs in Figure 5C, in which hypocritical cooperation produces a high level of the public good in altruist's dilemma games. That example fits Boyd and Richerson's (1992) observation that punishment-based strategies can cause behavior to become stable when these strategies are individually costly and confer no group benefit. Examples of the overproduction of social control include the suffocating conformity of small-town life depicted by Mark Twain and Sinclair Lewis, and the problem of "overcompliance" addressed in the organization literature (Heckathorn 1991). Such overproduction may provoke countermobilization because *reducing* the level of production then constitutes a pubic good. Countermobilization is a phenomenon that, according to Klandermans (1994:372), has been neglected in the collective action literature.

Superoptimal levels of social cooperation open a niche for *oppositional strategies* that attack these excessively controlling strategies and restore a more optimal level of collective action. This process is illustrated in Figure 7A, which depicts an altruist's dilemma ($F = 1.22$, $V/K_{c1} = .7$). Hypocritical cooperation leaps upward during the first several hundred generations, matched by full opposition. Full opposition continues to increase in representation until the level of collective goods production has been driven down to about 10 percent. The equilibrium outcome is a mix of full opposition and hypocritical cooperation. In essence, the group polarizes into pro- and anti-collective goods production subgroups.

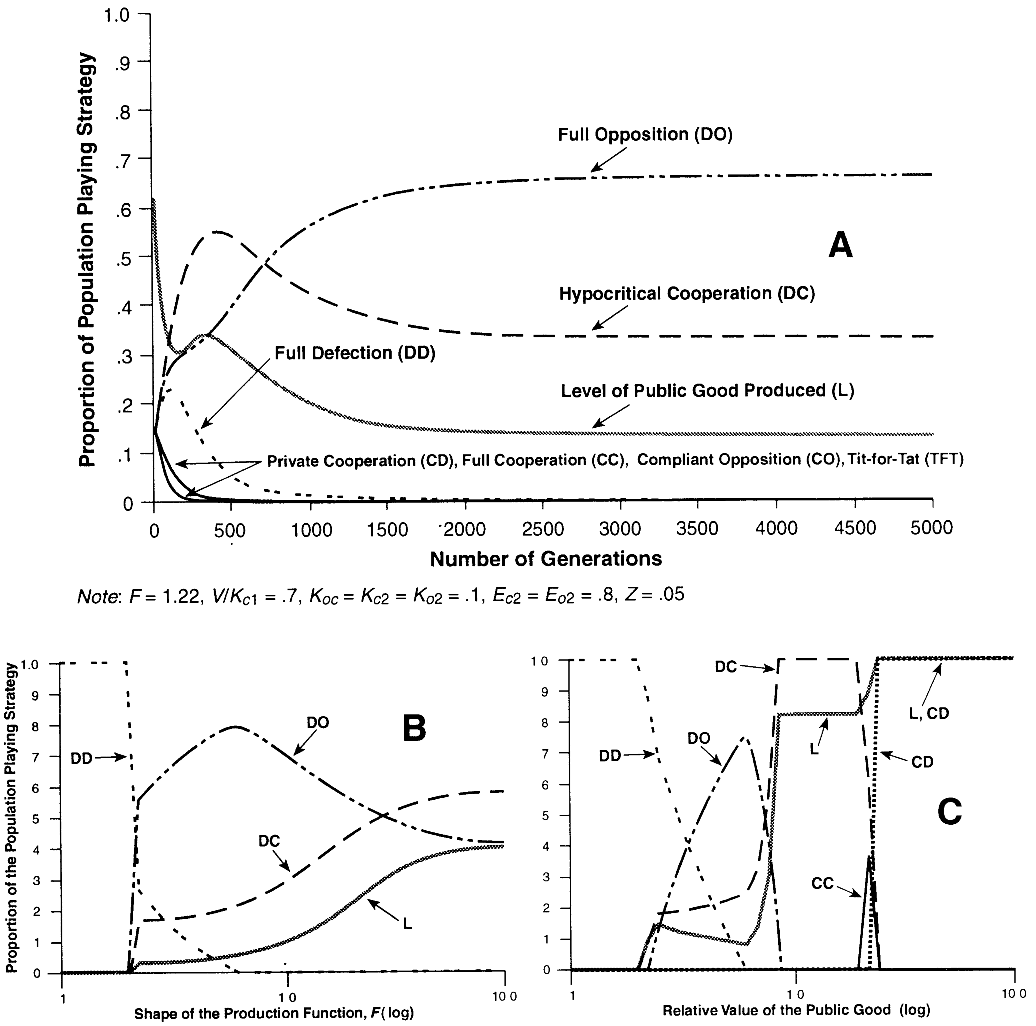
Figure 7B shows the effect of varying the shape of the production function in this altruist's dilemma game. The conflict between hypocritical cooperation and full opposition recurs in all systems except those with the most sharply accelerating production functions ($F < .22$). In these latter systems, hypocritical cooperation fails to arise, hence full opposition also cannot gain a purchase, so full defection wins. This failure of hypocritical cooperation reflects a limitation on compliant control that was identified in analyses using a forward-looking model (Heckathorn 1989). When the production function is sharply accelerating, almost unanimous contributions are required before any significant amount of collective good can be produced (e.g., when $F = .1$, a 50 percent contribution level would produce only 7 percent of the collective good). Hence, a more sharply accelerating production function progressively weakens incentives to use less-than-perfect means of control. So long as compliant control is costly and less than perfect, its use becomes counterproductive if the production function for the collective good is sufficiently sharply accelerating.

Figure 7C reveals further limits to oppositional strategies. These strategies thrive only in the "weaker" altruist's dilemmas ($.2 \leq V/K_{c1} < 1$). In the "stronger" altruist's dilemmas, the value of the collective good is so low that no compliant control emerges (see Figure 5C), so there is nothing for an oppositional strategy to oppose.

Oppositional strategies occupy a special niche: They retreat unless a collective good is being overproduced. Hence, they thrive only in the altruist's dilemma region when a moralistic strategy has produced a collective good with a negative net value.

In combination, moralistic strategies and oppositional strategies function as a system of checks and balances, with moralistic strategies stepping in to resolve free-rider problems and oppositional strategies emerging to hold moralistic strategies in check when collective goods production becomes excessive. The implication is that a complete system of selective incentives must include both moralistic and oppositional strategies.

Like any formal model, the model proposed here embodies many simplifying assumptions. Therefore, the question inevitably arises as to the robustness of its results. A definitive resolution of this question would require more extensive analysis. One prom-



Note: $F = 1.22$, $V/K_{C1} = .7$, $K_{OC} = K_{C2} = K_{O2} = .1$, $E_{C2} = E_{O2} = .8$, $Z = .05$

Figure 7. Proportion of the Population Playing Selected Strategies, by Number of Generations, Shape of the Production Function, and Relative Value of the Public Good: The Effect of Oppositional Strategies in the Altruist's Dilemma

ising sign, however, is the convergence of many of the model's conclusions with those based on forward-looking or backward-looking models.[6]

[6] This convergence occurs, in essence, because all three types of models view behavior as purposive, in that adaptive expectations govern behavior. The models differ only in the nature of these expectations—whether the future is calculable, will be like the past, or is best understood by those who are most successful. In informationally rich settings, the predictions of these models can be expected to converge, as was shown by Cross (1983) in the case of forward-looking and backward-looking models. The convergence of predictions therefore suggests that this model's limiting assumptions are no less adequate to represent core aspects of collective action than are those employed in other formal models of collective action.

THREE PHASES OF COLLECTIVE ACTION

Except in the smallest or most highly centralized systems, collective action is organized incrementally. During the *initial phase*, either "zealots" (Coleman 1990:490) or a "critical mass" (Marwell and Oliver 1993) make the initial contributions. During the *intermediate phase*, the ranks of contributing actors continue to grow because of greater marginal returns to contributors, strategic interaction, or the operation of selective sanctions. Finally, during the *mature phase*, the limits of collective good production are approached because of physical constraints on further production or a dwindling pool of additional contributors.

During each phase, promoters of collective action face a distinctive set of problems because of changes in the shape of that segment of the production function on which they operate. That is, each subgroup operates within a different *local portion* of the group's global production function. Realistic production functions typically are S-shaped, as illustrated in Figure 1A. Therefore, the local coalition that first seeks to initiate production faces only the leftmost portion of the curve where the function is accelerating ($F < 1$). During the intermediate phase, local groups face the middle part of the production function where the function is approximately linear ($F \approx 1$). Finally, in the mature phase, local groups face the rightmost portion where the production function is decelerating ($F > 1$). Thus, the process of collective goods production entails a progressive deceleration of the local production function faced by each successive group (F increases). This corresponds to movement from left to right in Figure 2's diagram of the game-space.

The process of collective action also entails changes in the value of what may be termed the "local collective good" that each subgroup potentially produces. This is the portion of the global collective good that can be produced by the subgroup, so it reflects the marginal gains attainable through contributions by the local group's members. During the start-up phase, the value of the "local" collective good is low because initially the slope of the production function is almost flat. During the intermediate phase, marginal gains increase, and so too does the value of the local collective good. Finally, during the mature phase, the slope decreases, reflecting diminishing marginal returns, so the value of the local collective good declines. Thus, organization of collective action is characterized by an increase in the value of the local collective good followed by a decline.

In combination, these changes in the shape of the production function and in the value of the collective good cause collective action, from its initial phase to its mature phase, to traverse an arc-shaped path through the game-space diagram.[7] Figure 2 shows such an arc corresponding to Figure 1A's production function. Here the system is assumed to contain a total of 25 actors with a collective good whose relative value is $V/K_{c1} = 20$, in which the subcoalition size is 2. In this system, the start-up phase begins in the altruist's dilemma zone with an accelerating production function because the marginal returns to the initial contributors are low and the local production function is accelerating. As more actors contribute, the system moves up and to the right, entering the PD zone, then the assurance game zone, then the privileged game zone. Eventually the trajectory moves down and to the right, entering the chicken game zone, then reentering the PD zone, and finally returning to the altruist's dilemma zone. This shows graphically both the complexity of the process of collective action and the importance of recognizing the distinctive types of strategic problems that arise along the way.

---

[7] Procedures for computing these arcs are straightforward. The first step is to specify the formula for the S-shaped production function. For example, Macy (1993:823) used a representative formula,

$$L_i = \frac{1}{1+e^{(.5-P_i)10}}, \qquad \text{(A)}$$

where P_i is the proportion of contributors in the group, and L_i is the level of collective good produced. This is the formula depicted in Figure 1A.

The second step is to identify the segment of the production function along which each subcoalition plays its local game. In the arc depicted in Figure 2, interaction is assumed to occur in pair-wise fashion, so subcoalitions have a size of 2. Therefore, when collective action is beginning, the first two actors can, through their collective decisions, produce a proportion of cooperators of 0 (if both defect), .04 (if only one contributes), or .08 (if both contribute). Therefore, the range from 0 to .08 is the interval along which they play their game.

The third step is to determine the marginal value of the local collective good. This gives the y-coordinate of the point in the game-space that corresponds to the subcoalition's game. The minimum amount of collective good that can be produced is found by substituting the minimum proportion of contributors (e.g., 0/25 = 0) into equa-

The shape and location of the curve are not immutable. Depending on how collective action is organized, the curve can be shifted up into areas of the game-space that favor the emergence of collective action, or down into areas that impede its development. Hence, there is a second option for organizing collective action: Rather than attempting to resolve the problems of trust, bargaining, coordination, or overcooperation that exist within the current game, actors can seek to *change that game*. For example, if the arc can be shifted upward into the privileged game zone, the success of collective action is guaranteed.

The arc can be shifted in three distinct ways. First, the value of the collective good can be altered. An increase in the value of the collective good moves the arc upward. However, it must be increased dramatically to escape the altruist's dilemma zone. For example, in Figure 2's system, V/K_{c1} must increase from 20 to 125. Groups sometimes do not have the capacity to augment directly the value of their collective goods. However, if those contributors who are first recruited are also those who most highly value the collective good, the effect is the same (Heckathorn 1993; Marwell and Oliver 1993).

A second approach is to reduce the size of the group. If the size of the group is reduced, the potential difficulties of organizing collective action are correspondingly decreased (Olson 1965), and the arc is moved upward. This also causes the curve to expand horizontally, thereby increasing the range through which the shape of the production function changes. Whether this approach is available depends in part on the distribution of resources among the potential contributors (Heckathorn 1993). If a subset of potential contributors has sufficient resources to produce the collective good, that effectively reduces the group's size. However, the group size must be reduced dramatically to escape the altruist's dilemma zone—in Figure 2's arc, group size must decrease from 25 to 9.

A third factor is the size of the subcoalitions through which collective action is incrementally organized. For example, consistent with my assumption that interaction is pair-wise, the arcs were computed assuming a subcoalition size of 2. That is, each additional contribution stemmed from a strategic interaction between two potential contributors. However, if the subcoalition size is increased, the effect resembles a decrease in group size because the arc moves upward and the legs move outward. Therefore, start-up problems can be eased. Whether this approach is feasible depends on whether there are preexisting relationships that allow contributors to motivate others to contribute. Start-up costs are more easily absorbed when potential contributors can also guarantee the participation of their followers. An example of this approach is Broadhead and Hecka-

tion A, which yields a production level of only slightly more than 0, (i.e., Min(L) = .0067). The maximum amount of collective good that can be produced is found by substituting the maximum proportion of contributors (e.g., 2/25 = .08) into equation A, which yields a level of Max(L) = .0148. When the difference between these two levels (.0148 − .0067 = .0081) is multiplied by the value of the collective good (20), the result gives the value of the local collective good that can be produced by the subcoalition (i.e., 20 * .0081 = .162).

The final step is to determine the value of the exponent, F. This gives the x-coordinate of the subcoalition's location in game-space. Each segment of the S-shaped production function can be approximated by the simpler function defined by equation 2 and depicted in Figure 1B. The aim is to match the extent to which the segment is convex, linear, or concave. This can be done by considering three points on the segment—the two endpoints, Max(L) and Min(L), and the midpoint between them, Mid(L). In the above example, when P_i = 1/25 = .04 of the players contribute, from equation A, the production level is Mid(L) = .01. If equation 2 is solved for F, an expression can be derived expressing F as a function of these three points:

$$F = \frac{\ln\left(\frac{\text{Max}(L)-\text{Mid}(L)}{\text{Max}(L)-\text{Min}(L)}\right)}{\ln(.5)}. \tag{B}$$

For example, when .0148, .01, and .0067 are substituted for Max(L), Mid(L), and Min(L) respectively, in equation B, F = .745.

The same procedure can be used to identify the subgames played by subsequent additions to the set of contributing actors. For example, when the subcoalition consists of potential contributors 11 and 12, they play along the P_i = .4 to P_i = .48 segment of the production function, a subgame with coordinates V/K_{c1} = 3.624 and F = .9193.

thorn's (1994) AIDS prevention intervention in which "augmenting the cohesion of high-risk groups [will] . . . increase their capacity for collective action and facilitate the emergence of AIDS prevention practices and norms" (p. 485).

As interpreted from the standpoint of the proposed model, these options provide the focus for some who are most critical of applications of the PD paradigm to collective action. For example, Fireman and Gamson (1979) conclude that collective action research should focus on three factors: ". . . how organizers raise consciousness of common interests, develop opportunities for collective action, and tap constituents' solidarity and principals" (p. 36). As interpreted using the proposed model, this corresponds to increases in the value of the public good (V), reductions in the cost of participation (K_{c1}), which they term a "search for more efficient ways of bringing about the collective good" (p. 33), and increases in subcoalition size—all of which shift the arc upward into more favorable regions of the game-space.

After means for shifting the arc into more favorable regions have been exhausted, as inevitably they must, the success of collective action depends on whether the organizational capacity of the group suffices to resolve any remaining dilemmas. If only easy dilemmas remain (i.e., dilemmas located high in game-space), even poorly organized groups can act collectively. When the dilemmas are more difficult, only better organized groups are successful.

Fireman and Gamson's (1979) argument that potential social dilemmas in collective action are resolved through appeals to identity and building group solidarity suggests that the arc typically shifts so far upward that it enters the privileged game zone. From the standpoint of the proposed model, this appears dubious, for it would require that collective goods always be of great value, that groups be tiny, or that groups be highly cohesive. Furthermore, from the fact that collective action traverses an arc-shaped path, it follows that the initiation process is most problematic. Except under highly favorable conditions, initiators face an altruist's dilemma (i.e., the legs of the arc extend downward into the altruist's dilemma region). This reflects the fact that their efforts will yield a net loss unless their contributions are subsequently augmented by other contributions.

The analysis shows that the range of strategies capable of initiating collective action when facing an altruist's dilemma is severely limited. TFT does not suffice, nor does full cooperation, because both always lose out to full defection. Among the strategies considered, only hypocritical cooperation can initiate cooperation in an altruist's dilemma (Figure 5A). Yet even here there is a potential problem because hypocritical cooperation can be neutralized by oppositional control (Figure 7A).

If oppositional strategies emerge too quickly, collective action is stillborn. Hence, the initiation of collective action must be relatively rapid if it is to avoid being crushed by oppositional control. Oppositional control necessarily emerges *in reaction to compliant control*, so it always emerges *after* compliant control. Furthermore, organizing oppositional control tends to be more difficult than organizing compliant control because oppositional control confronts a *coalition* of normative controllers, whereas compliant control confronts noncooperative individuals (Heckathorn 1990). Both these factors tend to provide a breathing space for compliant control. If collective action is initiated sufficiently quickly, it can move upward along the arc to escape the altruist's dilemma region before it is neutralized by oppositional control. In contrast, if collective action bogs down during the initiation phase, the organizing coalition may fragment. For example, the first states passed the ERA with little controversy, then opposition mounted. Had the ERA's supporters been able to act more quickly, they might have succeeded.

Oppositional control plays a different role during the mature phase of collective action. According to the proposed model, successful social movements are not self-limiting. In terms of Figure 2, when the group traverses the entire length of the arc, it eventually reaches the rightmost leg and reenters the altruist's dilemma region. That reflects diminishing marginal returns so severe that each new contribution yields a net loss. The collective good is then overproduced. It is then that oppositional control reappears to limit further collective goods production. If compliant control and oppositional control are well-matched in terms of cost and efficacy and the group is homogeneous, the effect is to stabilize production near the boundary between the PD and the altruist's dilemma regions of game-space on the right leg of the arc. That produces an approximately optimal level of collective goods production. However, if compliant control and oppositional control are not well-matched or the group is heterogeneous, the level may be far from optimum. For example, if valuations of the collective good are heterogeneous and production is organized through selective incentives, countermobilization may begin early, long before an optimal production level is attained (see Heckathorn 1993:343). Similarly, in the case of AIDS prevention, ". . . obstacles to containing the AIDS epidemic included the difficulty of mobilizing latent high-risk groups, and overcoming the high level of mobilization exerted by low-risk groups whose moral/political agenda conflicted with effective AIDS prevention" (Broadhead and Heckathorn 1994:475). As a result, levels of collective action to combat AIDS were initially quite low. Alternatively, if oppositional control is costly, countermobilization may begin late, only after overproduction has become substantial. In either case, mature systems of collective action can be identified qualitatively by a stalemate attained between proponents of greater production and lesser production of the collective good. Examples of such stalemates abound. For example, the environmental movement made great strides during the 1970s and early 1980s, but powerful coalitions are now deployed on both sides of environmental issues. Similarly, the movement to limit speech that is insensitive to issues of race, ethnicity, class, and gender achieved much success during the 1980s, but is now confronted by a reaction against "political correctness" in the mid-1990s. Substantial opposing coalitions are currently fighting either to expand or restrict defense expenditures, welfare benefits, consumer protection regulations, interstate highways, large-scale power plants, antitrust provisions, the rights of criminal suspects, and regulations protecting collective bargaining.

CONCLUSION

Theories of collective action have become more divergent because of the multifaceted character of the phenomenon. As in the fable involving the blind men and the elephant, different groups of analysts have focused on distinct forms of collective action. Some view collective action as N-person Prisoner's Dilemmas dominated by free-rider problems. Others identify different dilemmas that render social action problematic. Finally, some deny that collective action involves any form of social dilemma. According to my analysis, all three positions are correct when the analyses are viewed in their appropriate contexts.

Douglas Heckathorn *is Professor of Sociology at the University of Connecticut and Co-Director of the Eastern Connecticut Health Outreach Project. This paper grew out of his long-standing interest in developing formal theories of norms and collective action. He is currently collaborating with Robert Broadhead in efforts to apply these ideas to increasing the access to primary medical care by persons infected with HIV, and to preventing AIDS among drug injectors. Another recent work focuses on political rhetoric and public policy. He is currently completing a book on collective action.*

REFERENCES

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

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

———. 1992. "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups." *Ethology and Sociobiology* 13:71–195.

Broadhead, Robert S. and Douglas D. Heckathorn. 1994. "AIDS Prevention Outreach Among Injection Drug Users: Agency Problems and New Approaches." *Social Problems* 41:473–95.

Brown, Joel S., Michael J. Sanderson, and Richard E. Michod. 1982. "Evolution of Social Behavior by Reciprocation." *Journal of Theoretical Biology* 99:319–39.

Bush, Robert R. and Frederick Mosteller. 1964. *Stochastic Models of Learning*. New York: Wiley.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap Press.

Cross, John G. 1983. *A Theory of Adaptive Eco-*

nomic Behavior. Cambridge, England: Cambridge University Press.

Ferree, Myra Marx. 1992. "The Political Context of Rationality: Rational Choice Theory and Resource Mobilization." Pp. 29–52 in *Frontiers in Social Movement Theory*, edited by A. Morris and C. Mueller. New Haven, CT: Yale University Press.

Fireman, Bruce and William A. Gamson. 1979. "Utilitarian Logic in the Resource Mobilization Perspective." Pp. 8–45 in *The Dynamics of Social Movements*, edited by M. N. Zald and J. D. McCarthy. Cambridge, MA: Winthrop.

Hardin, Russell. 1971. "Collective Action as an Agreeable N-Prisoners' Dilemma." *Behavioral Science* 16:472–81.

Heckathorn, Douglas D. 1980. "A Unified Model for Bargaining and Conflict." *Behavioral Science* 25:261–84.

———. 1989. "Collective Action and the Second-Order Free-Rider Problem." *Rationality and Society* 1:78–100.

———. 1990. "Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control." *American Sociological Review* 55:366–84.

———. 1991. "Extensions of the Prisoner's Dilemma Paradigm: The Altruist's Dilemma and Group Solidarity." *Sociological Theory* 9:34–52.

———. 1993. "Collective Action and Group Heterogeneity: Voluntary Provision Versus Selective Incentives." *American Sociological Review* 58:329–50.

Hirshleifer, David and E. Rasmusen. 1989. "Cooperation in a Repeated Prisoner's Dilemma with Ostracism." *Journal of Economic Behavior and Organization* 12:87–106.

Hirshleifer, Jack. 1982. "Evolutionary Models in Economics and Law: Cooperation Versus Conflict Strategies." *Research in Law and Economics* 4:1–60.

Hirshleifer, Jack and Juan Carlos Martinez Coll. 1988. "What Strategies Can Support the Evolutionary Emergence of Cooperation?" *Journal of Conflict Resolution* 32:367–98.

Klandermans, Bert. 1988. "Union Action and the Free-Rider Dilemma." Pp. 77–92 in *Research in Social Movements, Conflict and Change*, vol. 10, *Social Movements as a Factor of Change in the Contemporary World*, edited by L. Kriesberg and B. Misztal. Greenwich, CT: JAI Press.

———. 1994. "Targeting the Critical Mass." *Social Psychology Quarterly* 57:368–72.

Macy, Michael W. 1990. "Learning Theory and the Logic of Critical Mass." *American Sociological Review* 55:809–26.

———. 1993. "Backward-Looking Social Control." *American Sociological Review* 58:819–37.

Martinez Coll, Juan Carlos, and Jack Hirshleifer. 1991. "The Limits of Reciprocity." *Rationality and Society* 3:35–64.

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

Maynard Smith, John. 1982. *Evolution and the Theory of Games*. Cambridge, England: Cambridge University Press.

Mueller, Dennis C. 1989. *Public Choice II: A Revised Edition of Public Choice*. Cambridge, England: Cambridge University Press.

Oberschall, Anthony. 1973. *Social Conflict and Social Movements*. Englewood Cliffs, NJ: Prentice Hall.

Oegema, Dirk and Bert Klandermans. 1994. "Why Social Movement Sympathizers Don't Participate: Erosion and Nonconversion of Support." *American Sociological Review* 59: 703–22.

Oliver, Pamela E. 1980. "Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations." *American Journal of Sociology* 85:1356–75.

Oliver, Pamela E., Gerald Marwell, and Ruy Teixeira. 1985. "A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action." *American Journal of Sociology* 91:522–56.

Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Opp, Karl-Dieter. 1989. *The Rationality of Political Protest*. Boulder, CO: Westview.

Rapoport, Anatol, Melvin J. Guyer, and David G. Gordon. 1976. *The 2x2 Game*. Ann Arbor, MI: University of Michigan Press.

Rasmusen, Eric. 1989. *Games and Information*. Cambridge, MA: Blackwell.

Rubinstein, Ariel. 1982. "Perfect Equilibrium in a Bargaining Model." *Econometrica* 50:97–109.