Y773
Professor Elinor Ostrom
Spring 1991

The Evolution of Institution: An Evolutionary Game Theory Approach

by
Myungsuk Lee

Introduction

For many years, it has been accepted by many scholars that rational and self-interested individuals will not act to achieve their common or group interests because the benefit of collective action is non-excludable (Olson 1965, Hardin 1968). This is so-called 'free-rider' hypothesis. Whenever one person cannot be excluded from the benefits that others provide, according to this hypothesis, each person is motivated not to contribute to the joint effort, but to free-ride on the efforts of others. If all participants choose to free-ride, the collective benefit will not be produced. Based upon this, it is predicted that participants cannot escape from this 'trap' by themselves since the bonds of words are too weak to bridle man's ambition, avarice, anger and other passions without fear of some coercive power which sets over men with both right and force sufficient to compel performance (Hobbes 1960).

Prisoners' Dilemma game has been used in representing this collective action problem (for example, see Hardin 1971, 1982; Campbell 1985; Ordeshook 1986).[1] In one-shot Prisoners' Dilemma game which is non-cooperative game where no binding agreement is possible, (i) each player has two strategies available, "cooperate (C)" and "defect (D)", and D is a dominant strategy;

---

[1] This situation is also explained by other games such as Chicken game (Taylor 1987), and Assurance game (Runge 1984, Taylor 1987; Isaac et.al. 1989; and Snidal 1985). But, because of its "frequency and conceptual importance" (Tsebelis 1990, 107), I focus on the Prisoners' Dilemma game in this paper.

(ii) if each player uses his/her dominant strategy, then the final outcome is Pareto-inferior in that both players can find some other outcome that they jointly and unanimously prefer; and (iii) even if the players can communicate beforehand and agree to avoid the Pareto-inferior outcome, unless they can somehow make a binding agreement, then each player will defect from it (Ordeshook 1986, 206-210)[2]. This line of logic ends up with maintaining that collective action problem cannot be resolved without "external authority with centralized coercion", that is, a 'Leviathan' in Hobbes's term.

Evolution of Cooperation through TIT FOR TAT

The theory of collective action based upon free-rider hypothesis, however, has serious problems. First, it has been repeatedly reported that self-organized institutions have been devised without reference to central authorities and sustained over long periods without enforcement by external agents (Fleishman 1988; E. Ostrom 1990a, 1990b; Feeny et al. 1990; Ostrom et.al 1991; Tang 1991; Weissing and Ostrom 1990). Experimental studies also have repeatedly found that participants consistently achieve better outcomes than are predicted by noncooperative game theory (see Dawes et al., 1986).

This anomaly seems to be overcome by a "super-game

---

[2] John Nash (1951) distinguished between cooperative and noncooperative game theory using the dual criteria - the availability of communication and enforceable agreements. Most contemporary game theorists, however, rely primarily on the second criteria. For more detail, see Harsanyi and Selten 1988.

approach". Robert Axelrod shows that voluntary cooperation can emerge among egoists (Axelrod 1981, 1984). Using computer tournaments, he tries to show that the super-strategy called TIT FOR TAT might be able to solve collective action problem if participants do not discount the future too much.[3] It seems to me that, however, the implication of his result must be interpreted very cautiously, since it tends to be somewhat over-estimated as if TIT FOR TAT solves collective action problems completely. His result seems problematic in several aspects.

First, according to "Folk Theorem", cooperation is nothing more than one of the infinite number of equilibrium which are available in infinitely repeated Prisoners' Dilemma (Fundenberg and Maskin 1986; Rasmusen 1989; Binmore and Samuelson 1990; Kreps 1990; Sened 1991). Players, therefore, fail to cooperate "not because they are stupid or particularly naughty, but because cooperative equilibria are not the only equilibria that exist in the repeated Prisoners' Dilemma super-games" (Sened 1991, 19).

Secondly, Axelrod shows that TIT FOR TAT can be collectively stable if the future is not discounted too much. That TIT FOR TAT is collectively stable, however, cannot guarantee the evolution of voluntary cooperation in collective action situation. When he mentions that TIT FOR TAT is collectively stable strategy, it means, "if everyone in a population is cooperating with everyone else because each is using TIT FOR TAT

[3] TIT FOR TAT plays in the following manner. It cooperates on the first round. After that, it always does what the other player did on the previous round.

strategy, no one can do better using any other strategy provided the discount parameter is high enough" (Axelrod 1981, 312). That is, he shows the condition under which cooperation can be sustained by TIT FOR TAT, but he does not show how cooperation through TIT FOR TAT can actually emerge.

Thirdly, it may be costly to monitor other's behavior and there can be some mistakes in interpreting other's behavior (Kreps 1990). These two possibilities can make voluntary cooperation through TIT FOR TAT extremely difficult (see Hirshleifer and Coll 1988).

Finally, experimental studies show that subjects tend not to use conditional cooperation strategy such as Trigger and TIT FOR TAT (Ostrom et al. 1990; Ostrom et al., 1991). It also has been criticized that human decision makers do not care for complete strategic plan (that is, super-game strategy like Trigger and TIT FOR TAT) but employ rather play-oriented strategic consideration (Güth 1990).

Evolution of Institution

For these reasons, it becomes interesting to ask how cooperation can emerge in Prisoners' Dilemma situation without using a super-game approach. One possible candidate for overcoming this anomaly is the evolution of institutions for resolving collective action problem. Contrary to the theoretical predictions based on "Free-Rider hypothesis" (for example, Taylor 1987, 158), there are many instances of social life governed by unwritten laws and customs where the 'Leviathan' plays little or no role (see E. Ostrom 1989, 1990a, 1990b, 1991; Ostrom et. al. 1990, 1991; Feeny et al. 1990; Rowe 1990; Snidal 1985; Sened 1991; Tang 1991). This implies that, cooperative solution can emerge in Prisoners' Dilemma situation through self-governing institutions.

The term "institution" refers to several different concepts. It can refer to a specific organization; it can describe established human relationships in society; or it can denote the rules that individuals use to order specific relationship with one another (for more detail, see Sened 1991). In this paper, I use the term "institution" in this latter sense: an institution is "simply the rules actually used (the working rule or rules-in-use) by a set of individuals to organize repetitive activities that produce outcomes affecting those individuals and potentially affecting others" (E. Ostrom 1990b, 11).

With this concept of institution, we can say that every dilemma situation represented by a game is determined by both physical regularities and institution (see Kiser and Ostrom 1982; Gardner and Ostrom 1989). This statement implies that some sorts of status-quo rule exist in every case, and players do not interact with each other in the "institutional-free Hobbesian world" (Levi 1990,1). Thus, the incentive structure of a game is the result of the interaction between physical regularities and rules-in-use. This line of logic is likely to be misunderstood as saying that the institution is exogenously given. We must remember that, however, social institutions are not always exogenous to human actions. In real world situation, players of a game would choose their actions considering not only the effect of their actions within the institutional framework, but also the affect of their action on the institutional framework itself (Gardner and Ostrom 1989; E. Ostrom 1989, 1990a; Rowe 1990). This implies that institution, which is defined as rules-in-use, can be imposed and changed by players themselves to the extent that status-quo rules permit. That is to say, players can change institutions, at least incrementally, and institutions can evolve through these changes. Collective action problem, then, can be resolved by this evolution of institutions.

Recently, E.Ostrom (1989, 1990a), Tsebelis (1990) and Sened (1991) pay attention to the possibility that players can change any of the rules-in-use that are potentially under the control of them in the status-quo rules. That players also consider the effect of their action on the rules-in-use, however, has not been captured in a formal game, and there is still no formal theory which explains how institution evolves. In the next sections, I introduce the third strategy - "changing rules-in-use"- into the Prisoners' Dilemma game, and try to answer the question "How institution evolves to solve collective action problems?" using evolutionary game theoretic approach.

Evolutionary Game Approach

Game theorist in social science, as well as in biology, have recently been devoting increasing attention to evolutionary approach (for examples, see, Maynard Smith 1982; Axelrod 1981, 1984, 1986; Reichert and Hammerstein 1983; Schuessler 1989; Gardner and Morris 1989; Binmore and Samuelson 1990; Miller and Andreoni 1990). This approach seems helpful in understanding complex dynamics of collective action problems. Such approaches are characterized by the common feature that more successful strategies grow while less successful strategies shrink in representation of the general population.

Most literatures using this approach, however, have some problems. First, they deal with only the evolution of strategies rather than the evolution of institutions. Players are almost always assumed to have only two choices - "C" and "D". Even though infinite numbers of strategies are introduced in the super-game approach, all of them are simply the different combinations of "C" and "D". By incorporating the third strategy of 'Time-Out (T)' changing rules-in-use into the Prisoners' Dilemma game, I think, we can deal with the evolution of institutions, given that institution refers to rules-in-use[4].

Secondly, it also has some problems in choosing equilibrium concept. As Hirshleifer and Coll (1988) point out, most of studies using this approach employ the equilibrium concept

---

[4] Notice that, however, it is extremely difficult to capture the evolution of institution in a simple formal game, and the concept of evolution used in biology cannot be applied to the study of institution without changes. For this reason, in this paper, I simply assume that the evolution of institution is a kind of cultural evolution (for more detail, see Boyd and Richerson 1985).

introduced by Maynard Smith (1982), that is, the "evolutionarily stable strategy (ESS)". ESS is a purely static specification of condition for equilibrium, defined only in terms of payoffs.[5] In other words, ESS condition implies that some strategy is not vulnerable to invasion by sufficiently small groups of mutant strategy (Binmore and Samuelson 1990). This concept of equilibrium cannot explain the complex dynamics of the evolution of strategies especially when there are more than two Nash equilibria. For this reason, this paper uses another equilibrium concept - the "evolutionary equilibrium distribution (EED)", based on the concept introduced by Hirshleifer and Coll (1988). EED refers to "the ultimate stable proportions to which an evolutionarily changing population converges, if such exist [with given initial distribution of strategies]" (Hirshleifer and Coll 1988, 368). In this paper, EED is denoted as $(p_1, p_2, \ldots, p_n)$, where p_i ($0 \le p_i \le 1$ and $\Sigma p_i = 1$) refers to the proportion of players who follow strategy i. In sum, EED enables us to analyze the dynamic process which is idiosyncratic to a particular initial condition.[6]

In the next section, I will develop a formal model of 'Time-Out Game' based upon what we have discussed so far.

---

[5] The formal criteria for 'a' to be an ESS are:
(i) $U(a,a) > U(b,a)$,
or(ii) $U(a,a) = U(b,a)$ and $U(a,b) > U(b,b)$
, where U(x,y) refers to payoff to player choosing x, when other player chooses y.

[6] There can be more than one ESS, but there is only one EED given specific initial distribution of strategies.

Time-Out Game

Time-Out game is based upon several basic assumptions:

(i) The world is populated by players that interact pair-wise in a Prisoners' Dilemma situation. A single player may be interacting with many others, but the player is assumed to be interacting with them one at a time. It means that only two player games are considered here[7];

(ii) Players are interacting with randomly-chosen partners. This random pairing assumption probably makes "the task of cooperators more difficult among egoists" (Schuessler 1989, 733). In addition, even though a single player may interact with more than one (randomly-chosen) partners, the number of partners he may interact with in a single period is assumed to be decided randomly and small enough to exclude the possibility of conditional cooperation strategies such as TIT FOR TAT. For this reason, in this hypothetical world, contrary to Axelrod's round-robin tournament world, players cannot identify, trace and punish defectors;

(iii) Players have another choice called "Time-Out (T)" in addition to "C" and "D". This strategy plays as follows: It tries to find some possible way of changing rules-in-use to the extent that the status-quo rule permits. It probably costs players to search this possibility. Time-Outer cannot make

---

[7] We need to pay close attention to the limitation of two person game. Delimiting inquiry to two-person game can be problematic, especially in dealing with the problem of provision of public good.

change unless the other player also chooses Time-Out strategy. If both players choose Time-Out, then they can get some changes in rules-in-use allowing them to get some sort of cooperative outcome, which are permitted in the status-quo rules. There can be several possible changes in rules-in-use which enables them to have some sort of cooperative outcomes. Here, however, I assume that there is only one, if any, possible changes in rules-in-use under given status-quo rule.[8] The implementation of this change is also costly to the players. If the other player chooses strategies other than Time-Out, then his Time-Out strategy works like "Defect". This assumes that rational players will not cooperate and search for the possibility of the change in rules-in-use at the same time. It is because they know that both cooperation and searching for the possibility of the change in rules-in-use are costly to them. For this reason, they will "defect" while they search for the possibility of the change in rules-in-use.

In sum, the game (G) is 3x3 rather than 2x2. Let c (c>0) and e (e>0) denote search cost and implementation cost[9]

---

[8] Notice that the existence of the possibility of some sort of cooperative outcome does not necessarily means that there can be a some mechanism available to the players to make enforceable threats or commitments, which are not allowed in non-cooperative game. Binding agreements or enforceable agreements is possible only when they are enforced <u>against</u> the participants' own will - i.e., there is an incentive to deviate from them. In our game, players have no incentive to deviate from (T,T) outcome as long as the change is efficient (it will be explained shortly after).

[9] These two costs are similar to what Professor E. Ostrom called "ex ante (transformation) cost" and "ex post (monitoring and enforcement) cost", respectively. For more details, see E.

respectively, then the game looks like the matrix in table 1:

(Table 1 about here)

The portion of this game surrounded by double line is Prisoners' Dilemma where T > R > P > S, as mentioned in assumption 1. As you can see, I also assume that payoffs are symmetric.

This game can have two Nash equilibrium outcomes - (D,D) and (T,T) - if and only if (R-e-c) is greater than or equal to P. In words, if the sum of the search cost and implementation cost does not exceed the payoff difference between reward and punishment, then this game can have two equilibria. This is reasonable in that if there exists some possible change which can bring some sort of cooperative outcomes but the costs of this change is too expensive, then players cannot afford this kind of changes and there exists an incentive to deviate from it even though they somehow arrive at this outcome. Based upon this, we can get the definition of the efficient change:

DEFINITION: A change in rule-in-use is efficient if and only if the sum of the search cost and implementation cost does not exceed the payoff difference between reward and punishment, formally (R-P) > (c+e).

How can we select one equilibrium from those two? To answer this question, I employ evolutionary game theoretic approach, as mentioned before. Here, I assume that players do not use super-game strategy such as TIT FOR TAT, and future payoffs are not discounted. It is because in evolutionary modelling, "the

---

Ostrom (1990a, 198-205).

survival of strategies is more important than utility maximization with a certain time preference" (Schuessler 1989, 736). Evolutionary approach is, to repeat, based upon the evolutionary principle that strategies shown to be relatively effective will be used more in the future than the less effective strategies. It can imply that (i) the more effective individuals are more likely survive and reproduce; (ii) the players learn by trial and error, that is the players observe each other, and those with poor performance tend to imitate the strategies of those they see doing better.[10]

To analyze the evolutionary process and find EEDs, I employ an explicit dynamic formula called the replicator dynamics, which is standard in models of this type (see, Hofbauer and Sigmund 1988; Hirshleifer and Coll 1988; Gardner and Morris 1989):[11]

$$X_i = K*X_i*(F_i(X) - F(X)), \qquad (1)$$

where X_i represents the change in any single period in the proportion X_i following strategy i, $F_i(X)$ refers to the respective average payoff to strategy i, F(X) is the average

[10] In general, the cultural evolution of rules-in-use refers only to the former case (Boyd and Richerson 1985). However, I include both cases without making clear the difference between them. It is because, by doing so, we also can simulate the iteration of game without super-game approach.

[11] I use a discrete-time-interval model (based on difference equation technique) rather than continuous model (based on differential equation technique). It is because, in our model, the size of next generation or round is assumed to be determined by that of the current one. This choice can make a big difference (see, Goldberg 1986, 46-49). Also, notice that this evolutionary mechanism is only one of many, and we cannot tell which one is better in describing the evolution of social phenomena.

payoff in the population, and k ($0<k\leq 1$) denotes the constant representing the 'sensitivity' or speed of response. This equation enables us to analyze how the distribution of strategies changes over time in the situation where each player interacts with randomly-chosen players, and tends to imitate others' strategy performed better than his strategy in the previous period[12].

As Axelrod (1986) points out, "the evolutionary approach...involves nonlinear effects...[Therefore,] it is often impossible to use deductive mathematics to determine the consequences of a given model" (p.1098). In our model, the replicator dynamics is composed of three nonlinear equations.[13] For this reason, I use computer simulation program called 'PHASER'(see Koçak 1989) to reveal the dynamics of a process. In the next section, I will analyze the results of simulation and

---

[12] This equation can also represent Axelrod's round robin tournament situation where each player interacts with <u>all</u> the other players in the population (for example, see Hirshleifer and Coll 1988). But, in this paper, I assume anonymous society in order to exclude the possibility of contingent strategies such as TIT FOR TAT. It means that (i) a single player may interact with <u>more than one</u> players at one round or time period, but the number of players he can interact with in one time period is not large enough to use contingent strategies; (ii) therefore, it is extremely difficult to identify, trace and punish defectors.

[13] They are:

$$x=Kx\{Rx+Sy+Sz-Rx^2-(S+T)xy-(S+T-c)xz-(2P-c)yz-Py^2-(R-c-e)z^2\}$$
$$y=Ky\{Tx+Py+Pz-Rx^2-(S+T)xy-(S+T-c)xz-(2P-c)yz-Py^2-(R-c-e)z^2\}$$
$$z=Kz\{(T-c)x+(P-c)y+(R-c-e)z-Rx^2-(S+T)xy-(S+T-c)xz-(2P-c)yz-Py^2-(R-c-e)z^2\}$$

,where x, y, z refers to the change in any single time period in the proportion of players following strategies C,D,T, respectively, and x, y, z refers to the proportion of players following strategies C,D,T, respectively.

the implication of these results.

Simulation of Time-Out Game

Before discussing simulation results, we need to point out two things. First, if the change in rules-in-use is not efficient, then (0, 1, 0)[14] is the only EED and there cannot be an evolution of institutions. It is because (T,T) outcome, as well as (C,C) outcome, cannot even be a Nash equilibrium and there exits one dominant strategy (D) if the change in rules-in-use is not efficient (see the definition in previous section)[15]. Secondly, even if the change in rules-in-use is efficient, (i) the ultimate extinction of "C" and a stable division of the population between "C" and "T"; and (ii) a stable division of the population between "C", "D", and "T" cannot be an EED (see Appendix I for proof). This discussion can be summarized by the following proposition:

PROPOSITION 1: If the change in rules-in-use is efficient, then both (0,1,0) and (0,0,1) can be an EED. However, if the change in rules-in-use is not efficient, (0,0,1) cannot be an EED, and the evolution of institution and the ultimate extinction of "D" are impossible.

In the following sections, I will restrict my analysis only to the cases where the change in rules-in-use is efficient. Now, let me discuss the results of simulation. In this simulation, I

---

[14] For the rest of this paper, (x, y, z) refers to the (either EED or initial) distribution where x%, y%, and z% of population follow "C", "D", and "T" strategies, respectively.

[15] Only Nash equilibrium can survive in an EED. See, Hirshleifer and Coll (1988, pp.379-380).

used Axelrod's payoff matrix[16]. To begin with, I assumed that both c, e are .5, and k is .3. Then, I tried one initial distribution, (.33, .34, .33). The result is shown in Figure 1.

(Figure 1 about here)

As you can see, (0,0,1) is an EED in this case. This is an encouraging result for me since this means that institution can evolve if the population is evenly divided into 3 groups of players following 3 strategies. This is not very satisfactory in that rational players will not choose dominated strategy "C". Therefore, I tried another initial distribution (0,1,0), and, as we can easily expect, (0,1,0) is an EED. Actually, it is rather meaningless result, since, in our replicator model, no strategy can recover from 0%. So, I tried another initial distribution (0,.5,.5) which denotes the case where half of the population follows one Nash equilibrium strategy (D) and another half of the population follows the other Nash equilibrium strategy (T). In this case, (0,0,1) is an EED. Observing this result, I guessed that (0,0,1) can be an EED if the proportion of T in the initial distribution exceeds some threshold value, which is denoted as T^*. Based upon this thought, I increased the proportion of "T" gradually[17] and examined the results. It is because I wanted to know under what conditions (0,0,1) can be an EED - that is, institution can evolve, when nobody follows "C" strategy. In

---

[16] T, R, P, S are 5, 3, 1, 0 respectively. See Axelrod (1981, 1984).

[17] I increase it by '.01'. I use this sort of 'grid search' when I examine the dynamic processes of our model.

addition, I thought that this process can simulate the situation where only small proportion of players notice the possibility of efficient change and choose "T". After trial of several initial distributions, I found that my guess is right. (0,0,1) can be an EED if more than 11% of players follow "T" strategy. To have a cooperative result through the evolution of institution, we need only 11% of players who notice the possibility of efficient change and follow "T" strategy, even though there is no cooperative player in the population. This result is shown in Figure 2. As you can see, the threshold value

(Figure 2 about here)

of T (T^*) is 11% in this case. It means that if the initial proportion of "T" is greater than 11%, then "D" goes extinct ultimately (see Figure 2(b)), whereas "T" goes extinct if the initial distribution of "T" is less than 11% (see Figure 2(a)). It is astonishing that such a relatively small portion of players is needed to have cooperative outcome through evolution of institution.

Next, I assumed that the initial proportion of "T" above which (0,0,1) can be an EED, T^*, may be affected by the initial proportion of players who follow "C" strategy. It is because relative performances of strategies are heavily dependent upon the distribution of strategies in the population. Based on this assumption, I searched T^* values by increasing the initial proportion of "C" by 10%, from 10% to 90%. The result is that: As "C" increases, T^* tends to decrease. If "C" is 90% initially, then T^* is 3%, whereas T^* is 11% if there is no "C". These results are shown in Table 2(a).

(Table 2 about here)

This discussion can be summarized as follows:

PROPOSITION 2: Even if nobody chooses dominated strategy "C", relatively small proportion of players following "T" strategy (in our parameter configuration, 11%) can make the evolution of institution possible.

PROPOSITION 3: There exists a threshold value of T in the initial distribution of the strategies, T^*, above which "D" goes extinct ultimately and only cooperative outcome through institution can survive. And, as the number of cooperative players increases in the population, the value T^* tends to become smaller.[18]

Then, I changed c, e values and examined the results of these changes. The findings are: (i) if c and e are very small, then T^* can be very small in comparison to the case where c and e are large;[19] (ii) if c and e are large (but, of course, not large enough to have inefficient change), then T^* need to be larger in comparison to the case when c and d are small. This result is shown in Table 2(b) and 2(c). Notice that if we search for T^* value without any restriction, then T^* can be very small. If, under some conditions, T^* is only .00001, then one "T" player can bring about the evolution of institution to solve the collective action problem, as long as the size of the population

---

[18] We can observe general tendency of decreasing T^* when C increases. This tendency, however, is not without exception. (see, Table 2 (c) and (d)).

[19] If c and e are .01, T^* needs to be only .01 (i.e., 1%) irrespective of the initial proportion of "C". Notice that it is the minimum value of T^* attainable in our search. It may be even smaller than this if we do not restrict our search only to two digits below zero.

is less than 100,000. These results can be summarized as follows:

PROPOSITION 4: As the search and implementation costs of change in rules-in-use increase, we need larger value of T^* to have cooperative outcomes through the evolution of institution. And, there may be values of c and e which are so small that only one player can bring about the evolution of institution.

Turning back to the condition for the efficient change in rules-in-use, we can have several sets of c and e which have identical (c+e) values satisfying this condition, since it only requires that (c+e) should be less than some value. We can expect that c and e have different effects on T^*. Let (c+e) be some kind of efficiency index, then we can have two hypothetical cases with same efficiency index, but one has higher search cost and lower implementation cost, on the other hand, the other has lower search cost and higher implementation cost. I assumed that (i) they may have different effects on T^* values; and (ii) the former case may need larger T^* value than the latter. It is because, as Professor E. Ostrom (1990a, 198) points out, if search cost is relatively high, then players may tend not to do further cost calculation, and the evolution of institution may become difficult. To find out this difference, I tried four cases - two with (c+e)=1.5, and two with (c+e)=2. The finding is shown in Table 2(d)-2(g). As you can see, the findings support my previous argument. T^* is larger when e is higher in comparison to the case where e is lower, even though (c+e) are identical in both cases. In particular, it turns out that (i) When search cost is very high, it is extremely difficult to get cooperative outcome. Unless there is no "D" in the initial distribution, under some conditions, "T" is doomed to go extinct; on the other hand (ii) When search cost is very low, very small T^* can bring out cooperative outcome, even though implementation cost is very high. This can be summarized as follows:

PROPOSITION 5: If we have two possible changes in rules-in-use, which have same (c+e) value, then the one with lower search cost is much more likely to bring about cooperative outcome, than the other with higher search cost.

A Model of the Evolution of Institution

All these results are summarized in Figure 3. In Figure 3,

(Figure 3 about here)

(0,0,1) is an EED in the shaded areas above the T^* lines, whereas (0,1,0) is an EED in the clean areas below the T^* lines. That (0,0,1) is an EED means that 'status-quo rules-in-use' can be changed and institution can evolve without external intervention. And that (0,1,0) is an EED means that players cannot escape from the dilemma situation. Notice that the shaded areas are larger than the clean areas in all cases. It may imply that the evolution of institution without external intervention is not so difficult as predicted by "free-rider hypothesis".

In our model, to recapitulate, we have two possible EEDs - (0,1,0) and (0,0,1). Either one of them can be an EED, depending on "payoff components (T, R, P, S, c and e)" and "initial distribution of three strategies". Payoff components are, as mentioned earlier, determined by both physical regularities and rules-in-use. These payoff components determine T^* value, the minimum proportion of "T" in the initial distribution required to make (0,0,1) an EED. Which factor, then, does affect the initial distribution of strategies, which also plays pivotal roles in determining the outcome of the game? Maybe the attribute of the population (or community) does. The initial distribution of strategies can vary across the communities. Some communities may have more cooperative players initially than other communities, even though they all face the same dilemma situation. Based upon these, I think we can say that (i) **Physical regularities** and **Rules-in-use** determine the incentive structure of the game; and (ii) **Attribute of the community** determine whether or not institution can evolve to solve collective action problems, under the incentive structure given by the physical regularities and the rules-in-use. Let the game with the third strategy be 'Time-Out Game (G)'. Then, G can be defined as a function of both 'physical regularities (π)' and 'status-quo rules-in-use (τ)' at a particular time point. If the outcome of G is "C" or "D", there may be no bottom-up (or decentralized) evolution of institution. It is because (i) if "C", then there is no reason to change status-quo rules-in-use (τ) at all; and (ii) if "D", then bottom-up evolution of institution is impossible and some sorts of external intervention may be inevitable. On the other hand, if the outcome of G is "T", then institution can evolve without external interventions. If all population ultimately follows the strategy of changing rules-in-use, then 'status-quo rule (τ)' is changed into τ' and G is also changed into G'. And, as discussed before, the outcome of G is determined by the initial distribution of the strategies (i.e., the attribute of the community). If initial proportion of T is greater than T^*, then "T" is an outcome of G. If initial proportion of T is smaller than T^*, then "C" is an outcome of G.

In G', then, players have different incentives and 'T' than in the previous game, G. If new incentive structure in G' brings about cooperative outcome, as expected before, no further change in rules-in-use is necessary. However, if, for whatever reason (for example, due to changes in π or higher-level rules which affect rules-in-use, τ), new incentive structure in G' cannot bring about cooperative outcomes, accordingly, further change in rules-in-use is required and the process is repeated. Through the iteration of this process, the rules-in-use (that is, institution) evolves. This discussion can be depicted as follows:

(Figure 4 about here)

Notice that this model assumes that even if two different societies or communities face the identical incentive structure of G initially given by physical regularities and rules-in-use, they can have totally different results and it depends on the attribute of the community (that is the initial distribution of the strategies). Therefore, we cannot expect that "one good institutional arrangement", which can solve the collective action problem of one community, can always solve the same problem in another community even though the physical regularities and rules-in-use are identical in both cases.

Concluding Remarks

So far, I have examined the results of Time-Out game using simulation. The major findings are: (i) Cooperative outcome is attainable through the evolution of rules-in-use, as long as there exists an efficient change in rules-in-use; (ii) if new incentive structure cannot bring about cooperative outcomes due to some changes in physical regularities or higher-level rules, then Time-Out game will be played again; (iii) through the iteration of (i) and (ii), institution may evolve to solve collective action problems without external interventions.

This study is, however, somewhat unsatisfactory in that it depends upon assumptions which seems questionable. They are assumptions on: (i) pairwise interaction with randomly chosen partners; (ii) the symmetry among players; and (iii) the uniqueness of the possible change in rules-in-use under given status-quo rule, and so on. Simplification through several assumptions is inevitable in modelling complex social phenomena. The problems are: (i) to what extent the result of the model depends on assumptions; and (ii) whether or not these assumptions are "right type of simplification" (McGinnis 1991, 3). The results may be very sensitive to some small changes in assumptions. I think we need further study on this problem.

In addition, I did not use deductive mathematics in analyzing my model. I think that, by formally describing T^* as the function of parameters using deductive mathematics, we may be able to capture more detailed dynamics which might be missed in the simulation analysis. As mentioned before, however, it is often impossible to use deductive mathematics in the evolutionary game involving nonlinear effects. We also need further study on the advantages and disadvantages of using deductive mathematics.

Figure 1: Dynamic Path of Time-Out Game

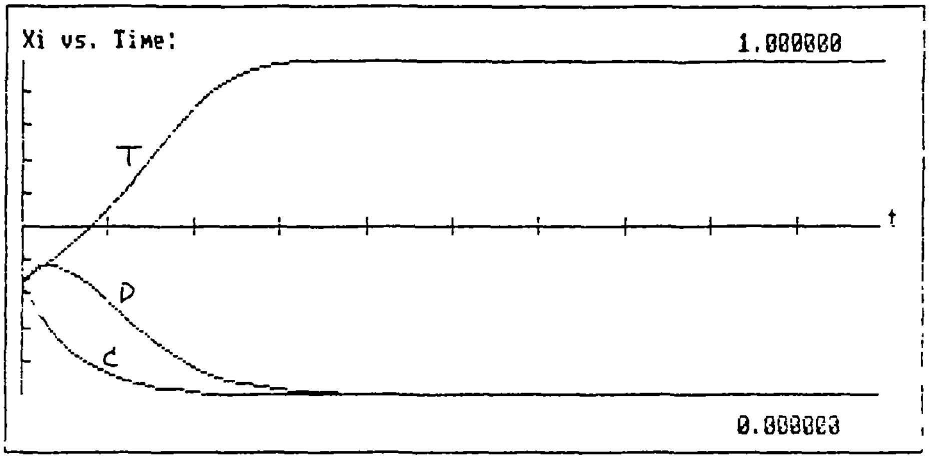
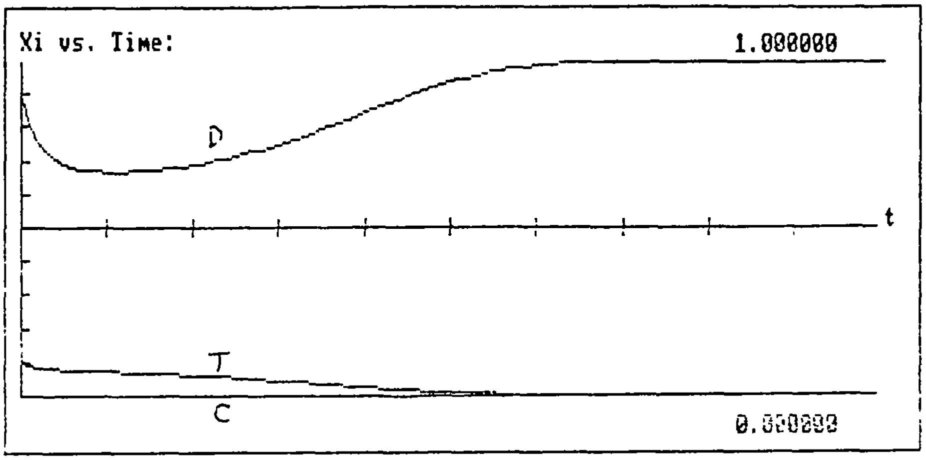


Figure 2: Dynamic Paths

(a) when T is smaller than the threshold value of T:



(b) when T is greater than the threshold value of T:

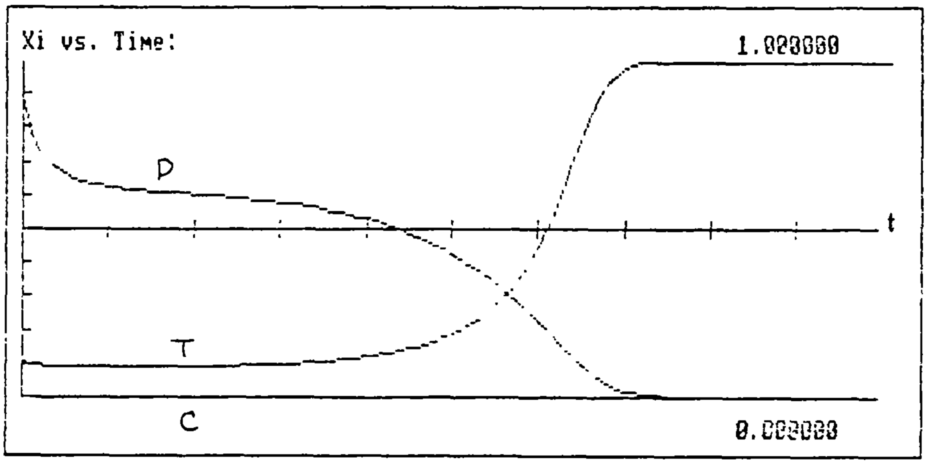
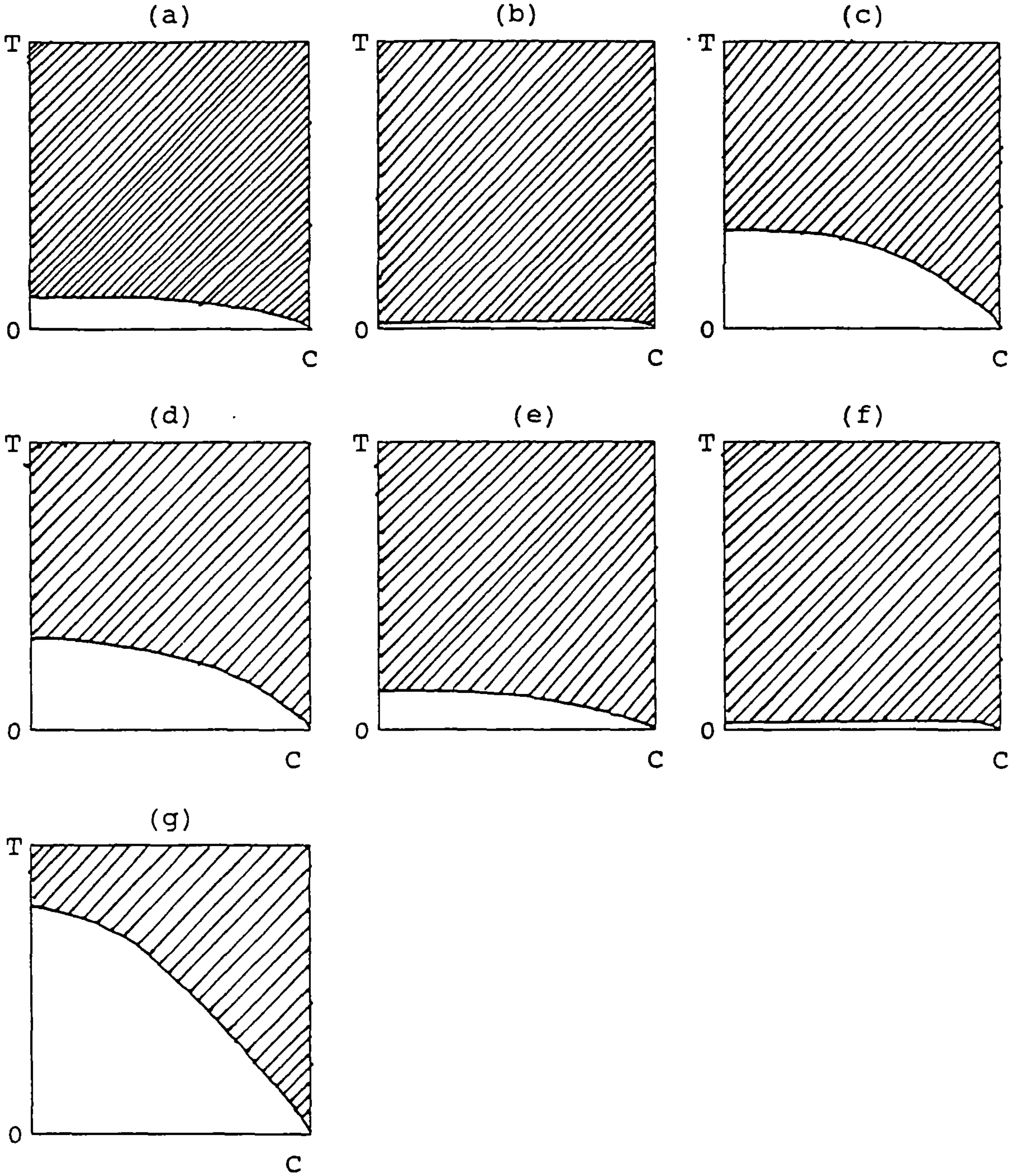


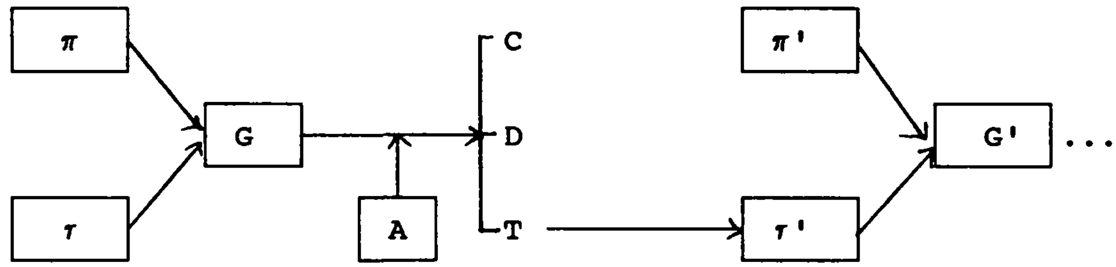
Figure 3: Graphic Representation of T^* lines



*Note: Both Axes are in % (from 0 to 100).

(a): (c,e) = (.5,.5)
(b): (c,e) = (.01,.01)
(c): (c,e) = (1,1)
(d): (c,e) = (1,.5)
(e): (c,e) = (.5,1)
(f): (c,e) = (.01,1.99)
(g): (c,e) = (1.99,.01)

Figure 4: Final Model of the Evolution of the Institution



,where π: physical regularities
τ: rules-in-use
G: Time-Out game
A: Attribute of the Community

Table 1. Prisoners' Dilemma + Time-Out

	C	D	T
C	R R	S T	S T-c
D	T S	P * P	P P-c
T	T-c S	P-c P	R-c-e * R-c-e

Table 2: PHASER Simulation Results

%"C"	(a)	(b)	(c)	(d)	(e)	(f)	(g)
0	11	1	29	28	13	1	77
10	11	1	30	29	13	1	77
20	11	1	32	29	12	1	74
30	10	1	32	29	12	1	67
40	9	1	31	28	11	1	59
50	9	1	30	26	10	1	50
60	7	1	27	24	9	1	40
70	6	1	23	20	7	1	30
80	5	1	17	15	6	1	20
90	3	1	9	8	4	1	10

* All entries are T^* in %.
*Note: (a): (c,e) = (.5,.5)
(b): (c,e) = (.01,.01)
(c): (c,e) = (1,1)
(d): (c,e) = (1,.5)
(e): (c,e) = (.5,1)
(f): (c,e) = (.01,1.99)
(g): (c,e) = (1.99,.01)

Appendix 1

This proof is based on the 'Conditions for 3X3 Evolutionary Equilibrium' provided by Hirshleifer and Coll (1988, 379-380).

1. (x, y, z), where x=0, and 0<y,z< 1):
This means that the ultimate extinction of "C" and a stable division of the population between "C" and "T" is an EED. To have this kind of EED, we need the following conditions:

(1) necessary condition:

$$P-c \geq P \tag{a.1}$$
$$P \geq R-e-c \tag{a.2}$$

(2) sufficient condition:

$$P-c > P > S \tag{a.3}$$
$$P > R-c-e > S \tag{a.4}$$

or

$$P = P-c > S \tag{a.5}$$
$$P = R-c-e \tag{a.6}$$

In our assumption, c, e > 0, so inequalities (a.1), (a.3), (a.4), and (a.5) cannot be satisfied. In addition, by definition, efficient change means P < R-e-c, so inequalities (a.2), and (a.6) cannot be satisfied. Therefore, we cannot have this kind of EED. Q.E.D.

2. (x, y, z), where 0<x,y,z<1:
This means that a stable division of the population between "C", "D", and "T" is an EED. To have this kind of EED, we need the following conditions:

(1) necessary condition:

$$R \geq T,\ T-c \tag{a.7}$$

(2) sufficient condition:

$$R > T,\ T-c \tag{a.8}$$

or

$$R = T > T-c \tag{a.9}$$
$$S > P \tag{a.10}$$

By assumption, c >0, and T > R > P > S, so inequalities (a.7), (a.8), (a.9), and (a.10) cannot be satisfied. Therefore we cannot have this kind of EED. Q.E.D.

Bibliography

Axelrod, Robert. (1981). "The Emergence of Cooperation among Egoists". American Political Science Review. 75:306-318.

__________. (1984). The Evolution of Cooperation. New York: Basic Books.

__________. (1986). "An Evolutionary Approach to Norms". American Political Science Review. 80:1095-1111.

Bartholdi III, John J., C. A. Butler and M. A. Trick. (1986). "More on the Evolution of Cooperation". Journal of Conflict Resolution. 30:1. 129-140.

Binmore, Ken and L. Samuelson. (1990). "Evolutionary Stability in Repeated Games Played by Finite Automata". Paper presented at the Micro Workshop at Indiana University, Nov. 15, 1990.

Boyd, R. and P.J. Richerson. (1985). Culture and Evolutionary Process. Chicago: University of Chicago Press.

Campbell, R. (1985). "background for the Uninitiated". in Paradoxes of Rationality and cooperation. eds. R. Campbell and L. Sowden. Vancouver: University of British Columbia Press. 3-41.

Feeny, David et.al. (1990). "The Tragedy of the Commons: Twenty Two Years Later". Human Ecology. 18:1. 1-17.

Fleishman, John A. (1988). " The Effect of Decision Framing and Others' Behavior on Cooperation in a Social Dilemma". Journal of Conflict Resolution. 32:1. 162-180.

Fundenberg, Drew and Erik Maskin. (1986). "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information". Econometrica. 54:3. 533-554.

Gardner, Roy and Elinor Ostrom. (1989). "Rules and Games". Public Choice. Forthcoming.

Gardner, Roy and M. Morris. (1989). "The Evolution of Bluffing in Animal Contests: An ESS Approach". Journal of Theoretical Biology. 137:235-243.

Goldberg, Samuel. (1986). Introduction to Difference Equations: With Illustrative Examples from Economics, Psychology, and Sociology. New York:Dover.

Güth, Werner. (1990). "Game Theory's Basic Question: Who is a Player? - Examples, Concepts, and Their Behavioral Relevance". Plenary Lecture at the First Summer School in

Economic Psychology, Linz/Austria.

Hamberger, Henry. (1973). "N-Person Prisoners' Dilemma". The Journal of Mathematical sociology. 3:27-48.

Hardin, Gerald. (1968). "The Tragedy of the Commons". Science. 162:1243-1248.

Harsanyi, John and R. Selten. (1988). A General Theory of Equilibrium Selection in Games. Cambridge: MIT Press.

Hirshleifer, Jack and Juan C. M. Coll. (1988). "What Strategies Can Support Evolutionary Emergence of Cooperation?". Journal of Conflict Resolution. 32:2 (June). 367-98.

Hobbes, Thomas. (1960). Leviathan. Oxford: Basil Blackwell. Originally published in 1651.

Hofbauer, J. and K. Sigmund. (1988). The Theory of Evolution and Dynamical Systems. cambridge: Cambridge University Press.

Issac, R. Mark., D. Schmidtz, and J.M. Walker. (1989). "The Assurance Problem in a Laboratory Market". Public Choice. 62:217-236.

Kiser, L.L. and E. Ostrom. (1982). "The Three world of Action. A Metatheoretical Synthesis of Institutional Approaches". in Strategies of Political Inquiry. ed. E. Ostrom, pp.179-222. Beverly Hills:Sage.

Kreps, D.M. (1990). A course in Microeconomic Theory. Princeton: Princeton University Press.

Koçak, Hüseyin. (1989). Differential and Difference Equations through Computer Experiments. New York: Springer-Verlag.

Levi, Margaret. (1990). "A Logic of Institutional Change" in Margaret Levi and Karen Cook. eds. The Limits of Rationality. Chicago: The University of Chicago Press. Forthcoming.

Maynard Smith, John. (1982). Evolution and the Theory of Games. New York: Cambridge University Press.

McGinnis, Michael D. (1991). "Richardson, Rationality, and Restrictive Models of the Arms Race: A Return to Simplicity". Journal of Conflict Resolution. Forthcoming.

Miller, J.H. and J. Andreoni. (1990). "A Coevolutionary Model of Free-Riding Behavior: Replicator Dynamics as an Explanation of the Experimental Results". unpublished manuscript.

Mueller, Ulrich. (1988). "Optimal Retaliation for Optimal Cooperation". Journal of Conflict Resolution. 31:4. 692-724.

Nash, John F. (1951). "Non cooperative Game". Annals of Mathematics. 54:2. 286-295.

Olson, M. (1965). The Logic of Collective Action. Public Goods and the Theory of Groups. Cambridge: Harvard University press.

Ordeshook, Peter. (1986). Game Theory and Political Theory: An Introduction. New York: Cambridge University Press.

Ostrom, Elinor. (1989). "Microconstitutional Change in Multiconstitutional Political System". Rationality and Society. V.1, No. 1 (July), 11-50.

_______________. (1990a). Governing the Commons. New York: Cambridge University Press.

_______________. (1990b). Crafting Irrigation Institutions: Social Capital and Development. Decentralization: Finance and Management Project Vol.1 No.1.

Ostrom, Elinor , J. Walker and R. Gardner. (1990). "Sanctioning by Participants in Collective Action Problems". Paper delivered at the 1990 Annual Meeting of the American Political Science Association.

_______________________________________. (1991). "Covenants with and without a Sword: Self-enforcement Is Possible". Paper presented at the Midwest Political Science Association Meetings, Chicago, IL.

Rasmusen, Erick. (1989). Games and Information. Cambridge: Basil Blackwell.

Raub, Werner. (1988). "Problematic Social Situations and the 'Large-Number Dilemma' A Game-Theoretical Analysis". The Journal of Mathematical Sociology. 13:4. 311-357.

Reichert, Susan E. and Peter Hammerstein. (1983). "Game Theory in the Ecological Context". Ann. Rev. Ecol. Syst.. 14:377-409.

Rowe, Nicholas. (1990). Rules and Institutions. Ann Arbor: The University of Michigan Press.

Runge, C. Ford. (1984). "Institution and the Free-Rider: The Assurance Problem in Collective Action". Journal of Politics. 46:154-181.

Schuessler, Rudolf. (1989). "Exit Threats and Cooperation under

Anarchy". Journal of Conflict Resolution. 33:4. 728-749.

Sened, Itai. (1991). "Contemporary Theory of Institutions in Perspective". Journal of Theoretical Politics. Forthcoming.

Snidal, Duncan. (1985). "Coordination versus Prisoners' Dilemma: Implication for International Cooperation and regime". American Political Science Review. 79:923-942.

Tang, Sui Yan. (1991). "Institutional Arrangements and Management of Common-Pool Resources". Public Administration Review. 51(1)(Jan./Feb.):42-51.

Taylor, Michael. (1987). The Possibility of Cooperation. Cambridge: Cambridge University Press.

Thomas, Ewart and M.W. Feldman. (1988). "Behavior-Dependent Contexts for Repeated Plays of the Prisoner's Dilemma". Journal of Conflict Resolution. 32:4. 699-726.

Tsebelis, George. (1990). Nested Games: Rational Choice in Comparative Politics. Berkeley: University of California Press.

Weissing, Franz and E. Ostrom. (1990). "Irrigation Institutions and the Games Irrigators Play: Rule Enforcement without Guards". in. Game Equilibrium Models II: Methods, Morals, and Markets. ed. Reinhard Selten. Berlin: Springer-Verlag.