Discussion Paper #2005-7

Group Formation and Heterogeneity in Collective Action Games

by

Toshimasa Maruta and Akira Okada

October, 2005

Group Formation and Heterogeneity in Collective Action Games[1]

Toshimasa Maruta[2] and Akira Okada[3]

October, 2005

Abstract: We present a simple model of voluntary groups in a collective action problem where individuals differ in their willingness to cooperate. The heterogeneity of individuals' preferences generally yields multiple equilibrium groups with different levels of cooperation. Voluntary participation in a binding contract to cooperate does not necessarily lead to the full cooperation. Applications to voluntary provision of public goods and cartel formation in oligopolistic markets are discussed.

Journal of Economic Literature Classification Numbers: C70, C72, D70.

Keywords: group formation, collective action, heterogeneous preferences, voluntary participation.

[1] We are grateful to anonymous referees for their useful comments. Okada acknowledges financial support from the Japan Society for the Promotion of Science under grant No. (A) 16203011 and the Matsushita International Foundation.

[2] Corresponding author: Advanced Research Institute for the Sciences and Humanities, Nihon University, 12-5 Goban-cho, Chiyoda, Tokyo 102–8251, Japan. E-mail: *tosh@arish.nihon-u.ac.jp* Phone: +81 3 5275 9607 Fax: +81 3 5275 9204.

[3] Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601 Japan. E-mail: *aokada@econ.hit-u.ac.jp* Phone: +81 42 580 8599 Fax: +81 42 580 8748

1 Introduction

The collective action problem arises in social situations in which it is a common interest for all individuals to act collectively but each of them has an incentive to free ride on the others' actions. Examples of collective action problems can be found in many social, political, and economic activities. They include: public goods provision, cartel formation, labor union, environmental pollution, common-pool resources management, participation in community work, trading associations, international organizations, etc. A central question in the collective action problem is how a group of individuals voluntarily cooperate, and how such a voluntary group is stable. The purpose of this paper is to consider the formation of voluntary groups in collective action. Modelling the collective action problem as an n-person prisoner's dilemma, our analysis focuses on how the heterogeneity of individuals' preferences affects the formation of voluntary groups.

In many real situations, individuals differ in their willingness to participate in collective action. For example, consider a problem of environmental pollution. Some individuals are concerned very much with environmental preservation, and they are willing to contribute for the prevention of pollution even if they have a small number of followers. Others may be reluctant to participate in such a collective activity. They might contribute for the activity only if a large number of individuals have already done or will do. In such heterogeneous situations, the size of a group is not a decisive factor in its stability. The configuration of all members' attitudes to collective action matters greatly. In the presence of conflicting incentives to cooperate and to free ride, the group formation of heterogeneous individuals becomes more complicated than that of homogeneous ones.

When individuals form a group of collective action, the group must have some appropriate mechanism to enforce the collective action on members. The enforcement mechanism of cooperation has been extensively studied in the literature. In the theory of repeated games, various types of mutual punishments among individuals to sustain cooperation have been analyzed in dynamic situations. The folk theorem of repeated games (e.g., Fudenberg and Maskin 1986) shows that a collective action can be enforced by decentralized punishing behavior (such as trigger strategies) under suitable conditions. The theory of mechanism design has investigated the design of mechanisms to implement a desirable collective choice. Besides the enforcement problem of collective actions, the problem of voluntary participation is important: do individuals voluntarily participate in a group (or a mechanism)? In comparison to voluminous works on the enforcement problem of collective actions, the voluntary participation problem has not been studied very much. The theory of mechanism design often assumes that individuals non-strategically participate in the mechanism if they become better-off than in the status-quo. It should be remarked that whatever enforcement mechanism is employed, the mechanism itself is a kind of public good and that every individual has an incentive to free ride on it. This problem has been called the second order dilemma in the provision of public goods (Ostrom 1998). There exists a strategic conflict among individuals regarding who participate in the mechanism of cooperation. The celebrated Coase Theorem (Coase, 1960), which states that rational agents necessarily achieve a Pareto-efficient allocation through voluntary bargaining if the agreement is costlessly enforced, does not pay enough attention to free riding.

To focus the analysis on the voluntary formation of groups, we leave the enforcement problem of cooperation out of the scope of this paper. We simply assume that a group, once formed, is endowed with some costly enforcement mechanism.[1]

The process of group formation is modelled as a two-stage game. In the first stage, individuals decide independently to participate in a group or not. In the second stage, participants negotiate on a collective action. The agreement of cooperation is reached only if all participants agree to do so. If the agreement is reached, then participants are bound to cooperate, bearing the group costs. Any non-participant is allowed to free ride. If the agreement is not reached, then the group breaks up, and the n-person prisoner's dilemma is played.

Each individual's incentive to cooperate is characterized by the minimum size of a group in which her participation to it can make her better off (even with bearing

[1] As Dixit and Olson (2000) argue, the Coase Theorem can be extended to the case of positive enforcement costs (transaction costs) if enforcement costs are taken into account in defining the Pareto frontier.

group costs) than in the non-cooperative equilibrium of the prisoner's dilemma. We will call this number the individual's *threshold of cooperation*. Individuals with smaller thresholds are more motivated to cooperate. It is shown that a group is formed in a Nash equilibrium of the second stage if and only if the group size exceeds thresholds of its members. Such a group is called *successful*. By solving backward, the two-stage group formation game is reduced to the following n-person strategic-form game, which we call the *group formation game*. In this game, all individuals decide independently to participate in a group or not. The group is formed if and only if it is successful, that is, all participants are better off by cooperating than in the non-cooperative equilibrium. Any non-member free rides on the group action.

We show that the group formation game with heterogeneous preferences has multiple strict Nash equilibria. Specifically, besides the non-participation equilibrium, there are multiple equilibria with different levels of cooperation. A successful group can be sustained in equilibrium if every member is "critical" in that her unilateral deviation makes the group unsuccessful. In this paper, we will pay attention only to pure-strategy Nash equilibria.

This paper is related to several works in the literature. Similar two-stage game models of voluntary participation are analyzed by Palfrey and Rosenthal (1984), Okada (1993), Saijo and Yamato (1999) and Dixit and Olson (2000) among others.[2] Saijo and Yamato (1999) consider a general class of provision mechanisms of public goods satisfying Pareto efficiency in an economy with Cobb-Douglas utility functions. A basic conclusion of these works can be stated as follows: all individuals do not necessarily participate in a

[2]Palfrey and Rosenthal (1984) consider a one-stage model of voluntary participation in providing binary public goods, which has a payoff structure similar to that of the voluntary group formation game presented in section 2. Okada (1993) considers a three-stage model including negotiations on punishment levels in a group. Okada and Sakakibara (1991) and Okada, Sakakibara and Suga (1995) extend the model of voluntary participation to consider the formation of a constitutional state in a dynamic economy where the accumulation of public goods affect the formation of the sate. In these works, a political system of the state determines rules of selecting an enforcer and of distributing surplus. Okada, Sakakibara and Suga (1995) study a constitutional choice between a centralized system (monarchy) and a decentralized system (democracy) by an extended model.

mechanism of cooperation, due to the incentive of free riding. In this respect, this paper complements these works. The important difference between the previous works and ours is that their analyses are restricted to a symmetric case that all individuals have the same utility functions. We believe that the heterogeneity sheds a new light on the study of group formation in a collective action problem. Chew (2000) considers the formation of a social (communication) network in a coordination game which is a model of collective action different from the prisoners' dilemma game. In his model, every individual may be one of two types, either willing or unwilling to participate in a collective action. Individuals can communicate their types to only others linked by a network. Chew characterizes a minimum network structure under which all individuals participate in a collective action, regardless of a prior belief about types. Since the classic work of Olson (1965), the group formation in collective action has been widely investigated. The group size effect, argued by Olson (1965), that larger groups are less successful in organizing collective action is not necessary true in our model. The success of collective action critically depends upon the benefit and the cost to participate. The largest group may form, depending on the distribution of individuals' thresholds of cooperation.[3] Our model of the group formation differs from other "threshold" or "critical mass" models (Schelling 1978, Granovetter 1978, Marwell and Oliver 1988). These models typically presume a simple behavioral rule of individuals, in which they are programmed to participate in collective actions whenever the number of participants exceeds their thresholds. On the other hand, individuals behave strategically in our model. That is, even if the number of participants exceeds their thresholds, they do not automatically participate, since their best responses in such a situation might be to free ride.

The paper is organized as follows. Section 2 presents a group formation game. Section 3 provides two characterizations of strict Nash equilibria of the game. It is shown that the group formation game is weakly acyclic. Section 4 discusses extensions of the model to cases of multiple groups and of general payoff functions. Applications to

[3]It will be shown that the group of all individuals can form in equilibrium if there exist at least two individuals whose thresholds of cooperation is equal to the number of all individuals.

voluntary contribution to public goods and to cartel formation in oligopolistic firms are presented. Section 5 concludes.

2 The Model

Consider an n-person prisoner's dilemma as follows. Let $N = \{1, 2, \cdots, n\}$ be the set of players. Every player $i \in N$ has two actions, C (cooperation) and D (defection). Player i's payoff is given by

$$u_i(a_i, h),\ \ a_i = C, D,\ \ h = 0, 1, \cdots, n-1, \tag{2.1}$$

where a_i is player i's action and h is the number of other players who select C. We make the following assumption.

Assumption 2.1. The payoff function of player $i \in N$ satisfies:

(1) $u_i(D, h) > u_i(C, h)$ for every $h = 0, 1, \cdots, n-1$,

(2) $u_i(C, n-1) > u_i(D, 0)$,

(3) $u_i(C, h)$ and $u_i(D, h)$ are increasing in h.

This assumption is standard in the literature of an n-person prisoner's dilemma (Schelling 1978), except that players are "heterogeneous" in the sense that they have different payoff functions. The heterogeneity of players is critical to the analysis of this paper. Property (1) means that every player is better off by defecting than cooperating, regardless of others' play. This implies that every player has an incentive to free ride on others' cooperative action. Thus, the action profile $(D, \cdots, D)$ is a unique Nash equilibrium of the game. On the other hand, property (2) says that if all players cooperate, they are all better off than in the unique equilibrium. Namely, the equilibrium is not Pareto efficient. Property (3) means that the more others cooperate, the higher payoff each player can receive, regardless of her action. The cooperative action by any player has positive externality on all others.

The prisoner's dilemma describes an anarchic situation in which players are free to choose their actions. In such a situation, a natural outcome of the game is the Nash equilibrium in which no one cooperates. The central question concerning the prisoner's dilemma is whether and how self-interested individuals voluntarily cooperate under the presence of temptations to defect. To escape from an undesirable state of non-cooperation, some suitable mechanisms for preventing opportunistic behavior are needed. The literature has considered diverse mechanisms attaining cooperation. However, as we have discussed in the introduction, any mechanism of cooperation itself is a kind of public good. Every individual has an incentive to free ride on the mechanism. To consider the problem of voluntary participation in the mechanism of cooperation in a pure form, we formulate the following game of group formation.

The process of group formation is defined as a two-stage game.

Participation decision stage: Every player $i \in N$ decides independently whether or not to participate in a group, to negotiate on the collective action. Participation takes some costs, say, for phone calls, mails and transportations. The participation cost is denoted by a small positive amount $\varepsilon_i > 0$. Let S be the set of all s participants. If $s = 0$ or $s = 1$, then no group is possible.[4]

Group negotiation stage: All participants negotiate on their cooperation according to the unanimity rule, knowing the outcome of the participation decision stage. Each of them either accepts or rejects independently to cooperate. The agreement of cooperation is reached if and only if all participants accept it. When the agreement of cooperation is reached, it is enforced and all participants are bound to cooperate. The enforcement is costly. Every participant i bears the group costs $c_i(s)$ (including participation costs ε_i) where s is the number of all participants. All non-participants are free to defect. When an agreement is not reached, all individuals, both participants and non-participants, play the original prisoner's dilemma.

When all participants agree to cooperate, we assume that the group is endowed with some mechanism to enforce the agreement. The mechanism has various functions such as

[4]When $s = 1$, the single participant has no incentive to cooperate in the prisoner's dilemma.

monitoring members' actions and punishing defecting members. Obviously, it is costly for group members to have such a mechanism, and a cost allocation problem may arise. In what follows, to keep our game model as simple as possible, we do not present a formal model of cost allocations in a group. Rather, we simply assume that the group cost born by each member is exogenously given, and formulate it by a group cost function $c_i(s)$. We remark that the precise form of $c_i(s)$ is irrelevant to the result of the paper as long as the following assumption 2.2 is held.

In the process of group formation, individuals decide whether they should participate in a group or not, anticipating rationally the outcome of the group negotiation stage. To consider the participation decision of individuals, we characterize a subgame perfect equilibrium of the two-stage game of group formation. By backward induction, we first analyze the group negotiation stage. When a group of s members agree to cooperate, every member receives utility

$$g_i(C, s-1) \equiv u_i(C, s-1) - c(s). \tag{2.2}$$

We call $g_i(C, s-1)$ the *group payoff* of player i where s is the number of group members. Concerning the group payoff, we assume the following property.

Assumption 2.2. For every $i \in N$, the group payoff $g_i(C, s-1)$ of player i is monotonically increasing in s, and there exists a unique integer s_i $(2 \leq s_i \leq n)$ such that

$$g_i(C, s_i - 2) < u_i(D, 0) < g_i(C, s_i - 1). \tag{2.3}$$

This assumption means that even if we replace the original cooperative payoff $u_i(C, h)$ with the group payoff $g_i(C, h)$, the properties (Assumption 2.1) of the n-person prisoner's dilemma still hold true. If Assumption 2.2 does not hold, the problem of group formation becomes trivial. For example, if $g_i(C, s-1) \leq u_i(D, 0)$ for all $s \leq n$, then no players have incentive to participate in a group. The positive integer s_i in (2.3) shows the minimum size of a group in which member i can be better off than in the non-cooperative equilibrium of the prisoner's dilemma. We call s_i the *threshold of cooperation* of i. Player i can benefit by cooperating whenever at least $(s_i - 1)$ others cooperate. In this sense, players with smaller thresholds have higher motivation to cooperate. The threshold of cooperation plays a critical role in our analysis.

Definition 2.1. A subset S of N is called a *successful group* if $|S| \geq s_i$ for every $i \in S$.

The size of a successful group is greater than or equal to all members' thresholds of cooperation. By definition, every member of a successful group can receive higher payoff than the non-cooperative payoff in the prisoner's dilemma. The naming of a successful group is explained by the following proposition which characterizes a strict Nash equilibrium of the group negotiation stage. In a strict Nash equilibrium, every player's best response to all other players' actions is unique.

Proposition 2.1. *In the group negotiation stage, an agreement of cooperation is reached in a strict Nash equilibrium if and only if the group of participants is successful.*[5]

Proof. Suppose that all s participants agree to cooperate. Then, every participant receives the group payoff $g_i(C, s-1)$. If any member rejects to cooperate, the negotiation breaks down by the unanimity rule, and she receives the non-cooperative payoff $u_i(D, 0)$. Therefore, the agreement of cooperation in a group S is reached in a strict Nash equilibrium if and only if for all $i \in S$, $g_i(C, |S|-1) > u_i(D, 0)$. From assumption 2.2, this is equivalent to that the group is successful. □

Given the (strict) Nash equilibrium of the group negotiation stage, the whole two-stage game can be reduced to the following one-stage game. In this game, every player $i \in N$ chooses simultaneously and independently either $\sigma_i = 1$ (participation) or $\sigma_i = 0$ (non-participation). Let $\Sigma_i = \{0, 1\}$ be the set of actions of player i, and let $\Sigma = \prod_{i \in N} \Sigma_i$ be the set of action profiles of n players. In the following analysis, we do not consider

[5] There exist many "trivial" non-strict Nash equilibria leading to the disagreement. For example, any action profile where at least two participants reject to cooperate is such an equilibrium. These equilibria are peculiar to the unanimity rule where everyone has a veto power. We remark that from the viewpoint of each participant the action of agreement (weakly) dominates that of disagreement. For this reason, we only consider strict Nash equilibria leading to the agreement of cooperation in every successful group.

randomized actions (mixed strategies). For an action profile $\sigma = (\sigma_1, \cdots, \sigma_n) \in \Sigma$, the set $S(\sigma)$ of participants is given by

$$S(\sigma) = \{i \in N | \sigma_i = 1\}.$$

The payoff $f_i(\sigma)$ of player i for an action profile $\sigma = (\sigma_1, \cdots, \sigma_n) \in \Sigma$ is defined as follows.

(i) When a group $S(\sigma)$ of participants is successful,

$$f_i(\sigma) = \begin{cases} g_i(C, |S(\sigma)| - 1) & \text{if } \sigma_i = 1, \\ u_i(D, |S(\sigma)|) & \text{if } \sigma_i = 0. \end{cases}$$

(ii) When $S(\sigma)$ is not successful,

$$f_i(\sigma) = \begin{cases} u_i(D, 0) - \varepsilon_i & \text{if } \sigma_i = 1, \\ u_i(D, 0) & \text{if } \sigma_i = 0, \end{cases}$$

where $\varepsilon_i > 0$ is a participation cost.

Formally, the process of group formation reduces to the n-person game $\Gamma = (N, \{\Sigma_i, f_i\}_{i \in N})$ in strategic form. We call it the *group formation game*. The group formation game Γ differs from the original n-person prisoner's dilemma in critical ways. In Γ, non-participation does not dominate participation, nor vice versa. By this reason, a non-empty group of participants may arise in equilibrium, in which non-participants free ride on the group action. Participants, however, can guarantee their equilibrium payoffs in the prisoner's dilemma. In this sense, cooperators are never exploited by free-riders in the group formation game.

3 The Nash Equilibria in the Group Formation Game

In this section, we characterize the Nash equilibria of the group formation game Γ. For an action profile $\sigma = (\sigma_1, \cdots, \sigma_n) \in \Sigma$, let σ_{-i} be the action profile obtained from σ

by deleting σ_i. As usual, an action profile $\sigma = (\sigma_1, \cdots, \sigma_n)$ is sometimes denoted by $\sigma = (\sigma_{-i}, \sigma_i)$. Let $S(\sigma)$ be the set of participants in σ.

Definition 3.1. For an action profile $\sigma = (\sigma_{-i}, \sigma_i) \in \Sigma$ in Γ, player i's action σ_i is called a *best response* to σ if $f_i(\sigma_{-i}, \sigma_i) = \max_{\sigma'_i \in \Sigma_i} f_i(\sigma_{-i}, \sigma'_i)$. An action profile $\sigma = (\sigma_1, \cdots, \sigma_n)$ is a *(strict) Nash equilibrium* of Γ if every σ_i is a (uniquely) best response to σ.

Definition 3.2. For a successful group S, member i of S is called *critical* if $S \setminus \{i\}$ is not successful, and *non-critical* otherwise.

In a successful group, no critical member has an incentive to deviate from the group since the remaining group is not successful and so the deviation gives her the noncooperative payoff in the prisoners' dilemma. On the contrary, every non-critical member has an incentive to deviate from the group because, by doing so, she can free ride on the remaining (successful) group. In an unsuccessful group, every participant has an incentive to deviate from the group for saving participation costs. A non-participant has an incentive to join the unsuccessful group if her participation makes the group successful. By these incentives of individuals, we can characterize strict Nash equilibria of the group formation game Γ.

Theorem 3.1. *The group formation game Γ has the following strict Nash equilibria $\sigma = (\sigma_1, \cdots, \sigma_n)$ with the set $S(\sigma)$ of participants.*

(1) $\sigma = (0, \cdots, 0)$, *i.e.,* $S(\sigma) = \varnothing$.

(2) $S(\sigma)$ *is a successful group with every member of it being critical.*

Proof. (1) For $\sigma = (0, \cdots, 0)$, every individual i's payoff is $u_i(D, 0)$. If she unilaterally participates in a group, she receives a smaller payoff $u_i(D, 0) - \varepsilon_i$ since any group with only one member is not successful. Thus, $\sigma = (0, \cdots, 0)$ is a strict Nash equilibrium.

(2) Suppose that $S(\sigma)$ is a successful group with every member of it being critical. We will show that every individual is worse-off if she deviates unilaterally from σ. If any participant i deviates from the group $S(\sigma)$, then the remaining group $S(\sigma) \setminus \{i\}$

is not successful by assumption, and thus she will receive payoff $u_i(D,0)$. Since the group $S(\sigma)$ is successful, $u_i(D,0)$ is smaller than $f_i(\sigma) = u_i(C, |S(\sigma)| - 1)$. In σ, every non-participant i receives payoff

$$f_i(\sigma) = u_i(D, |S(\sigma)|).$$

If she participates in the group $S(\sigma)$, her payoff will be

$$f_i(\sigma_{-i}, 1) = u_i(C, |S(\sigma)|) \ \text{ or } \ u_i(D,0) - \varepsilon_i,$$

depending on whether or not the enlarged group $S(\sigma) \cup \{i\}$ is successful. In any case, non-participant i's payoff will be smaller than $f_i(\sigma)$.

(3) It remains to show that there exist no other strict Nash equilibria of the group formation game. Suppose that $S(\sigma)$ is a successful group where some member i is not critical. Since the group $S(\sigma) \setminus \{i\}$ remains successful, we have

$$f_i(\sigma_{-i}, 1) = v_i(C, |S(\sigma)| - 1) < u_i(D, |S(\sigma)| - 1) = f_i(\sigma_{-i}, 0).$$

Therefore, $\sigma_i = 0$ is a best response to σ for non-critical member i of $S(\sigma)$. Finally, suppose that $S(\sigma)$ is not a successful group. Then, for all participants $i \in S$,

$$f_i(\sigma_{-i}, 1) = u_i(D,0) - \varepsilon_i < u_i(D,0) \leq f_i(\sigma_{-i}, 0),$$

where $f_i(\sigma_{-i}, 0)$ is equal to either $u_i(D, |S(\sigma)| - 1)$ or $u_i(D,0)$, depending on whether the remaining group except player i is successful or not.

$\square$

Although the logic supporting the theorem is simple and intuitive, it can be helpful for us to explain it by an alternative definition of a Nash equilibrium in the group formation game. A group of participants in a Nash equilibrium satisfies two stability properties:

Internal stability: No single member wants to opt out of the group.

External stability: No single outsider wants to join the group.

It is clear that the action profile $\sigma = (0, \cdots, 0)$ is a Nash equilibrium because no one is willing to cooperate unilaterally. Furthermore, this is a strict Nash equilibrium when participation cost is positive. When a group is not successful, the internal stability is violated because all participants want to opt out of the group for saving participation costs. When a group is successful, the external stability always holds because all non-participants have incentive to free ride. If there exists any non-critical member, then she has an incentive to opt out of the group since the remaining group remains successful. Thus, the internal stability implies that every participant is critical.

We can derive another characterization of Nash equilibria in the group formation game. It is given in terms of players' thresholds of cooperation. For $S \subset N$ and $m = 2, \cdots, n$, we define $F_S(m)$ by the number of all members in S whose thresholds of cooperation are given by m. That is, $F_S(m) = |\{ i \in S \mid s_i = m \}|$. F_S represents the distribution of members' thresholds of cooperation in the group S. Its definition implies that:

Lemma 3.1. *For $S \subset N$,*

(1) $F_S(2) + \cdots + F_S(|S|) \leq |S|$.

(2) *A group S is successful if and only if $F_S(2) + \cdots + F_S(|S|) = |S|$.*

Theorem 3.1 shows two conditions for a (nonempty) group of participants to be the result of a strict Nash equilibrium in a group formation game: (i) it is successful, and (ii) every member is critical. The latter condition has a simple formulation in terms of the threshold of cooperation as follows.

Theorem 3.2. *A nonempty subset S of N is the set of participants in a strict Nash equilibrium of the group formation game Γ if and only if*

$$F_S(2) + \cdots + F_S(|S|) = |S| \quad \textit{and} \quad F_S(|S|) \geq 2.$$

Proof. From Theorem 3.1 and Lemma 3.1, it suffices to show that every member of a successful group S is critical if and only if $F_S(|S|) \geq 2$. Suppose that $F_S(|S|) \geq 2$. For every $i \in S$, the group $S \setminus \{i\}$ is not successful because $F_{S \setminus \{i\}}(|S|) \geq 1$. Thus, every member i of S is critical. If $F_S(|S|) = 1$, then a unique member i with $s_i = |S|$ is not critical because $S \setminus \{i\}$ is a successful group. If $F_S(|S|) = 0$, all members j of S have thresholds s_j of cooperation with $s_j \leq |S| - 1$. Therefore, they are not critical. $\square$

The theorem enables us to see clearly how the heterogeneity of individuals affects the group formation in collective action. When individuals are so homogeneous that they have identical thresholds, the size of an equilibrium group is uniquely determined by the common threshold. In contrast, when individuals are heterogeneous, there typically exist many equilibrium groups with different sizes. For example, if all individuals' thresholds of cooperation are "widely" distributed over a range $Z = \{2, 3, \cdots, m_1\}$ $(2 \leq m_1 \leq n/2+1)$ so that $F_N(s) \geq 2$ for every integer s in Z, then groups of all sizes s in the range Z can be formed in equilibrium. The heterogeneity of individuals causes the multiplicity of equilibrium groups. Different equilibria may reveal different levels of cooperation from partial cooperation to full cooperation.

The heterogeneous thresholds also affect the likelihood of the full cooperation. In the homogeneous case, the group of n players can be formed under a stringent condition that all individuals' thresholds of cooperation are equal to the number n of players. To put it differently, the full cooperation can be attained among homogeneous individuals only if they happen to have the largest possible thresholds, namely $F_N(n) = n$. In contrast, the largest group can be sustained in equilibrium among heterogeneous individuals under a much weaker condition that $F_N(n) \geq 2$.

With help of Theorem 3.2, we can prove the existence of a strict Nash equilibrium of the group formation game with (possibly partial) cooperation.

Theorem 3.3. *There exists a strict Nash equilibrium of the group formation game* Γ *where a non-empty group of individuals forms.*

Proof. With no loss of generality, we assume that $s_1 \leq s_2 \leq \cdots \leq s_n$ where s_i is the threshold of cooperation for individual i. Let N^* be the set of integers $m \in N$ such that there exist at least two individuals $i \in N$ such that $s_i = m$. That is, $N^* = \{m \in$

$N|F_N(m) \geq 2\}$. Since $s_i \geq 2$ for every $i \in N$ by assumption 2.1, N^* is not empty. Let m^* be the maximum integer of N^*, and let $M^* = \{1, \cdots, m^*\}$. If $F_{M^*}(2) + \cdots + F_{M^*}(m^*) \geq m^*$, then we can choose a subset U of M^* satisfying $F_U(2) + \cdots + F_U(m^*) = m^*$ and $F_U(m^*) = 2$. Then, it follows from Theorem 3.2 that there exists a strict Nash equilibrium of Γ in which all individuals in U participate in a group . On the contrary, suppose that $F_{M^*}(2) + \cdots + F_{M^*}(m^*) < m^*$. Then, it must be true that there are more that $n - m^*$ individuals i such that $s_i = m^* + 1, \cdots, n$. This implies that there exists some integer $t > m^*$ with $F_N(t) \geq 2$. This contradicts the definition of m^*. □

Let us illustrate the results by means of a numerical example. For simplicity, we assume that group costs are zero. That is, $g_i = u_i$.

Example 3.1. Consider a seven-person prisoner's dilemma with the player set $N = \{1, \ldots, 7\}$. Figure 1 defines payoffs. For each player i, the payoff by defection is commonly given in row $u(D, h)$, where h $(= 0, 1, \ldots, 6)$ is the number of other cooperators. Payoffs by cooperation are heterogeneous. Player i's payoff by cooperation is shown in row $u_i(C, h)$. The players' thresholds of cooperation are given by $s_1 = 2$, $s_2 = s_3 = s_4 = 4$, $s_5 = 6$, and $s_6 = s_7 = 7$. Player 1 is the most willing to cooperate. She has an incentive to cooperate if only one other player cooperates. Players 6 and 7 are the most reluctant to cooperate. They will cooperate only if all other players cooperate. Players $i = 2, \cdots, 5$ are in intermediate positions. The distribution F_N of players' thresholds of cooperation is given in Figure 2. Theorems 3.1 and 3.2 show that there exist three strict Nash equilibria: (1) (no cooperation) no players participate in a group, (2) (partial cooperation) Four players 1, 2, 3, 4 participate in a group, and (3)(full cooperation) All players participate in a group.

To conclude this section, we show an important property of the group formation game Γ. The property says that, starting from any action profile, some strict Nash equilibrium can be reached by successive plays of individuals' best responses. To be precise, we need the following definitions (see Young (1993)).

h	0	1	2	3	4	5	6
$u(D,h)$	0	2	4	6	8	10	15
$u_1(C,h)$	-1	1	2	3	4	5	6
$u_2(C,h)$	-3	-2	-1	1	2	3	4
$u_3(C,h)$	-3	-2	-1	1	2	3	4
$u_4(C,h)$	-3	-2	-1	1	2	3	4
$u_5(C,h)$	-5	-4	-3	-2	-1	1	2
$u_6(C,h)$	-6	-5	-4	-3	-2	-1	1
$u_7(C,h)$	-6	-5	-4	-3	-2	-1	1

Figure 1: A seven-person prisoner's dilemma.

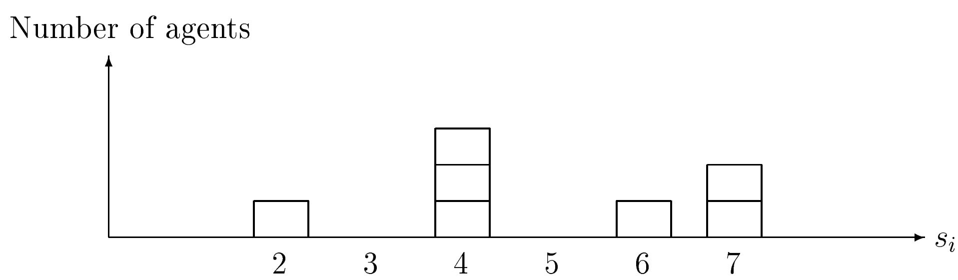

Figure 2: Threshold distribution.

Definition 3.3. (1) The *best response graph* V of Γ is a binary relation on the set of action profiles Σ such that, for every σ, $\sigma' \in \Sigma$, $(\sigma, \sigma') \in V$ if and only if there exists one player i such that (i) $\sigma_i \neq \sigma'_i$ and $\sigma_{-i} = \sigma'_{-i}$, and (ii) σ'_i is a best response to σ for i. When $(\sigma, \sigma') \in V$, we write $\sigma \to \sigma'$ and call it an *edge* from σ to σ'.

(2) Γ is called *weakly acyclic* if for every action profile σ^1, there is a path of action profiles $\sigma^1 \to \cdots \to \sigma^L$ in the best response graph V such that σ^L is a strict Nash equilibrium of Γ.

The weak acyclicity of the group formation game Γ intuitively means the following situation. Suppose that Γ is played repeatedly where the opportunity of changing actions may be given to every player with positive probability in every round. Suppose furthermore that whenever the opportunity of adjustment comes, every player responds to the current action profile myopically and rationally. Then, starting from any action profile, the game play can reach some strict Nash equilibrium with positive probability. Once a strict Nash equilibrium is reached, it is played forever, that is, it is an absorbing state of the process. The notion of weak acyclicity plays a critical role in the stochastic evolutionary theory developed by Young (1993) and Kandori, Mailath and Rob (1993).

Theorem 3.4. *The group formation game* Γ *is weakly acyclic.*

Proof. By the same proof as in (3) of Theorem 3.1, we can show that an edge of the best response graph V of Γ must be one of the following types. Let $\sigma \in \Sigma$.

(1) Assume that $S(\sigma)$ is successful. For every $\sigma' \in \Sigma$, $\sigma \to \sigma'$ if and only if there is a non-critical $i \in S(\sigma)$ such that $\sigma = (\sigma_{-i}, 1)$ and $\sigma' = (\sigma_{-i}, 0)$.

(2) Assume that $S(\sigma)$ is not successful. For every $\sigma' \in \Sigma$, $\sigma \to \sigma'$ if and only if either $\sigma = (\sigma_{-i}, 1)$ and $\sigma' = (\sigma_{-i}, 0)$ for some $i \in S(\sigma)$ or $\sigma = (\sigma_{-i}, 0)$ and $\sigma' = (\sigma_{-i}, 1)$ for $i \notin S(\sigma)$ such that $S(\sigma) \cup \{i\}$ is successful.

Take σ^1 and assume first that it is successful. By Theorem 3.1, we can assume that there are non-critical members in $S(\sigma^1)$. (1) implies that $\sigma^1 \to \sigma^2$ iff $\sigma^2 = (\sigma^1_{-i}, 0)$, where $i \in S(\sigma^1)$ is not critical. Since i is not critical, $S(\sigma^2)$ is successful. By induction, in any sequence $\sigma^1 \to \cdots \to \sigma^L$ in V, $S(\sigma^l)$ is successful for every $l = 1, \ldots, L$ and the number of players who choose action 1 is strictly decreasing. Thus the length of such a sequence is at most $|S(\sigma^1)| + 1$. Take the longest such sequence $\sigma^1 \to \cdots \to \sigma^{L^*}$. Then σ^{L^*} is successful and every member of $S(\sigma^{L^*})$ is critical. That is, σ^{L^*} is a strict Nash equilibrium.

Assume next that σ^1 is not successful and take a sequence $\sigma^1 \to \cdots \to \sigma^L$ in V. If there is l ($l = 2, \ldots, L$) such that σ^l is successful, then the arguments in the above paragraph applies. Thus we can assume that every σ^l is not successful. By (2), $\sigma^{l+1} = (\sigma^l_{-i}, 0)$ for $l = 1, \ldots, L-1$. Thus the number of 1-players is strictly decreasing, and the length of such a sequence is at most $L^* = |S(\sigma^1)| + 1$. Take the longest such sequence $\sigma^1 \to \cdots \to \sigma^{L^*}$. It is clear that σ^{L^*} is the non-cooperation equilibrium. $\square$

4 Extensions and Applications

We have presented a two-stage game model to investigate the formation of a voluntary group in a collective action problem in a simple and pure form. The n-person prisoners' dilemma has been considered extensively as a basic model of a collective action problem in the literature. In this section, we discuss how we can relax some restricted assumptions in the model.

4.1 Multiple Groups

In the group formation game Γ, it is assumed that individuals are allowed to form only one group for collective actions. When one group forms, any non-participant does not have an incentive to joint it, as we have shown. On the other hand, non-participants may have an incentive to form the second group. Consider an example of a four-person prisoners' dilemma where the thresholds of cooperation are two for all players.[6] As we have shown, only a two-person group can be formed, and the two other persons free ride. But, since the thresholds of cooperation for non-participants are two, it may be profitable for them to form another two-person group, instead of free-riding. In real situations, there often exist multiple groups each of which exhibits positive externality on others. For example, several groups form to protect environmental pollution in a community. In a faculty of a university, different groups of members organize different seminars to activate their research.

To consider a possibility of multiple groups, we extend the two-stage game model of group formation presented in section 2, and consider an additional stage of choosing locations at the start of the game. The extended model has three-stages as follows.

Location choice stage: Every player $i \in N$ chooses independently one of locations $k = 1, \cdots K$ where $K(> 1)$ is a fixed integer. For each $k = 1, \cdots K$, let $N_k(\subset N)$ denote the set of individuals who choose location k where $N_1 \bigcup \cdots \bigcup N_K = N$ and all N_i and

[6] This example and the model discussed in this section was suggested by an anonymous referee. We are grateful to his/her helpful suggestions.

N_j $(i \neq j)$ are disjoint.

Participation decision stage/Group negotiation stage: In every location $k = 1, \cdots K$, the group formation game Γ with the player set N_k is played independently. The payoff of every individual is given by (2.1) and (2.2) where the number of cooperators are counted over all locations.

We denote this extended model of group formation by Γ^*. Γ^* is a multi-stage game model with perfect information. In every stage, all individuals choose their actions independently with perfect knowledge of the outcomes in all previous stages.

We examine under what conditions the formation of multiple groups is possible. To do this, we need to generalize the notion of the threshold of cooperation as follows.

Definition 4.1. For every $i \in N$, the *generalized threshold of cooperation* for individual i is a function s_i^* which assigns to every $m \in \{0, 1, \cdots, n-2\}$ the minimum integer $s_i^*(m)$ $(2 \leq s_i^*(m) \leq n - m)$ such that

$$g_i(C, m + s_i - 2) < u_i(D, m) < g_i(C, m + s_i - 1). \tag{4.1}$$

In the extended game Γ^*, individuals in every location $k = 1, \cdots K$ play the group formation game, given the number of cooperations in all other locations. For $m \in \{0, 1, \cdots, n-2\}$, the generalized threshold $s_i^*(m)$ of cooperation for individual i means the minimum number of cooperators (including i) so that i is better-off than in the outcome of no cooperation, given the number m of cooperators.

The *cooperation structure* τ in the extended game Γ^* shows who choose each location and who participate in a group in each location. Formally, τ is represented by $[(N_1, S_1), \cdots, (N_K, S_K)]$ where $(N_1, \cdots, N_K)$ is a partition of N and each S_k is a subset of N_k. For every $k = 1, \cdots K$, N_k denotes the set of individuals who choose location k and S_k denotes the set of individuals who participate in a group in location k. With help of Theorem 3.1, we can prove the equilibrium condition of multiple groups in the extended game Γ^*.

Theorem 4.1. *There exists a subgame perfect equilibrium of the extended game* Γ^* *generating a cooperation structure* $\tau = [(N_1, S_1), \cdots, (N_K, S_K)]$ *if and only if for every* *location* $k = 1, \cdots K$ *with* $S_k \neq \emptyset$,

$$F_k^\tau(2) + \cdots + F_k^\tau(|S_k|) = |S_k| \quad \textit{and} \quad F_k^\tau(|S_k|) \geq 2,$$

where $F_k^\tau(s)$ *is the number of individuals* i *in* N_k *satisfying*

$$s_i^*(|S_1| + \cdots + |S_{i-1}| + |S_{i+1}| + \cdots + |S_K|) = s.$$[7]

Proof. The only-if-part can be proved by the same way as in Theorem 3.2. In order to prove the if-part, we will construct a subgame perfect equilibrium of the extended game Γ^* which generates the cooperation structure $\tau = [(N_1, S_1), \cdots, (N_K, S_K)]$. For every $k = 1, \cdots K$, define the following strategies in Γ^*.

(1) All individuals in N_k choose location k.

(2) In every location k with $S_k \neq \emptyset$, all individuals in S_k participate in a group and any other individual does not participate if and only if the set of individuals who choose location k is equal to N_k. Otherwise, no individuals participate. In every location k with $S_k = \emptyset$, no individuals participate in a group.

(3) In location k, all individuals in a group agree to cooperate if and only they are better-off by doing so than in the no-cooperation outcome, given the number of cooperators in all other locations.

Let $h = (h_1, \cdots, h_n)$ denote the strategy profile for individuals in Γ^*. It is clear that h generates the cooperation structure $\tau = [(N_1, S_1), \cdots, (N_K, S_K)]$. By the same proof as of Theorem 3.1, we can show without much difficulty that h induces a strict Nash equilibrium for the group formation game in every location k. Therefore, it suffices us to show that the location choice of every individual is a best response to all other individuals' location choice. For every location k, consider first every participant i. If she chooses any location $k'(\neq k)$, the strategy h prescribes that there are no groups in locations k and k' and all groups in all other locations remain to form. Therefore,

[7] Remark that the distribution F_k^τ of generalized thresholds in N_k depends on the cooperation structure τ.

participant i is worse-off by deviation from h. The same arguments can be applied to every non-participant. $\square$

The theorem shows that multiple groups may form in the voluntary formation game when free-riders may find it beneficial to form new subgroups among themselves. In the theorem, the multiple groups are supported by a punishing behavior of cooperators that they dissolve groups either if any new individuals come in their locations, or if any incumbents go out of their locations. Although the multiple groups enhance the level of cooperation in collective actions, we note that the multiplicity of strict Nash equilibria remains unchanged (or becomes even more) in the extended model Γ^* with location choice. The no cooperation equilibrium still arises.

Application 4.1. Consider a voluntary provision game of public goods as follows. Every individual $i = 1, \cdots, n$ has initial endowments w and decides her contribution g_i to public goods where $0 \leq g_i \leq w$. For a contribution profile $(g_1, \cdots, g_n)$ of individuals, the payoff of individual i is given by

$$u_i(g_1, \cdots, g_n) = \omega - g_i + a_i \sum_{j=1}^{n} g_j, \quad a_i < 1 < na_i,$$

where a_i is the marginal per capita return (MPCR) for individual i from a contribution to the public good. The condition $a_i < 1$ means that zero contribution is the dominant action for every individual i. The condition $1 < na_i$ means that the full contribution profile $g = (\omega, \cdots, \omega)$ Pareto-dominates the Nash equilibrium $g = (0, \cdots, 0)$. For simplicity, we assume that group cost is zero. Then, the threshold of cooperation for individual i is given by the minimum integer s_i^* such that $sa_i\omega > \omega$, that is, $s > 1/a_i$. The threshold of cooperation increases if the MPCR a_i decreases. Individuals with lower MPCRs need a larger number of cooperators to become willing to cooperate. Owing to the linearity of the payoff function, the threshold s_i^* of cooperation for individual i is independent of the number of cooperators in other locations. Consider a symmetric case that all individuals have the common thresholds, say $s^*(< n)$, of cooperation. When there is the only one opportunity to form a group, Theorem 3.1 shows that only s^* individuals contribute.

However, when there is no limit of opportunities to form groups, Theorem 4.1 shows that multiple s^*-person groups for voluntary contribution may form in equilibrium with the number of free-riders smaller than s^*.

4.2 General Payoff Functions

Another restrictive assumption in our model is that the payoff functions of individuals depend on only the number of cooperators as well as their actions. In a general situation, however, what is critical to individuals' decision to cooperate or not is with whom they cooperate, not simply how many others cooperate. To consider this kind of situation with generalized payoff functions, we extend the payoff functions (2.1) of individuals to

$$u_i(a_i, S), \ a_i = C, D, \ S \subset N \setminus \{i\}, \tag{4.2}$$

where S is the set of cooperators except i. Assumption 2.1 can be easily extended to this generalized payoff function. Regarding Assumption 2.1.(3), the increase in the number h of other cooperators is replaced with the growing of the set S of other cooperators in terms of set inclusion. The group payoff (2.2) is extended to

$$g_i(C, S \setminus \{i\}) = u_i(C, S \setminus \{i\}) - c_i(S)$$

where $c_i(S)$ is the group cost for $i \in S$.

In this general setup, the thresholds of cooperation for individuals are not simply defined as an integer. The notion of a successful group, however, can be defined in the same manner as in Definition 2.1.

Definition 4.2. A subset S of N is called a *successful group* if $g_i(C, S \setminus \{i\}) > u_i(D, \varnothing)$ for every $i \in S$.

It can be easily shown that Theorem 3.1 still holds under the generalized payoff function (4.2). That is, a group can be formed in a strict Nash equilibrium of the group formation game if and only if it is a successful group in which every member is critical.

As an example of a collective action problem with the generalized payoff function, consider the cartel formation in the Cournot oligopolistic market.

Application 4.2. Consider a linear Cournot game with asymmetric costs as follows. Let $N = \{1, \cdots, n\}$ denote the set of firms. Firms produce homogeneous goods with a linear (inverse) demand function, $p = 1 - \sum_{i=1}^{n} x_i$ (after normalization), where p is the price of the goods and x_i is the output of firm i. Firm i has a linear cost function $c_i(x_i) = c_i x_i$. For an output profile $x = (x_1, \cdots, x_n)$, the profit $\pi_i(x)$ for firm i is given by $\pi_i(x) = (p - c_i)x_i$.

It is well-known that the linear Cournot game has a unique Nash equilibrium (assuming an interior solution) as follows.

- firm i's output: $x_i^* = \frac{1 - nc_i + \sum_{j \neq i} c_j}{n+1}$
- firm i's profit: $\pi_i(x^*) = \frac{(1 - nc_i + \sum_{j \neq i} c_j)^2}{(n+1)^2}$
- price: $p = \frac{1 + \sum_{i=1}^{n} c_i}{n+1}$.

We assume that firms are free to form a cartel. To simplify the analysis, we further assume that if a cartel of firms forms, the total production is delegated to the member firm of which unit cost is the lowest among its members. The total profit for the cartel is distributed among members. Suppose that the members of a cartel is given by a subset S of N. Let c_S denote the unit cost for the cartel S, i.e., $c_S = \min\{c_i | i \in S\}$. Then, by the formula above, we can see that the Cournot game with cartel S has a unique Nash equilibrium in which the profits of firms are given as follows.

- the total profit of cartel: $\pi^S = \frac{(1 - mc_S + \sum_{j \notin S} c_j)^2}{(m+1)^2}$
- non-member i's profit: $\pi_i^{-S} = \frac{(1 - mc_i + \sum_{j \notin S, j \neq i} c_j + c_S)^2}{(m+1)^2}$

where $m = n - |S| + 1$. The profit π^S of cartel corresponds to the sum of group payoffs $g_i(C, S \setminus \{i\})$ for $i \in S$ in the general setup (assuming zero group cost), and the non-member i's profit π_i^{-S} corresponds to the free rider's payoff $u_i(D, S \setminus \{i\})$. With this in mind, we can apply our results to the cartel formation in the Cournot game.

As a benchmark, we first consider a symmetric case where all firms have the common unit cost c. We assume that the total profit π^S is distributed equally among cartel members. In this case, member i's profit π_i^s is given by

$$\pi_i^S = \frac{(1-c)^2}{(n-s+2)^2 s} \ ,$$

and non-member i's profit π_i^{-S} is given by

$$\pi_i^{-S} = \frac{(1-c)^2}{(n-s+2)^2} \ .$$

Observe that non-members of the cartel enjoy higher profits than members. They can free ride on the coordination of outputs in the cartel. The threshold s^* of cooperation for a firm is given by the minimum integer s such that

$$\pi_i^S = \frac{(1-c)^2}{(n-s+2)^2 s} > \frac{(1-c)^2}{(n+1)^2}.$$

where the left-hand side is the firm's Cournot equilibrium profit in the case of no cartel. Rearranging the inequality above, we can see that the threshold s^* of cooperation is the minimum integer satisfying $(n+1)^2 > (n-s+2)^2 s$. It can be shown after some manipulation that the threshold s^* of cooperation is given by

$$s^* = n + \frac{3}{2} - \sqrt{n + \frac{5}{4}}$$

(ignoring the integer constraint). By this formula, we can see that the cartel of all firms form for $n = 2, 3, 4, 5$ and for all n (> 5), the equilibrium size of a cartel is less than n (but larger than $n/2$).

Finally, we discuss how the heterogeneity of firms affects cartel formation with help of a numerical example. Consider three firms 1, 2, 3 with unit costs $c_1 = c_2 = 0.1, c_3 = 0.2$. By the formula of firms' profits π_i with a cartel given above, we can observe firms' profits under various cartels:

$$[\{1\},\ \{2\},\ \{3\}] : \ \pi_1 = \pi_2 = 1/16 = 0.063,\ \ \pi_3 = 0.023$$

$$[\{1,\ 3\},\ \{2\}] : \ \pi_1 + \pi_3 = 0.090,\ \ \pi_2 = 0.090$$

$$[\{1,\ 2\},\ \{3\}] : \ \pi_1 + \pi_2 = 1/9 = 0.111,\ \ \pi_3 = 0.054$$

$[\{1,\,2,\,3\}]:\ \pi_1 + \pi_2 + \pi_3 = 0.203$

where a set of more than one firm represents a cartel. Suppose that firms 1 and 3 form a cartel. Since the total profit 0.090 of the cartel is larger than the sum 0.89 of their Cournot equilibrium profits, the cartel $\{1, 3\}$ is successful. In contrast to this, the cartel $\{1, 2\}$ is not successful since the total profit 0.111 of firms 1 and 2 is less than the sum 0.126 of their Cournot equilibrium profits. There is no merit for firms 1 and 2 to form a cartel since they both has the least unit costs 0.1. The coordination of outputs by the cartel damages them, whereas non-member firm 3 free rides on it. $\{1, 2\}$ is a disadvantage cartel. The largest cartel $\{1, 2, 3\}$ is successful since the cartel profit 0.203 is larger than the sum 0.149 of all firms' profits in the Cournot equilibrium.

Which of successful cartels forms in an equilibrium of the group formation game depends on a distribution of the total profit in the cartel $\{1, 2, 3\}$. Suppose that the total profit is distributed according to the Nash bargaining solution with the status-quo point of the Cournot equilibrium profits. Then, the three firms receive profits $\pi_1 = \pi_2 = 0.081$, $\pi_3 = 0.041$, respectively, in the cartel. Then, since firm 2 enjoys a higher profit 0.090 than 0.081 by deviating from the cartel, the largest cartel $\{1, 2, 3\}$ can not be sustained in equilibrium. In the example, we conclude that the two-firm cartel $\{1, 3\}$ is stable. Since firm 1 and firm 2 are symmetric, the cartel $\{2, 3\}$ is also an equilibrium of the group formation game.

5 Conclusion

We have investigated the formation of voluntary groups in a collective action problem with heterogeneous individuals. We have shown that the heterogeneous preferences yield a genuine multiplicity of strict Nash equilibria in the group formation game. Different levels of cooperation are possible in equilibrium, ranging from non-cooperation to universal cooperation. Typically, there exist many partial cooperation equilibria in which cooperators and free-riders co-exist. In real situations, individuals often differ in their willingness to cooperate. Furthermore, the heterogeneous preference model is suitable even in the situation that individuals' material payoff functions are identical if one takes into account their social and psychological preferences studied by Fehr and Schmidt (1999), for example. Our model of group formation is static. A promising research in future is to develop a dynamic model of group formation and to explore which of equilibrium groups is stable in the long-run.[8]

References

Chwe, M. S.-Y. (2000), "Communication and Coordination in Social Networks," *Review of Economic Studies*, Vol. 67, 1–16.

Coase, R. (1960), "The Problem of Social Cost," *Journal of Law and Economics*, Vol. 3, 1–44.

Dixit, A. and M. Olson (2000), "Does Voluntary Participation Undermine the Coase Theorem?," *Journal of Public Economics*, Vol. 76, 309–335.

Fudenberg, D. and E. Maskin (1986), "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, Vol. 54, 533–554.

Granovetter, M. (1978), "Threshold Models of Collective Behavior," *American Journal of Sociology*, Vol. 83, 1420–1443.

Fehr, E. and K. Schmidt (1999), "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics* Vol. 114, 817–868.

Kandori, M., G. Mailath, and R. Rob (1993), "Learning, Mutation and Long-Run Equilibria in Games," *Econometrica* Vol. 61, 29–56.

Marwell, G. and P. Oliver (1993), *The Critical Mass in Collective Action: A Micro-Social Theory*, Cambridge: Cambridge University Press.

Okada, A. (1993), "The Possibility of Cooperation in an n-Person Prisoners' Dilemma with Institutional Arrangements," *Public Choice*, Vol.77, 629–656.

[8] In a recent work (Maruta and Okada, 2001), we consider an equilibrium selection of group formation by applying the stochastic evolutionary theory developed by Young (1993) and Kandori, Mailath and Rob (1993).

Okada, A. and T. Maruta (2001), "Stochastic Stability of Group Formation in Collective Action Games," Discussion Paper No.536, Institute of Economic Research, Kyoto University.

Okada, A. and K. Sakakibara (1991), "The Emergence of the State - A Game Theoretic Analysis of the Theory of Social Contract, " *The Economic Studies Quarterly*, Vol. 42, 315-333.

Okada, A., K. Sakakibara and K. Suga (1997), "The Dynamic Transformation of Political Systems through Social Contract: A Game Theoretic Approach," *Social Choice and Welfare*, Vol. 14, 1-21.

Olson, M. (1965), *The Logic of Collective Action*, Cambridge: Harvard University Press.

Ostrom, E. (1998), "A Behavioral Approach to the Rational Choice Theory of Collective Action," *American Political Science Review* Vol. 92, 1–22.

Palfrey, R.T. and H. Rosenthal (1984), "Participation and the Provision of Discrete Public Goods: A Strategic Analysis," *Journal of Public Economics*, Vol. 24, 171–193.

Saijo, T. and T. Yamato (1999), "A Voluntary Participation Game with a Non-Excludable Public Good," *Journal of Economic Theory*, Vol. 84, 227–242.

Schelling, T.C. (1978), *Micro-Motives and Macro-Behavior*, New York: W.W. Norton.

Young, P.H. (1993), "The Evolution of Conventions," *Econometrica* Vol. 61, 57–84.