Coalition-proof Equilibria in a Voluntary Participation Game

Ryusuke Shinohara*

May 12 2010

(forthcoming in *International Journal of Game Theory***)**

Abstract

We examine the coalition-proof equilibria of a participation game in the provision of a (pure) public good. We study which Nash equilibria are achieved through cooperation, and we investigate coalition-proof equilibria under strict and weak domination. We show that under some incentive condition, (i) a profile of strategies is a coalition-proof equilibrium under *strict domination* if and only if it is a Nash equilibrium that is not strictly Pareto-dominated by any other Nash equilibrium and (ii) every strict Nash equilibrium for non-participants is a coalition-proof equilibrium under *weak domination.*

JEL classification: C72, D62, D71, H41.

Key Words: Participation game; Coalition-proof equilibrium; Dominance relation.

* Faculty of Economics, Shinshu University, 3-1-1, Asahi, Matsumoto, Nagano, 390-8621, JAPAN. E-mail: ryusukes@shinshu-u.ac.jp Telephone: +81-263-37-2951 Fax: +81-263-37-2344

1 Introduction

We examine the coalition-proof equilibria of a participation game in the provision of a (pure) public good. The situation is as follows. There exist one private good and one public good, and the public good is produced from the private good. There are $n \geq 2$ agents who simultaneously choose to either participate or not participate in the joint production of the public good. The agents who participate choose the level of public good and distribute its production cost in accordance with some rule. The agents who do not participate can free-ride. Such a game has been studied by several researchers. Saijo and Yamato (1999, 2010) and Shinohara (2009) study the participation issue in mechanisms such as those presented by Hurwicz (1979) and Walker (1981). An important application of our model is to the ratification game of international environmental agreements (Barrett, 1994; Carraro and Siniscalco, 1993). Most preceding studies investigate the Nash equilibria of the participation game.

These authors consider the case in which all agents have the same preference relations. Thus, the Nash equilibria of the participation game tend to have the same characteristic (e.g., equilibrium number of participants and equilibrium level of the public good) even if multiple Nash equilibria exist. However, when preferences differ, there may exist multiple Nash equilibria with different characteristics. In this paper, we allow differences in preferences and determine which Nash equilibria can be achieved through cooperation, as modelled through the coalition-proof equilibrium introduced by Bernheim et al. (1987). Two notions of coalition-proof equilibrium can be defined depending upon which notion of a dominance relation is adopted. The first is *strict domination*, and the second is *weak domination*. Let N be the set of agents and $T \subseteq N$ be a coalition. A strategy profile s_T strictly dominates a strategy profile s'_T for T if all members of T are made better off by switching from s'_T to s_T. The domination is weak if no member of T is made worse off and at least one member of T is made better off by the switch. The set of a coalition-proof equilibrium under weak domination, *wd coalition-proof equilibrium* for short, and that under strict domination, *sd coalition-proof equilibrium* for short, do not necessarily intersect, as Konishi et al. (1999) point out. We examine these two notions and clarify the relationship between Nash equilibria and coalition-proof equilibria for each dominance relation.

Our results involve a condition that relates to participation incentives — *participation inducements* (Condition 1). It says that if agent i does not gain by joining a set of participants P, then i does not gain by joining a set of participants that produces the public good at a higher level than P. First, we show that if participation inducements hold, the set of sd coalition-proof equilibria coincides with the set of Nash equilibria that are not strictly Pareto-dominated by any other Nash equilibrium.

A Nash equilibrium is *strict for non-participants* if no agent who does not participate gains by switching to participating. Second, we prove that if Condition 1 holds, every Nash equilibrium that is strict for non-participants is a wd coalition-proof equilibrium.

This paper is organized as follows. In Section 2, we introduce the participation game and the coalition-proof equilibrium under the two dominance relations. Section 3 considers the basic properties of our model, and Section 4 presents the main results. In Section 5, we apply our results. Section 6 concludes.

2 The model

2.1 A participation game in the provision of a public good

We consider the problem of providing a (pure) public good and distributing its cost. There exist one private good and one public good. The level of the public good can be any non-negative real number. The set of agents is denoted by $N = \{1, \ldots, n\}$ with $n \geq 2$. Each agent $i \in N$ has a preference relation that is representable by a quasi-linear utility function. If y and x_i designate the level of the public good and the contribution to public good production from agent $i \in N$, respectively, then agent i's utility is $V_i(y, x_i) = v_i(y) - x_i$. We assume that $v_i(0) = 0$, v_i is twice continuously differentiable, $v_i' > 0$, and $v_i'' < 0$. For each $y \geq 0$, $c(y)$ is the amount of private good required to produce y units of the public good. We assume that $c(0) = 0$, c is twice continuously differentiable, $c' > 0$, and $c'' \geq 0$.

We consider a situation in which there exists an opportunity for the joint production of a public good and each agent can decide whether or not to participate in the production. We consider the following two-stage game. In the first stage, agents simultaneously decide whether or not to participate. In the second stage, the agents who chose to participate jointly produce the public good and distribute its cost of production. Let $P \subseteq N$ be a set of participants, and let $(y^P, (x_j^P)_{j\in P})$ be the outcome of the second stage. We assume that the *ratio allocation* introduced by Kaneko (1977a, 1977b) is achieved in the second stage. Formally, $y^{\emptyset} := 0$, and $(y^P, (x_j^P)_{j\in P})$ satisfies the following conditions for each non-empty subset P of N:

$$y^P \in \arg\max_{y\in\mathbb{R}_+} \sum_{j\in P} v_j(y) - c(y) \text{ and } x_i^P = \frac{v_i'(y^P)}{\sum_{j\in P} v_j'(y^P)} c(y^P) \text{ for each } i \in P.$$

We further assume that $y^P > 0$ for each non-empty subset P of N.[*1]

We are not concerned with how the ratio allocation is attained in the second stage. However, if participants play the mechanisms constructed by Hurwicz (1979), Walker (1981), and Corchon and Wilkie (1996), the ratio allocation can be achieved at equilibrium of the mechanism.

Agents who select non-participation use the public good at no cost because of its non-excludability.

Assumption 1 For each $P \subseteq N$ and for each $i \notin P$, $x_i^P = 0$ and i consumes y^P.

Given the outcome of the second stage, the participation-decision stage can be reduced to the following simultaneous game. Each agent i chooses either $s_i = I$ (participation) or $s_i = O$ (non-participation). Let $P(s) := \{i \in N | \, s_i = I\}$ be the set of participants at profile $s = (s_1, \ldots, s_n)$. Then, each agent i obtains the utility

[*1] This assumption is not essential. In the case of $y^P = 0$ for some non-empty subset P of N, if our discussion is applied to each Q with $y^Q > 0$, the same conclusion can be obtained.

$V_i\left(y^{P(s)}, x_i^{P(s)}\right)$. In other words, the participants produce the public good and share the cost of the public good as above. Each non-participant uses the public good at no cost. We call this reduced game the *participation game*. It is formally defined as follows.

Definition 1 The *participation game* induced from $(V_i)_{i\in N}$ is the list $G = \left[N,\, S^n = \{I,\, O\}^n,\, (U_i)_{i\in N}\right]$, where U_i, payoff function of $i \in N$, is defined by $U_i(s) = V_i\left(y^{P(s)}, x_i^{P(s)}\right)$ for each $s \in S^n$.

2.2 The definition of coalition-proof equilibria

We limit our attention to pure strategy profiles.

A *Nash equilibrium* of a participation game is defined as usual.

Next, we define the notion of a *coalition-proof equilibrium* (Bernheim et al., 1987). It is a refinement of Nash equilibria based on stability against self-enforcing coalitional deviations. It is based on the notion of a *restricted game*. A restricted game is defined for each subset of agents and for each strategy of the agents outside the subset. For each $D \subseteq N$, we denote the complement of D by $-D$ and a strategy profile for D by $s_D \in S^{\#D}$.[*2] We simply write $s_N = s$. Let $T \subsetneq N$. Let $s_{-T} \in S^{n-\#T}$. The *restricted game* $G|s_{-T}$ is the game in which T plays G, taking it as given that the other agents choose s_{-T}. That is, the set of agents is T, set of strategy profiles is $S^{\#T}$, and payoff function for each $i \in T$ is the function $\widetilde{U}_i : S^{\#T} \to \mathbb{R}$ defined by $\widetilde{U}_i(s_T) = U_i(s_T,\, s_{-T})$ for each $s_T \in S^{\#T}$.

A coalition-proof equilibrium is defined under two different notions of dominance. First, given $T \subseteq N$ and $s_{-T} \in S^{n-\#T}$, a strategy profile $s_T \in S^{\#T}$ *strictly dominates* a strategy profile $\widetilde{s}_T \in S^{\#T}$ for T at s_{-T} if all members of T are made better off by switching from $\widetilde{s}_T$ to s_T. Second, a strategy profile s_T *weakly dominates* a strategy profile $\widetilde{s}_T$ for T at s_{-T} if no member of T is made worse off and at least one member

[*2] For each set $D \subseteq N$, $\#D$ means the cardinality of D.

of T is made better off by switching from $\widetilde{s}_T$ to s_T. Definitions 2 and 3 below are for coalition-proof equilibria under strict domination and under weak domination, respectively.

Definition 2 A coalition-proof equilibrium under strict domination, or *sd coalition-proof equilibrium* for short, $(s_1, \ldots, s_n)$ is defined inductively with respect to the number of agents t:

- When $t = 1$, for each $i \in N$, s_i is an sd coalition-proof equilibrium for $G|s_{-i}$ if $s_i \in \arg\max\ U_i(s_i', s_{-i})$ s.t. $s_i' \in S$.
- Let $T \subseteq N$ with $t = \#T \geq 2$. Assume that sd coalition-proof equilibria have been defined in the restricted games $G|s_{-Q}$ for each $Q \subsetneq T$.
- Consider the restricted game $G|s_{-T}$ with t agents.
  - A strategy profile $s_T \in S^t$ is *self-enforcing under strict domination*, or *sd self-enforcing*, if s_Q is an sd coalition-proof equilibrium of $G|s_{-Q}$ for each $Q \subsetneq T$.
  - A strategy profile s_T is an sd coalition-proof equilibrium of $G|s_{-T}$ if it is an sd self-enforcing strategy profile and no other sd self-enforcing strategy profile $\widehat{s}_T \in S^t$ exists such that $U_i(\widehat{s}_T, s_{-T}) > U_i(s_T, s_{-T})$ for each $i \in T$.

Definition 3 The definition of a coalition-proof equilibrium under weak domination, or *wd coalition-proof equilibrium* for short, is derived from Definition 2 by substituting "weak domination" for "strict domination."

The sd coalition-proof equilibria are defined as the weakly Pareto-efficient frontier within the set of sd self-enforcing strategy profiles. The sd self-enforcing strategy profiles are recursively defined with respect to the number of agents in coalitions. At an sd self-enforcing strategy profile of N, no proper coalition of N can coordinate its members' strategies in such a way that all members of the coalition are made better off and no proper subset of the coalition can further deviate in a self-enforcing way. The wd coalition-proof equilibria are similarly defined based on weak domination. These two notions of coalition-proof equilibrium do not necessarily intersect. See Konishi et al. (1999) and Shinohara (2005) for details.

3 Basic properties of the participation game

3.1 Properties for payoff functions

We define the payoff functions for each $P \subseteq N$ and each $i \in N$.

$$u_i(P) = \begin{cases} v_i(y^P) - \dfrac{v_i'(y^P)}{\sum_{j \in P} v_j'(y^P)} c(y^P) & \text{if } i \in P, \\ v_i(y^P) & \text{otherwise.} \end{cases}$$

Lemma 1 states that the level of the public good increases as the number of participants increases.

Lemma 1 For each pair P, $Q \subseteq N$, if $Q \subsetneq P$, then $y^P > y^Q$.

The proof is immediate.

Lemma 2 For each pair P, $Q \subseteq N$, if $y^P > y^Q$, then

$$u_i(P) > u_i(Q) \text{ for each } i \notin P \cup Q \text{ and} \tag{1}$$

$$u_j(P) > u_j(Q) \text{ for each } j \in P \cap Q. \tag{2}$$

Proof. Condition (1) is trivial. We show (2). Let $P, Q \subseteq N$ be such that $y^P > y^Q$, and let $i \in P \cap Q$. Since $(y^P, (x_j^P)_{j \in P})$ is a ratio equilibrium for P, $V_i\left(y, \frac{v_i'(y^P)}{\sum_{j \in P} v_j'(y^P)} c(y)\right)$ is maximized at y^P. Hence,

$$u_i(P) = v_i(y^P) - \frac{v_i'(y^P)}{\sum_{j \in P} v_j'(y^P)} c(y^P) \geq v_i(y^Q) - \frac{v_i'(y^P)}{\sum_{j \in P} v_j'(y^P)} c(y^Q). \tag{3}$$

Note that since $y^P > y^Q$, then $\sum_{j \in P} v_j'(y^P) = c'(y^P) \geq c'(y^Q) = \sum_{j \in Q} v_j'(y^Q)$.

Because

$$\begin{aligned}\frac{v_i'(y^P)}{\sum_{j\in P} v_j'(y^P)} - \frac{v_i'(y^Q)}{\sum_{j\in Q} v_j'(y^Q)} &= \frac{v_i'(y^P)[\sum_{j\in Q} v_j'(y^Q)] - v_i'(y^Q)[\sum_{j\in P} v_j'(y^P)]}{[\sum_{j\in P} v_j'(y^P)][\sum_{j\in Q} v_j'(y^Q)]} \\ &\leq \frac{v_i'(y^P)[\sum_{j\in P} v_j'(y^P)] - v_i'(y^Q)[\sum_{j\in P} v_j'(y^P)]}{[\sum_{j\in P} v_j'(y^P)][\sum_{j\in Q} v_j'(y^Q)]} \\ &= \frac{[v_i'(y^P) - v_i'(y^Q)][\sum_{j\in P} v_j'(y^P)]}{[\sum_{j\in P} v_j'(y^P)][\sum_{j\in Q} v_j'(y^Q)]} < 0,\end{aligned}$$

we have

$$\frac{v_i'(y^P)}{\sum_{j\in P} v_j'(y^P)} < \frac{v_i'(y^Q)}{\sum_{j\in Q} v_j'(y^Q)}. \tag{4}$$

Therefore, by (3) and (4), $u_i(P) = v_i(y^P) - \frac{v_i'(y^P)}{\sum_{j\in P} v_j'(y^P)}c(y^P) > v_i(y^Q) - \frac{v_i'(y^Q)}{\sum_{j\in Q} v_j'(y^Q)}c(y^Q) = u_i(Q)$. ■

Lemma 2 is a basic property of the payoff functions. From Lemma 2, the payoffs to all agents increase with the level of the public good. This property will play an important role in showing the main results.

3.2 Nash equilibria and Pareto domination

We introduce two notions of Pareto domination.

Definition 4 A strategy profile $s \in S^n$ *weakly Pareto-dominates* a strategy profile $\widetilde{s}$ if $U_i(s) \geq U_i(\widetilde{s})$ for each $i \in N$ and $U_i(s) > U_i(\widetilde{s})$ for some $i \in N$. Profile $s \in S^n$ *strictly Pareto-dominates* $\widetilde{s}$ if $U_i(s) > U_i(\widetilde{s})$ for each $i \in N$.

Lemma 3 provides a sufficient condition for a strategy profile to be Pareto-dominated by some Nash equilibrium.

Lemma 3 Let $s \in S^n$ be a Nash equilibrium. Then, s weakly Pareto-dominates every $\widehat{s}$ such that $P(\widehat{s}) \subsetneq P(s)$. Moreover, s strictly Pareto-dominates $\widehat{s}$ if $\#[P(s) \backslash P(\widehat{s})] \geq 2$.

Proof. By Lemma 1, $y^{P(s)} > y^{P(\widehat{s})}$. From $y^{P(s)} > y^{P(\widehat{s})}$, (1), and (2), we obtain that $u_i(P(s)) > u_i(P(\widehat{s}))$ for each $i \in P(\widehat{s}) \cup [N \backslash P(s)]$. From the definition of a Nash equilibrium and (1), we obtain that $u_i(P(s)) \geq u_i(P(s)\backslash\{i\}) \geq u_i(P(\widehat{s}))$ for each $i \in P(s)\backslash P(\widehat{s})$. The second inequality holds strictly if $P(\widehat{s}) \subsetneq P(s)\backslash\{i\}$. ■

4 Main results

Condition 1 (Participation inducements) For each pair P, $Q \subseteq N$ with $y^P > y^Q$ and each $i \in P \cap Q$, if $u_i(Q\backslash\{i\}) - u_i(Q) \geq 0$, then $u_i(P\backslash\{i\}) - u_i(P) > 0$.

Here is an interpretation of Condition 1. An increase in the level of the public good does not change the incentive to participate. For example, let $Q \subseteq N$ and $i \in Q$ such that $u_i(Q\backslash\{i\}) \geq u_i(Q)$. If $u_i(Q\backslash\{i\}) > u_i(Q)$, agent i has the incentive to withdraw from Q and free-ride. Condition 1 says that agent i would also want to withdraw from every P that produces the public good at a level higher than y^Q. If $u_i(Q\backslash\{i\}) = u_i(Q)$, agent i is indifferent between joining and free-riding on $Q\backslash\{i\}$. Agent i gains by not joining if the level of the public good increases. Thus, if an agent benefits from not participating in providing the public good at level $y > 0$, this agent also benefits from not participating in providing the public good of any level higher than y.

4.1 Characterization of sd coalition-proof equilibria

We show that the set of sd coalition-proof equilibria coincides with the weakly Pareto-efficient frontier of the set of Nash equilibria.

Proposition 1 In the participation game, under Condition 1, a strategy profile is an sd coalition-proof equilibrium if and only if it is a Nash equilibrium that is not strictly Pareto-dominated by any other Nash equilibrium.

Proof. We first show that if $s \in S^n$ is a Nash equilibrium that is not strictly Pareto-dominated by any other Nash equilibrium, then it is an sd coalition-proof equilibrium. Assume, on the contrary, that s is not an sd coalition-proof equilibrium. Let $D \subsetneq N$ be a deviating coalition, and $\widetilde{s}_D$ be a profile of deviating strategies of D. Let us denote $\widetilde{s} := (\widetilde{s}_D, s_{-D})$. Note that $\widetilde{s}_D$ is an sd coalition-proof equilibrium in $G|s_{-D}$ and $u_i(P(\widetilde{s})) > u_i(P(s))$ for each $i \in D$. Since s is a Nash equilibrium, then $\#D \geq 2$.

Claim 1 If $y^{P(s)} \geq y^{P(\widetilde{s})}$, there is a member of D that is worse off after the deviation.

Proof of Claim 1. From Lemma 1, if $P(s) \subsetneq P(\widetilde{s})$, $y^{P(s)} < y^{P(\widetilde{s})}$. Hence, $P(s) \subsetneq P(\widetilde{s})$ does not hold. There is $i \in P(s)\backslash P(\widetilde{s})$ or $P(\widetilde{s}) \subseteq P(s)$. From Lemma 3, if $P(\widetilde{s}) \subseteq P(s)$, the deviation by D is not profitable. If $i \in P(s)\backslash P(\widetilde{s})$ and $j \in P(\widetilde{s})\backslash P(s)$ exist, agent j switches from O to I by the deviation of D. The payoff to j before the deviation is $u_j(P(s)) = v_j(y^{P(s)})$, and the payoff to j after the deviation is $u_j(P(\widetilde{s})) = v_j(y^{P(\widetilde{s})}) - \dfrac{v_j'(y^{P(\widetilde{s})})}{\sum_{k \in P(\widetilde{s})} v_k'(y^{P(\widetilde{s})})} c(y^{P(\widetilde{s})})$. Since $y^{P(s)} \geq y^{P(\widetilde{s})}$, we have $v_j(y^{P(s)}) \geq v_j(y^{P(\widetilde{s})})$, which, together with $\dfrac{v_j'(y^{P(\widetilde{s})})}{\sum_{k \in P(\widetilde{s})} v_k'(y^{P(\widetilde{s})})} c(y^{P(\widetilde{s})}) > 0$, implies $u_j(P(s)) > u_j(P(\widetilde{s}))$. Therefore, agent i is made worse off by the deviation. **(End of Proof of Claim 1)**

From Claim 1, if all members of D are better off after the deviation, $y^{P(\widetilde{s})} > y^{P(s)}$.

Claim 2 If $y^{P(\widetilde{s})} > y^{P(s)}$, then $\widetilde{s}_D$ is not a Nash equilibrium of $G|s_{-D}$.

Proof of Claim 2. If $k \in P(s)$ for each $k \in P(\widetilde{s})$, then $P(\widetilde{s}) \subseteq P(s)$, which implies that $y^{P(\widetilde{s})} \leq y^{P(s)}$. Hence, $k \notin P(s)$ for some $k \in P(\widetilde{s})$. Notice that $k \in D$. Since the deviation by D is improving, $u_k(P(\widetilde{s})) > u_k(P(s))$. By the definition of a Nash equilibrium, $u_k(P(s)) \geq u_k(P(s) \cup \{k\})$. Thus, $u_k(P(\widetilde{s})) > u_k(P(s) \cup \{k\})$. Since $k \in P(\widetilde{s})$, then $y^{P(\widetilde{s})} > y^{P(s) \cup \{k\}}$. Otherwise, $u_k(P(\widetilde{s})) \leq u_k(P(s) \cup \{k\})$ from (2). By Condition 1, $u_k(P(s)) - u_k(P(s) \cup \{k\}) \geq 0$ implies $u_k(P(\widetilde{s})\backslash\{k\}) - u_k(P(\widetilde{s})) > 0$. Hence, $u_k(P(\widetilde{s})\backslash\{k\}) > u_k(P(\widetilde{s}))$, which implies that $\widetilde{s}_D$ is not a Nash equilibrium of $G|s_{-D}$. **(End of Proof of Claim 2)**

From Claims 1 and 2, no coalition can deviate in a self-enforcing and payoff-improving way. Consequently, every Nash equilibrium that is not strictly Pareto-dominated by any other Nash equilibrium is an sd coalition-proof equilibrium. By the definition of an sd coalition-proof equilibrium, no Nash equilibrium that is strictly Pareto-dominated by an sd coalition-proof equilibrium is coalition-proof. Thus, no Nash equilibrium that is strictly Pareto dominated by another Nash equilibrium is an sd coalition-proof equilibrium. ■

The intuition that no coalition has a profitable and self-enforcing deviation from any point in the Pareto-efficient frontier of the set of Nash equilibria is as follows. From Lemma 2, payoffs are positively correlated with the level of the public good. Hence, if a coalition deviates in a way that reduces this level, some members of the coalition are made worse off. This is proved in Claim 1. Therefore, a deviation of a coalition is profitable if the public good level increases. However, if a coalition deviates in such a way that the level of the public good increases, an agent j who chooses non-participation before the deviation needs to join in the deviation and switch from not participating to participating. From Condition 1, it follows that agent j wants to switch back to not participating after this coalitional deviation, because the level of the public good rises. Thus, no deviation that increases the level of the public good is self-enforcing. This is shown in Claim 2.

From Proposition 1, in the participation game under Condition 1, if there exist multiple Nash equilibria, one of which strictly Pareto-dominates the others, an sd coalition-proof equilibrium Pareto-dominates all other Nash equilibria. This does not apply to all strategic games. In a strategic game with more than two players, the set of sd coalition-proof equilibria and the set of Pareto-superior Nash equilibria do not necessarily intersect (Bernheim et al., 1987).

4.2 Relationship between a Nash equilibrium and a wd coalition-proof equilibrium

In this subsection, we provide a sufficient condition for a Nash equilibrium to be a wd coalition-proof equilibrium in the participation game, and we address the question of which Nash equilibria are wd coalition-proof equilibria.

Definition 5 A Nash equilibrium $s \in S^n$ is *strict for non-participants* if $u_i(P(s)) > u_i(P(s) \cup \{i\})$ for each $i \notin P(s)$.

At a strict Nash equilibrium for non-participants, the participants may be indifferent between participating and not participating, whereas each non-participant is better off not participating.

Proposition 2 Assume that Condition 1 holds. Every Nash equilibrium that is strict for non-participants is a wd coalition-proof equilibrium as well as a Nash equilibrium that is not *weakly* Pareto-dominated by any other Nash equilibrium.

Proof. The proof of this proposition is very similar to that of Proposition 1. Let $s \in S^n$ be a Nash equilibrium that is strict for non-participants. We show by contradiction that s is a wd coalition-proof equilibrium. Assume that a coalition $D \subseteq N$ deviates from s by using $\widetilde{s}_D \in S^{\#D}$. Denote $\widetilde{s} := (\widetilde{s}_D, s_{-D})$. Notice that (i) $\widetilde{s}_D$ is a wd coalition-proof equilibrium of $G|s_{-D}$ and (ii) no member of D is made worse off and at least one member of D is made better off by this deviation.

We can prove, similarly to Claim 1, that $y^{P(\widetilde{s})} > y^{P(s)}$. If $y^{P(\widetilde{s})} > y^{P(s)}$, then $i \in P(\widetilde{s}) \backslash P(s)$. By (ii), $u_i(P(\widetilde{s})) \geq u_i(P(s))$. Since s is a Nash equilibrium that is strict for non-participants, $u_i(P(s)) > u_i(P(s) \cup \{i\})$. Therefore, $u_i(P(\widetilde{s})) - u_i(P(s) \cup \{i\}) > 0$, which implies $y^{P(\widetilde{s})} > y^{P(s) \cup \{i\}}$. From Condition 1 and the strictness of s, it follows that $u_i(P(s)) - u_i(P(s) \cup \{i\}) > 0$ implies $u_i(P(\widetilde{s}) \backslash \{i\}) - u_i(P(\widetilde{s})) > 0$. Thus, $u_i(P(\widetilde{s}) \backslash \{i\}) > u_i(P(\widetilde{s}))$. By this inequality, $\widetilde{s}_D$ is not a Nash equilibrium of

$G|s_{-D}$, which is a contradiction.

Substituting N for D in the proof above, we show that s is a Nash equilibrium that is not weakly Pareto-dominated by any other Nash equilibrium. ■

The intuition for Proposition 2 is almost the same as that for Proposition 1. The difference is that in Proposition 2, we must treat joint deviations in which at least one agent in a coalition is made better off and the others are not made worse off. If a Nash equilibrium is not strict for non-participants, it may not be a wd coalition-proof equilibrium for the following reason: let $s \in S^n$ denote a Nash equilibrium that is not strict for non-participants: there exists $i \notin P(s)$ such that $u_i(P(s)) = u_i(P(s) \cup \{i\})$. That is, agent i is indifferent between participating and not participating when $P(s)$ is the set of participants. Then, let $j \in P(s)$, and consider a joint deviation by i and j in which i deviates from not participating to participating and j continues to participate. After this deviation, i and j receive payoffs $u_i(P(s) \cup \{i\})$ and $u_j(P(s) \cup \{i\})$, respectively. By Lemma 1, $y^{P(s)\cup\{i\}} > y^{P(s)}$. By (2), $u_j(P(s) \cup \{i\}) > u_j(P(s))$. However, nothing prevents the possibility that $u_j(P(s) \cup \{i\}) \geq u_j((P(s) \cup \{i\}) \backslash \{j\})$. Therefore, if $u_j(P(s) \cup \{i\}) \geq u_j((P(s) \cup \{i\}) \backslash \{j\})$, the deviation of $\{i, j\}$ is self-enforcing under weak domination.

Definition 6 A strategy profile $s \in S^n$ is a *strict Nash equilibrium* if $U_i(s_i, s_{-i}) > U_i(\widetilde{s}_i, s_{-i})$ for each $i \in N$ and each $\widetilde{s}_i \in S \backslash \{s_i\}$.

The following corollary is immediate from Proposition 2.

Corollary 1 In the participation game, under Condition 1, every strict Nash equilibrium is a wd coalition-proof equilibrium.

From Proposition 2, it follows that a Nash equilibrium that is strict for non-participants is a wd coalition-proof equilibrium. This Nash equilibrium is not weakly Pareto-dominated by any other Nash equilibrium. Thus, from Proposition 1, it is also an sd coalition-proof equilibrium. These results mean that under Condition 1, the sets of wd and sd coalition-proof equilibria intersect. This phenomenon is noteworthy since these sets do not intersect in general.

Finally, Propositions 1 and 2 hold as long as Lemmas 1 and 2 are satisfied. Therefore, the assumption that all agents have quasi-linear preferences does not matter. An example of a preference domain that satisfies Lemmas 1 and 2 is the symmetric Cobb-Douglas domain, examined by Saijo and Yamato (1999). On this domain, all agents have preference relations that are representable by the same Cobb-Douglas utility function and they have the same endowments of the private good. Refer to Shinohara (2007) for a detailed discussion.

4.3 Related literature

Yi (1999) studies the relation between the set of sd coalition-proof equilibria and the weakly Pareto-efficient frontier of the set of Nash equilibria. He focuses on a class of games in which the strategy space of each player is a subset of the real line and the payoff to each player depends on the sum of his opponents' strategies. Our participation game is in this class only when, designating by 1 and 0 participation and non-participation, $u_i(P(s_i, s_{-i})) = u_i(P(s_i, \widetilde{s}_{-i}))$ for each $i \in N$, each $s_i \in \{0, 1\}$, and each pair s_{-i}, $\widetilde{s}_{-i} \in \{0, 1\}^{n-1}$ with $\sum_{j \neq i} s_j = \sum_{j \neq i} \widetilde{s}_j$. This condition means that the payoffs to all players depend on the number of participants and not on their identities. It is satisfied only when the preferences are the same. Since our model allows preferences to differ, Yi (1999)'s results cannot be applied to our model to characterize the set of sd coalition-proof equilibria.

Thoron (1998) examines the wd coalition-proof equilibria of a cartel-formation game, which is similar to the participation game in a public good provision. In her game, each firm decides whether or not to join a cartel. Only the firms that join the cartel follow its agreements, and the other firms behave independently. The main differences between Thoron (1998) and this paper are as follows. First, Thoron (1998) assumes that firms are identical, whereas we allow players to differ. Second, the condition that is satisfied in Thoron (1998)'s model differs from our condition. Conversely, our condition (2) in Lemma 2 does not hold in Thoron (1998)'s model. Instead, she assumes that the cartel members' payoffs are less than the non-members' payoffs. In our model, since we allow players to differ, this assumption does not necessarily hold. Hence, we cannot use her results to clarify the properties of coalition-proof equilibria in the participation game.

Furusawa and Konishi (2009) study a participation game that is similar to ours. They examine the relationship between the wd coalition-proof equilibrium and the *free-riding-proof core*. Let T be a coalition. Suppose that T produces the public good. If agent $i \in T$ withdraws from T, i benefits from the public good produced by $T \backslash \{i\}$. A *free-riding-proof allocation for* T is an allocation such that no member of T gains by withdrawing from T. It is assumed that if a coalition deviates from some allocation, it takes its free-riding-proof allocation. The free-riding-proof core is the set of all free-riding-proof allocations from which no coalition deviates by taking its free-riding-proof allocations. Furusawa and Konishi (2009) show that the set of wd coalition-proof equilibrium allocations coincides with the free-riding-proof core. They do not examine sd coalition-proof equilibria. In order to identify the wd coalition-proof equilibria, they focus on the core-based solution. However, we focus on Nash equilibria.

5 Application of the main results

Applying Propositions 1 and 2, we characterize the sets of wd and sd coalition-proof equilibria when $v_i(y) = \alpha_i \sqrt{y}$, where $\alpha_i > 0$ for each $i \in N$ and $c(y) = y$.

Let $P \subseteq N$. Then, $y^P = \left(\frac{\sum_{j \in P} \alpha_j}{2} \right)^2$. The payoff functions are as follows:

$$u_i(P) = \begin{cases} \dfrac{\alpha_i(\sum_{j \in P} \alpha_j)}{4} & \text{if } i \in P, \text{ and} \\ \dfrac{\alpha_i(\sum_{j \in P} \alpha_j)}{2} & \text{if } i \notin P. \end{cases} \tag{5}$$

From (5), it follows that

$$u_i(P\backslash\{i\}) - u_i(P) = \frac{\alpha_i(\sum_{j\in P\backslash\{i\}}\alpha_j - \alpha_i)}{4} \lesseqgtr 0 \text{ if } \sum_{j\in P\backslash\{i\}}\alpha_j \lesseqgtr \alpha_i. \tag{6}$$

From (6), the Nash-equilibrium sets of participants are characterized as follows:

Lemma 4 A set of participants $P \subseteq N$ is a Nash-equilibrium set of participants in the participation game if and only if (i) $\sum_{j\in P\backslash\{i\}}\alpha_j \leq \alpha_i$ for each $i \in P$ and (ii) $\sum_{j\in P}\alpha_j \geq \alpha_i$ for each $i \notin P$.

Lemma 5 Let $n \geq 2$. There is no Nash equilibrium in which more than two agents participate.

Proof. Assume by contradiction that there is a Nash equilibrium in which P with $\#P \geq 3$ is the set of participants. Then, $\alpha_i \geq \sum_{j\in P\backslash\{i\}}\alpha_j$ for each $i \in P$. Summing up these inequalities over $i \in P$ yields $\sum_{i\in P}\alpha_i \geq \sum_{i\in P}\sum_{j\in P\backslash\{i\}}\alpha_j = (\#P - 1)\sum_{i\in P}\alpha_i$. However, since $\#P \geq 3$, $\sum_{i\in P}\alpha_i < (\#P-1)\sum_{i\in P}\alpha_i$. This is a contradiction. ■

Proposition 3 In the participation game with $n \geq 3$, there exists a Nash equilibrium that is strict for non-participants.

Proof. First, suppose that $\arg\max_{l\in N}\alpha_l$ is a singleton. Let $\{i\} = \arg\max_{l\in N}\alpha_l$. From (6), $\{i\}$ is supported at a Nash equilibrium at which $u_j(\{i\}) > u_j(\{i,j\})$ for each $j \in N\backslash\{i\}$. Second, suppose that $\arg\max_{l\in N}\alpha_l$ is not a singleton. Let $\{i,j\} \subseteq \arg\max_{l\in N}\alpha_l$. From (6), $\{i,j\}$ is attained at a Nash equilibrium at which $u_k(\{i,j\}) > u_k(\{i,j,k\})$ for each $k \in N\backslash\{i,j\}$. Therefore, there is a Nash equilibrium that is strict for non-participants. ■

When $v_i(y) = \alpha_i\sqrt{y}$ with $\alpha_i > 0$ and $c(y) = y$, Condition 1 is satisfied. Let P, $Q \subseteq N$ with $y^P > y^Q$, and let $i \in P \cap Q$. Since $y^P > y^Q$, we obtain $\sum_{j\in P}\alpha_j > \sum_{j\in Q}\alpha_j$, which implies $\sum_{j\in P\backslash\{i\}}\alpha_j - \alpha_i > \sum_{j\in Q\backslash\{i\}}\alpha_j - \alpha_i$. From (6), it follows that if $u_i(Q\backslash\{i\}) - u_i(Q) \geq 0$, then $u_i(P\backslash\{i\}) - u_i(P) > 0$. Hence, by Propositions 1 and 2,

(i) a strategy profile is a Nash equilibrium that is not strictly Pareto-dominated by any other Nash equilibrium if and only if it is an sd coalition-proof equilibrium, and (ii) there exist sd and wd coalition-proof equilibria.

Finally, using the results in this subsection, we characterize the set of coalition-proof equilibria in the following example.

Example 1 Let $N = \{1, 2, 3, 4\}$, with $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (3, 3, 2, 2)$. By Lemma 4, $\{1\}$, $\{2\}$, $\{1, 2\}$, and $\{3, 4\}$ are supported at Nash equilibria. Table 1 represents the payoffs at Nash equilibria. This table shows that no Nash equilibrium dominates any other Nash equilibrium in the sense of strict Pareto domination. Hence, these four sets of participants can be supported at sd coalition-proof equilibria. Since $\{1, 2\}$ and $\{3, 4\}$ are attained at strict Nash equilibria for non-participants, they are supported at wd coalition-proof equilibria. Sets $\{1\}$ and $\{2\}$ are not supported at wd coalition-proof equilibria.

⟨Insert Table 1 here⟩

If $n = 4$, we need to consider one-agent games, two-agent games, three-agent games, and the whole game, in that order. Since there exist four one-agent coalitions, six two-agent coalitions, and four three-agent coalitions, it is time-consuming to identify the set of coalition-proof equilibria according to its definition. However, from the results in this paper, we can characterize the set of coalition-proof equilibria by just checking Condition 1 and the strictness of Nash equilibria. Therefore, by applying our results, we can more easily characterize the equilibrium set.

6 Concluding remarks

The definition of a coalition-proof equilibrium is recursive and complicated. Therefore, in general, it is difficult to identify this equilibrium. This applies to simple games as well, such as the participation game. This paper clarifies which Nash equilibrium is coalition-proof in the participation game under Condition 1. Condition 1 may be restrictive, and our results may not be applicable to all participation games. The characterization of coalition-proof equilibria in the participation game without Condition 1 is left for future work.

Acknowledgements This paper is based on Chapters 3 and 4 of my doctoral dissertation submitted to Hitotsubashi University. I thank Koichi Tadenuma for a number of helpful comments, suggestions, and constant support. I am also grateful to William Thomson, Tatsuro Ichiishi, Shin-ichi Takekuma, Akira Okada, Motohiro Sato, Tatsuyoshi Saijo, Takehiko Yamato, Hideo Konishi, Takashi Shimizu, and participants of the seminars at the autumn meeting of the Japanese Economic Association at Meiji University and the Decentralization Conference in Japan. I also thank the anonymous referees. This research was supported by the Grand-in-Aid for Young Scientists (18830026, 21730156) from the Japan Society for the Promotion of Science. Of course, any remaining errors are my own.

References

[1] Barrett S (1994) Self-enforcing international environmental agreements. Oxford Econ Pap **46**: 878-894.

[2] Bernheim D, Peleg B, Whinston M (1987) Coalition-proof Nash Equilibria I. Concepts. J Econ Theory **42**: 1-12.

[3] Carraro C, Siniscalco D (1993) Strategies for the International Protection of the Environment. J Public Econ **52**: 309-328.

[4] Corchon L, Wilkie S (1996) Double Implementation of the Ratio Correspondence by a Market Mechanism. Econ Design **2**: 325-337.

[5] Furusawa T, Konishi H (2009) Contributing or Free-riding?: Voluntary Participation in a Public Good Economy. mimeo, Boston College.

[6] Hurwicz L (1979) Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points. Rev Econ Stud **46**: 217-225.

[7] Kaneko M (1977a) The Ratio Equilibrium and a Voting Game in a Public Goods Economy. J Econ Theory **16**: 123-136.

[8] Kaneko M (1977b) The Ratio Equilibria and the Core of the Voting Game $G(N, W)$ in a Public Goods Economy. Econometrica **45**: 1589-1594.

[9] Konishi H, Le Breton M, Weber S (1999) On Coalition-proof Nash Equilibria in Common Agency Games. J Econ Theory **85**: 122-139.

[10] Saijo T, Yamato T (1999) A Voluntary Participation Game with a Non-excludable Public Good. J Econ Theory **84**: 227-242.

[11] Saijo, T., and T. Yamato (2010) Fundamental Impossibility Theorems on Voluntary Participation in the Provision of Non-excludable Public Goods. Rev Econ Design **14**: 51-73.

[12] Shinohara R (2005) Coalition-proofness and Dominance Relations. Econ Letters **89**: 174-179.

[13] Shinohara R (2007) Coalition-proof Equilibria in a Voluntary Participation Game. Staff Paper #2007-07, Faculty of Economics, Shinshu University. (http://www.shinshu-u.ac.jp/faculty/economics/research/pdf/sp_2007-7.pdf)

[14] Shinohara R (2009) The Possibility of Efficient Provision of a Public Good in Voluntary Participation Games. Soc Choice Welfare **32**: 367-387.

[15] Thoron S (1998) Formation of a Coalition-proof Stable Cartel. Can J Econ **31**: 63-76.

[16] Walker M (1981) A Simple Incentive Compatible Scheme for Attaining Lindahl Allocations. Econometrica **49**: 65-71.

[17] Yi S (1999) On Coalition-proofness of the Pareto Frontier of the Set of Nash Equilibria. Games Econ Behav **26**: 353-364.

Payoffs / Nash-equilibrium sets of participants	Agent 1's payoff	agent 2's payoff	agent 3's payoff	agent 4's payoff
$\{1\}$	$9/4$	$9/2$	3	3
$\{2\}$	$9/2$	$9/4$	3	3
$\{1,2\}$	$9/2$	$9/2$	6	6
$\{3,4\}$	6	6	2	2

Table. 1 Payoff table of Example 1