Staff Paper Series ’07-04 May 2007

A Voluntary Participation Game through a Unit-by-Unit Cost Share Mechanism of a Non-excludable Public Good

Yukihiro Nishimura
and
Ryusuke Shinohara

Faculty of Economics
Shinshu University
Matsumoto 390-8621 Japan
Phone: +81-263-35-4600
Fax: +81-263-37-2344

A Voluntary Participation Game through a Unit-by-Unit Cost Share Mechanism of a Non-Excludable Public Good

Yukihiro Nishimura[a] * and Ryusuke Shinohara[b]

[a] Yokohama National University and Queen's University, Department of Economics, Dunning Hall, Queen's University, Kingston, Ontario, K7L 3N6, CANADA.

E-mail: ynishimu@ynu.ac.jp

[b] Department of Economics, Shinshu University, 3-1-1, Asahi, Matsumoto, Nagano, 390-8621, JAPAN.

E-mail: ryusukes@shinshu-u.ac.jp

December 12, 2006

Key Words: Public Good; Participation; Mechanism Design

JEL Classification: C72; D71; D78; H41

Abstract

The strategic analysis of voluntary participation in the public good provision has shown two distinct results. First, when the provision of public goods is binary, there are Nash equilibria supporting efficient allocations, and these are Strong Nash equilibria of the game. On the other hand, in the model of a continuous public good, Saijo and Yamato (1999, Journal of Economic Theory) showed that the participation of all agents is not an equilibrium in many situations. This paper considers the provision of a public good that is discrete and multi-unit, and considers a unit-by-unit participation game. Namely, people are asked to participate in each unit of public good provision, and those who chose to participate share the marginal cost of public good. In this game of public good provision, unlike the case of Saijo-Yamato, there are subgame-perfect equilibria that are Pareto efficient. We also use the refinement concepts to eliminate inefficient subgame-perfect equilibria and also to characterize the efficient subgame-perfect equilibria.

*Corresponding author. Tel: +1 (613) 533-2262, Fax: +1 (613) 533-6668.

1 Introduction

The strategic analysis of the public good provision (known as the free-rider problem) has two theoretical issues. The first, one that has gathered much interest, is the *demand revelation* problem. Groves and Ledyard (1977), Walker (1981) and others showed that there is a mechanism that achieves a Pareto efficient allocation through the Nash equilibrium. On the other hand, Palfrey and Rothenthal (1984), Saijo and Yamato (1999), and Dixit and Olson (2000) pointed out another issue of *participation*. Due to the nature of a public good, called non-excludability, each member of a group is tempted to free-ride, hoping that the other members of the group will pay the cost of the public good. Dixit and Olson (2000), for example, say that the conventional mechanism design approaches "start by assuming that the power to make and implement the mechanism has been handed over to someone, presumably by a duly constituted meeting of all participants. They do not consider individual incentives to attend this meeting; thus, they skip what we called the vital first stage of voluntary participation" (p. 313-314).

The significance of the participation problem is recognized in various economic situations. Many international treaties, including the withdrawal of the United States and other countries from the Kyoto Protocol, face this problem. Although it is in the interest of all to work towards common concerns, such as solving global environmental problems and promoting world peace,[1] we cannot prevent non-ratifiers of treaties to free-ride on the non-excludable benefits created by ratifiers,

[1] For example, the Nuclear Non-Proliferation Treaty is known to have nonsignatories that possess (or have ambition to possess) nuclear weapons. This results in the underprovision of a public good, i.e., world peace.

because of the absence of coercive powers to punish non-ratifiers. Another example of the participation problem is public broadcasting. For example, a public service broadcaster in Japan (Nihon Hoso Kyokai) is supported by the broadcasting fee. But many households choose not to pay the fee, as there are no penalties in place.

The analysis of voluntary participation has shown two distinct results. First, when the provision of public goods is binary (Palfrey and Rothenthal (1984), Dixit and Olson (2000), Cavaliere (2001), Shinohara (2004, 2005)), there are Nash equilibria supporting efficient allocations, and these are Strong Nash equilibria (Aumann (1959)) of the game. On the other hand, in the model of a continuous public good, Saijo and Yamato (1999) examined the following two-stage game with voluntary participation: In the first stage, each agent simultaneously decides whether she participates in the mechanism or not, and in the second stage, knowing the other agents' participation decisions, the agents who chose to participate in the first stage select their messages in the mechanism. They showed that the participation of all agents is not an equilibrium in many situations. Okada (1993) and Ray and Vohra (1997) derived similar conclusions under different settings.

This paper considers whether there is a mechanism that takes account of agents' voluntary participation and that implements Pareto efficient allocations. We examine the provision of multi-unit public good where the unit is discrete. The direct application of Saijo-Yamato's two-stage mechanism results in too little participation and inefficient allocations. We then construct the following *unit-by-unit participation game*. Rather than a group of participants choosing the level of public good provision as in Saijo-Yamato's framework, we consider the case where agents make the decision to participate in *each unit* of public good provision. For each unit, the agents who choose to participate share the marginal cost of public good. If the sum of the contributions by participants can afford the marginal cost, then that unit of public good is provided and the game moves to the next unit of the public good provision.

An intriguing real-world example of the unit-by-unit participation game is the negotiation process at abatement level of chlorofluorocarbon emission at the Montreal Protocol and the subsequent amendments and adjustments. The Protocol in 1987 established regulations on substances that would deplete the ozone layer. Subsequently, signatory countries held several conferences (at London in 1990, at Copenhagen in 1992, at Vienna in 1995, and at Montreal in 1997), where amendments were adopted to tighten existing control schedules and to add controls for further reduction of emissions of ozone-depleting substances. Amendments required ratification by a defined number of parties before their entry into force. The successful results of the Protocol have been recognized by many as an example of solving global environmental problems.

In this game of public good provision, unlike the case of Saijo-Yamato, there are subgame-perfect equilibria that are Pareto efficient. Moreover, these efficient subgame-perfect equilibria satisfy a stronger concept of a strict perfect equilibrium (Muto (1986)). We then characterize the set of strict perfect equilibria of this game. We also show that the concepts of a strong perfect equilibrium (Rubinstein (1980)) and a Perfectly Coalition-Proof Nash equilibrium (Bernheim, Peleg and Whinston (1987)) are useful to eliminate inefficient subgame-perfect equilibria and also to characterize the efficient subgame-perfect equilibria.

2 Model

Consider a model with one public good that can be produced from a single private good. The public good is produced in discrete units: the unit of public good y can take only the value of integers. Let $C(y)$ be the total cost of producing y units of the public good, and let $c(y) \equiv C(y) - C(y-1) > 0$, $y = 1, 2, ...$, denote the cost of producing the y'th unit of the public good (hereafter referred to as 'the marginal cost'). We assume $C(0) = 0$.

There are n agents. Each agent i ($i \in \{1, ..., n\} \equiv N$) has the preference relation that is represented by a quasi-linear utility function $V_i(x_i, y) = U_i(y) - x_i$, where x_i denotes agent i's contribution of the private good. We normalize to $U_i(0) = 0$ for all i. Let $u_i(y) \equiv U_i(y) - U_i(y-1) > 0$, $y = 1, 2, ...$, denote the marginal utility. Let $\theta_P(y) \equiv \sum_{i \in P} u_i(y)$ be the sum of the marginal utilities of the y'th unit of the agents in $P \subseteq N$, with $\theta_\emptyset(y) \equiv 0$ for all y.

The following assumptions simplify the proof of the propositions.

Assumption 1 *The public good can be provided at most $\bar{y} < \infty$ units.*

Namely, we consider a situation in which there is a capacity for the provision of the public good, beyond which the additional provision is either infeasible or it entails prohibitively high marginal costs.

The next assumption is standard:

Assumption 2 *There exists a unique value of y^*, $0 < y^* \leq \bar{y}$ such that:*

$$\theta_N(y) > c(y) \text{ for all } y \in \{1, 2, \ldots, y^*\}, \text{and } \theta_N(y) < c(y) \text{ for all } y \geq y^* + 1.$$

That is, y^* is the Pareto efficient amount of public good.[2] This includes a standard assumption of diminishing marginal utility and non-decreasing marginal cost.

3 Participation Game

As in the setup by Palfrey and Rothenthal (1984), Saijo and Yamato (1999) and others, we consider a situation where there is no coercive power to enforce the mechanism by compulsory participation of the agents, and any mechanism must take account of participation as a choice variable of the agents.[3] In this section, we first review Saijo and Yamato's (1999) framework and results. We then introduce our unit-by-unit participation game.

A Two-Stage Participation Game

Saijo and Yamato (1999) formally treated the public good provision problem with voluntary participation. They considered the following two-stage game:

Stage 1. *Agents decide* IN *or* OUT.

Stage 2. *Let $T \subseteq N$ be the set of agents who choose* IN *at Stage* 1. *Then a game (S^T, g^T) with the strategy space $S^T = \times_{i \in T} S_i^T$ and the outcome function $g^T : S^T \to R^{\#T+1}$ is played, where*

[2] The model can include the case where there is a high start-up cost in which $\theta_N(1) < c(1)$ but there exists $\hat{y} > 1$ such that $\sum_{i \in N} U_i(\hat{y}) > C(\hat{y})$, with the reformulation that, for $y > 0$, $\hat{U}_i(y) \equiv U_i(\hat{y} - 1 + y)$ and $\hat{C}(y) \equiv C(\hat{y} - 1 + y)$.

[3] As in conventional studies of the public good provision, the structure of the participation game is common knowledge, so that it is a complete information game.

$g^T(s) = (g_x^T(s), g_y^T(s)) = ((x_i)_{i \in T}, y)$ such that $C(y) = \sum_{i \in T} x_i$ determines the payoff of the agents such that $V_i(x_i, y) = U_i(y) - x_i$ for $i \in T$ and $V_i(x_i, y) = U_i(y)$ for $i \in N \setminus T$.

They showed that the participation of all agents is not an equilibrium in many situations. Their result applies to the current environment, as we will illustrate by the following example.

Example 1 Consider the economy with $n = 2$ and the agents have identical preferences represented by $V_i(x_i, y) = U(y) - x_i$, where $u(y) \equiv U(y) - U(y-1)$ is non-increasing in y, and the cost function is $C(y) = y$ for all y.

Let y^f be the allocation that satisfies:

$$y^f = \arg\max_y U(y) - y.$$

This is an allocation in which one agent *free-rides*. Consider the case where y^f is uniquely determined, in which:

$$u(y) > c(y) \text{ if } y \leq y^f \text{ and } u(y) < c(y) \text{ if } y > y^f. \tag{1}$$

Suppose also that $\theta_N(y^f + 1) > c(y^f + 1)$, so that $y^f < y^*$ (y^f is not Pareto efficient level of the public good).

Consider first the game with Saijo-Yamato's setting where g^T is the *voluntary contribution mechanism*, where $x_i = s_i \geq 0,\ y = \sum_{i \in T} s_i$. Suppose that $T = N$ at Stage 1. Then x_1, x_2 and y satisfy:

$$y = x_1 + x_2,\ x_i = \arg\max_{x_i \geq 0} U(x_i + x_j) - x_i\ (i, j = 1, 2). \tag{2}$$

An inspection of (1) shows that the solution of (2) is $y = y^f$ and $\{(x_1, x_2)|x_1 + x_2 = y^f,\ 0 \leq x_1 \leq y^f,\ 0 \leq x_2 \leq y^f\}$. In this setting, if agent i ($i = 1, 2$) participates, agent j ($j = 1, 2, j \neq i$) never has an incentive to participate, since by choosing OUT in Stage 1, agent j can enjoy $V_j(x_j, y) = U(y^f)$, which is at least as good as any outcome when $T = N$. Given that agent j does not contribute to the public good, agent i chooses to participate and produce y^f. Therefore, utility vectors $(V_1, V_2) = (U(y^f) - y^f, U(y^f))$ and $(V_1, V_2) = (U(y^f), U(y^f) - y^f)$ (where only one agent participates, or $T = N$ and only one agent contributes at Stage 2) constitute the Nash equilibria. The game essentially has the structure of the *'Hawk-Dove'* game, such that only one agent contributes at any subgame-perfect equilibrium, so that *the voluntary contribution of all agents is not obtained at the voluntary contribution mechanism.*

Consider next the case where g^T is a *Pareto efficient mechanism*, where $g_y^N(s^*) = y^*$ and $g_y^{\{i\}}(s^*) = y^f$ at any s^* that constitutes an *equilibrium* of the game (S^T, g^T). Examples are the mechanisms by Groves-Ledyard (1977) and Walker (1981) which implement y^* at Nash equilibria. In this case, if y^* and y^f satisfy:

$$2U(y^f) > 2U(y^*) - y^*, \tag{3}$$

it is impossible for $T = N$ in Stage 1 to be an equilibrium.[4] Nothing in the model precludes (3) to

[4] In order to induce the participation of agent i ($i = 1, 2$) in Stage 1, $U(y^f) \leq U(y^*) - x_i$ has to hold. Since $x_1 + x_2 = y^*$, this is incompatible with (3).

hold, so that the participation of all agents and the efficient provision of the public good are not obtained in an equilibrium.[5]

Our aim in this paper is to show that there is a mechanism that takes account of agents' voluntary participation and that implements Pareto efficient allocations.

A Unit-by-unit Participation Game

We now formally introduce our unit-by-unit participation game. It is a multi-stage game, and each stage of the game consists of two sub-stages that correspond to participation and cost-share decisions, respectively. The first stage is the following:

Stage 1.1. *Agents decide* IN *or* OUT.

Stage 1.2. *Let $P \subseteq N$ be the set of agents who choose* IN *at Stage* 1.1. *When $P = \emptyset$, the game ends. When $P \neq \emptyset$, we consider the following stage game among the participants. Each agent $i \in P$ announces $\gamma_i \in \mathbb{R}_{++}$ which determines her contribution.*[6] *The outcome of the stage game is: (i) If $\sum_{j\in P} \gamma_j \geq c(1)$, then the cost share of agent $i \in P$ is $g_i^1 \equiv \dfrac{\gamma_i}{\sum_{j\in P}\gamma_j} c(1)$, and the game continues to Stage 2. (ii) Otherwise, the game ends.*

In Stage 1.1, each agent's *participation decision* is taken for the first unit of the public good. Stage 1.2 is *the cost-share decision among the participants* for the first unit of the public good.

[5]The results are the same with $n > 2$. Saijo and Yamato show through the class of symmetric Cobb-Douglas utility profiles that the measure of the set of the economies for which all agents have participation incentives becomes smaller and converges to zero as n grows large. The same property can hold in our economy, for example, in the class of $U(y) = \dfrac{1}{1-\alpha} y^{1-\alpha}$, $\alpha \in (0, 1)$. (3) holds when $\alpha > 1/2$. The critical value of α which is compatible with full participation is decreasing in n as in Saijo-Yamato's (1999) Figure 2.

[6]Inclusion of zero contribution at the cost-share stage does not affect the nature of the following analysis.

Following the literature of the public good participation game, non-participants do not contribute to the marginal cost $c(1)$ where they can enjoy the benefit if the public good is provided. The game ends if nobody participates, or if the contribution falls short of $c(1)$.

Stages y, $y = 2, 3, ..., \bar{y}$, are defined analogously. When the game continues to Stage y, then:

Stage y.1. *Agents decide* IN *or* OUT.

Stage y.2. *Let $P \subseteq N$ be the set of agents who choose* IN *at Stage y.1. When $P = \emptyset$, the game ends. When $P \neq \emptyset$, we consider the following stage game among the participants. Each agent $i \in P$ announces $\gamma_i \in I\!R_{++}$ which determines her contribution. The outcome of the stage game is: (i) If $\sum_{j \in P} \gamma_j \geq c(y)$, then the cost share of agent $i \in P$ is $g_i^y \equiv \frac{\gamma_i}{\sum_{j \in P} \gamma_j} c(y)$, and the game continues to Stage $y + 1$. (ii) Otherwise, the game ends.*

There are several remarks.

- An important difference from Saijo and Yamato's (1999) game is the notion of participation. Rather than a group of participants choosing the level of public good provision as in Saijo-Yamato's framework, we consider the case where agents make the decision to participate in *each unit* of public good provision. This allows the non-participants of earlier stages to contribute at subsequent stages. A situation where re-negotiation is possible applies to this case.

- The structure of the game is set up in a way that allows a re-entry and an exit in later stages. For a re-entry (or participation at later stages), the agents who choose to participate in the earlier stages may want to ask for a high contribution from those who participate later. This is possible if the former continues to choose IN and get such an agreement in the cost-share stage. The possibility against punishment for exiting substantially complicates the analysis, and we do not consider this here.

Let us introduce some notations to facilitate the formal treatments in the following sections. A (pure) strategy of agent i at Stage y, s_i^y, specifies: (1) agent i's action that specifies IN or OUT at Stage $y.1$, and (2) a function $\gamma_i^y : \{P \subseteq N | i \in P\} \rightarrow I\!R_{++}$ that describes agent i's action γ_i at Stage $y.2$ when she participates at Stage $y.1$ and the set of participants is P. In general, such actions may depend on a history (a sequence of actions in previous stages). However, in order to simplify the analysis, we introduce the following condition of *History-Independence* to the strategy. Let $s^y \equiv (s_i^y)_{i \in N}$, and let S^y denote the set of s^y. Let $P^y : S^y \rightarrow 2^N$ specify the set of participants at Stage $y.1$.

History-Independence: $P^y(s^y)$ and $\gamma_i^y(P)$ are independent of the history for all y, all i and all $P \subseteq N$ $(i \in P)$.[7]

[7] Formally, let A_i^1 be a choice set of agent i at Stage 1. An action $a_i^1 \in A_i^1$ specifies: (1) IN or OUT at Stage 1.1, and (2) a function $\gamma_i^1 : \{P \subseteq N | i \in P\} \rightarrow I\!R_{++}$ that describes agent i's action γ_i at Stage 1.2 when she participates and the set of participants is P. The set of possible actions at Stage y $(y \geq 2)$ is defined iteratively. Given any sequence of actions in previous stages, i.e., $(a_i^{\hat{y}})_{i \in N}$, $\hat{y} \leq y-1$, let $h^y \equiv ((a_i^1)_{i \in N}, ..., (a_i^{y-1})_{i \in N}) \in H^y$ be the history at the end of Stage $y-1$, and we let $A_i^y(h^y)$ denote agent i's feasible actions in Stage y. A (pure) strategy of agent i at Stage y, s_i^y, specifies: (1) agent i's action that specifies IN or OUT at Stage $y.1$ (which may depend on h^y), and (2) a function $\gamma_i^y : H^y \times \{P \subseteq N | i \in P\} \rightarrow I\!R_{++}$ that describes agent i's action γ_i at Stage $y.2$ when the history is given by h^y, she participates at Stage $y.1$, and the set of participants is P. Let $P^y : H^y \times S^y \rightarrow 2^N$ specify the set of participants at Stage $y.1$. History-Independence implies that $P^y(h^y, s^y)$ and $\gamma_i^y(h^y, P)$ are independent of h^y for all y, all i and all $P \subseteq N$ $(i \in P)$.

Namely, we consider a situation where the actions of each agent at any Stage y are not dependent on the set of participants or the cost sharing at previous stages. There are potentially many possibilities for history-dependent strategies. However, it is hard to predict how other agents may alter their strategies contingent on the past, especially in a non-cooperative situation. The assumption of history-independence gives a simple guidance to agents, and thus it is a natural starting point.[8]

Associated with s^y, let $\kappa^y(s^y|P) \equiv (\kappa_i^y(s^y|P))_{i\in N}$ be an outcome function such that, if $\sum_{j\in P}\gamma_j \geq c(y)$ in Stage $y.2$ associated with a set of participants P in Stage $y.1$, then $\kappa_i^y(s^y|P) = g_i^y$ for all $i \in P$ and $\kappa_i^y(s^y|P) = 0$ for all $i \notin P$, and $\kappa^y(s^y|P) = 0^{\#N}$ otherwise. Let $g^y(s^y) \equiv \kappa^y(s^y|P^y(s^y))$. Given strategy $s \equiv (s^y)_{y\leq\bar{y}}$, when the participants choose to continue up to Stage $\tilde{y}$ and the game ends at Stage $\tilde{y}+1$, the payoff of agent i is $U_i(\tilde{y}) - \sum_{y=1}^{\tilde{y}} g_i^y(s^y)$.

4 Efficiency Is Compatible with Voluntary Participation

To examine the equilibrium allocations of this game, we apply the subgame-perfect equilibrium. Our first result is that efficiency is compatible with agents' voluntary participation decisions. We first introduce our first proposition and a sketch of the proof (which consists of three steps) to illuminate how the unit-by-unit game works for an efficient allocation. We also illustrate with a numerical example employed at Example 1.

[8] The points we want to make in the following analyses are: (1) efficiency is compatible with agents' voluntary participation; (2) only efficient equilibria are robust against refinements. If we allow history-dependent strategies, the set of efficient subgame-perfect equilibria would expand, but history-independence is sufficient to prove the existence of efficient subgame-perfect equilibrium. Also, inefficient equilibria can be eliminated by refinement concepts we examine below, even if we allow history-dependent strategies.

Proposition 1 *There is a subgame-perfect equilibrium supporting an efficient allocation.*

In an Appendix, we construct the following subgame-perfect equilibrium strategy s^*.

Step 1 The game ends by Stage $y^* + 1$ at s^*.

By Assumption 2, the aggregate benefit falls short of the marginal cost beyond $(y^* + 1)$'th unit, so that nobody has an incentive to participate and bear the cost of the public good.

Step 2 The following strategy exists, which constitutes the equilibrium strategy at Stage y^*:

Stage $y^*.1$: $P^{y^*}(s^{*y^*}) = P^{y^*}$ such that:

$$\theta_{P^{y^*}}(y^*) \geq c(y^*) \text{ and } \theta_{P^{y^*}\setminus\{i\}}(y^*) < c(y^*) \text{ for all } i \in P^{y^*}. \tag{4}$$

Stage $y^*.2$: for all $i \in P^{y^*}$, $\gamma_i^{y^*}(P^{y^*}) = g_i^{y^*}$ such that $\sum_{i \in P^{y^*}} g_i^{y^*} = c(y^*)$ and $0 < g_i^{y^*} \leq u_i(y^*)$ for all $i \in P^{y^*}$.

An intuition of voluntary participation induced by the set of participants depicted at (4) is the following. For *every* $i \in P^y(s^{*y^*})$, if she does not join, the y^*'th unit of the public good is not provided: since the total benefits that $P^y(s^{*y^*}) \setminus \{i\}$ can obtain fall short of the marginal cost, $P^y(s^{*y^*}) \setminus \{i\}$ cannot allocate the cost-sharing in order for everyone to receive a non-negative net benefit by participation. Since the net benefit of participation $(u_i(y^*) - g_i^{y^*})$ is set non-negative, no agent in P^{y^*} can be strictly better off by non-participation. In this sense, every $i \in P^y(s^{*y^*})$ is 'pivotal', in that her incentive is compatible with the efficient provision of the public good. In the case of the binary provision of the public good, such a set of participants is crucial for efficient provision with voluntary participation (see Palfrey and Rothenthal (1984) and Shinohara (2004)).

The strategies at Stage $\hat{y}$ $(0 < \hat{y} < y^*)$ are constructed by backward induction, given the sequentially-rational behavior at the following stages. Associated with s^*, let $g^y \equiv (g_i^y(s^{*y}))_{i\in N}$ be the equilibrium cost sharing at Stage y, $\hat{y}+1 \leq y \leq y^*$, and let $g^{\hat{y}+1,y^*} \equiv (g^{\hat{y}+1}, ..., g^{y^*})$ denote the list of equilibrium cost shares. Let $\phi_i(\hat{y}, y^*; g^{\hat{y}+1,y^*}) \equiv u_i(\hat{y}) + \sum_{y=\hat{y}+1}^{y^*}(u_i(y) - g_i^y)$ be agent i's benefit of the subsequent stages in addition to her gross marginal benefit at $\hat{y}$'th unit. Let $\psi_P(\hat{y}, y^*; g^{\hat{y}+1,y^*}) \equiv \sum_{i\in P}\phi_i(\hat{y}, y^*; g^{\hat{y}+1,y^*})$. As the last step of the proof, the following is shown:

Step 3 The following strategy exists, which constitutes the equilibrium strategy at Stage $\hat{y}$, $0 < \hat{y} < y^*$:

Stage $\hat{y}$.1: $P^{\hat{y}}(s^{*\hat{y}}) = P^{\hat{y}}$ such that:

$$\psi_{P^{\hat{y}}}(\hat{y}, y^*; g^{\hat{y}+1,y^*}) \geq c(\hat{y}) \text{ and } \psi_{P^{\hat{y}}\setminus\{i\}}(\hat{y}, y^*; g^{\hat{y}+1,y^*}) < c(\hat{y}) \text{ for all } i \in P^{\hat{y}}. \tag{5}$$

Stage $\hat{y}$.2: $\gamma_i^{\hat{y}}(P^{\hat{y}}) = g_i^{\hat{y}}$ for all $i \in P^{\hat{y}}$ such that $\sum_{i\in P^{\hat{y}}} g_i^{\hat{y}} = c(\hat{y})$ and $0 < g_i^{\hat{y}} \leq \phi_i(\hat{y}, y^*; g^{\hat{y}+1,y^*})$ for all $i \in P^{\hat{y}}$.

Here, each agent takes account of the net benefit, including those of subsequent stages, for their participation and cost-share decisions. The logic of voluntary participation by $P^{\hat{y}}(s^{*\hat{y}})$ is the same as $P^{y^*}(s^{*y^*})$: for every $i \in P^{\hat{y}}(s^{*\hat{y}})$, if she does not join, the total benefits including the benefits at subsequent stages by $P^{\hat{y}}(s^{*\hat{y}}) \setminus \{i\}$ $(\psi_{P^{\hat{y}}\setminus\{i\}}(\hat{y}, y^*; g^{\hat{y}+1,y^*}))$ fall short of the marginal cost $(c(\hat{y}))$, so that the public good is no longer provided. Again, every $i \in P^{\hat{y}}(s^{*\hat{y}})$ is pivotal. When agents make the decision to participate in each unit of public good provision, this logic applies until it reaches the y^*'th unit, so that efficiency is compatible with voluntary participation.

Let us illustrate our points through a numerical example.

Example 1, Continued Consider the case of $n = 2$ with identical utility functions and a linear production technology. Condition (3), which prevents the participation of all agents, holds if, for example,

$$u(y) = 1.65 - 0.15y \text{ for } y \leq 10 \text{ and } u(y) = 0.15 \text{ for } 11 \leq y \leq \bar{y}. \tag{6}$$

What about our case of unit-by-unit participation? It can be shown that $y^f = 4$ and $y^* = 7$. Following Steps 2 and 3 of Proposition 1, one can find an efficient subgame-perfect equilibrium by backward induction. Start with $y = y^* = 7$. By (4), $P^7 = N$. When $y = 6$, $\psi_N(6, 7; g^{7,7}) > c(6)$ and $\phi_i(6, 7; g^{7,7}) < c(6)$ for any $g^{7,7}$ consistent with the equilibrium cost sharing,[9] so that $P^6 = N$ by (5). When $y = 5$, it has to be the case for any $g^{6,7}$ that $\phi_i(5, 7; g^{6,7}) > c(5)$ for some $i = 1, 2$,[10] so that $P^5 = \{1\}$ or $\{2\}$, depending on the subsequent cost sharing. When $y \leq 4$, since the construction of the equilibrium strategy implies $\phi_i(y, 7; g^{y+1,7}) = u_i(y) + \phi_i(y + 1, 7; g^{y+2,7}) - g_i^{y+1} \geq u_i(y) > c(y)$, so that $P^y = \{1\}$ or $\{2\}$ by (5). In summary, the following set of participants is consistent with an

[9]Consider agent $j \neq i$'s contribution. Since $u_j(7) = 0.6$, the highest contribution that agent j is willing to contribute at Stage 7.2 is 0.6, where i can receive the highest marginal benefit $u_i(7) - 0.4 = 0.2$. Therefore, $\phi_i(6, 7; g^{7,7}) \leq u_i(6) + 0.2 = 0.95 < 1 = c(6)$, so that $P^6(s^{*6}) = N$ has to be the case at any equilibrium.

[10]It can be shown that $\psi_N(5, 7; g^{6,7}) = 2u(5) + \psi_N(6, 7; g^{7,7}) - 1 = 2.5$. Since $\psi_N(5, 7; g^{6,7}) = \sum_{i=1}^{2} \phi_i(5, 7; g^{6,7})$, it has to be the case that $\phi_i(5, 7; g^{6,7}) > 1$ for some $i = 1, 2$.

efficient subgame-perfect equilibrium:

$$
\begin{aligned}
P^y(s^{*y}) &= \{1\} \text{ or } \{2\} \text{ if } y \leq y^f + 1 \\
&= N \text{ if } y^f + 1 < y \leq y^*.
\end{aligned}
$$

As you see, the set of the subgame-perfect equilibria is substantially larger than the one examined in Saijo and Yamato. Condition (3) which prevents the *participation* of all agents (hence implementation of $y = y^*$ in Saijo and Yamato) is irrelevant for the *implementation* of $y = y^*$ in our case. The difference is that, whereas an inefficient provision of $y = y^f$ when $T \neq N$ is enforced in Saijo-Yamato's mechanism, such an allocation is subject to re-negotiation, as long as there is an efficiency gain. The unit-by-unit participation mechanism allows to exploit an efficiency gain in a way which is compatible with agents' voluntary participation.

From the above example, one can easily see that there is an *inefficient* subgame-perfect equilibrium in general. An outcome with $P^y(s^y) = \emptyset$ for Stage y, $y > y^f$ is consistent with a subgame-perfect equilibrium, since each agent will choose OUT given the other agent chooses so. In the following, we justify that the equilibrium outcomes shown in Proposition 1 are more reasonable to achieve.

5 Strict Subgame-Perfect Equilibria

The nature of non-excludability inevitably generates multiple equilibria. We propose here to strengthen the equilibrium. To clarify our point, consider the following example:

Example 2 Consider the case of $n = 3$, $y^* = 1$, $u_i(1) = u$ for all i, and $c(1) \in (u, 2u)$. Consider the following strategy $\tilde{s}$:

Stage 1.1: $P^1(\tilde{s}^1) = N$.

Stage 1.2: For all $i \in N$, $\gamma_i^1(N) = \dfrac{c(1)}{3}$, and $\gamma_i^1(P) = \epsilon < c(1) - u$ if $P \neq N$.

If any agent chooses non-participation, the sum of the announced cost-share does not cover $c(1)$, so that the public good is not produced. Since $u > c(1)/3$, this strategy constitutes a subgame-perfect equilibrium. However, N does not satisfy (4).

A questionable behavior in this example is the one at P with $|P| = 2$. Since $u > c(1)/2$, there is a scope for cooperative gain at Stage 1.2 which is not fully exploited, but cooperation emerges at $P = N$. The games examined by Muto (1986) and Okada (1993) have similar features, and they impose conditions stronger than subgame-perfection. Here, we propose a refinement concept of a *strict perfect equilibrium* which is similar to Muto (1986).

Let us take any subgame starting from Stage y.1, and denote the payoff of agent i starting from Stage y by $R_i^y : N \times I\!R_+^n \times (S^{\hat{y}})_{y<\hat{y}\leq\bar{y}} \to I\!R$ that makes: (1) a set of participants $P \subseteq N$ at Stage y.1, (2) the outcome (cost-sharing) g^y of Stage y, and (3) the strategies at subsequent stages $b^y \equiv (s^{\hat{y}})_{y<\hat{y}\leq\bar{y}}$ (with $b^{\bar{y}} \equiv \emptyset$), correspond to a real number $R_i^y(P, g^y, b^y)$. When the game first ends at Stage $\tilde{y} \geq y + 1$ after Stage y, $R_i^y(P, g^y, b^y) \equiv \sum_{\hat{y}=y}^{\tilde{y}-1}(u_i(\hat{y}) - g_i^{\hat{y}}(s^{\hat{y}}))$. When the game ends at Stage y, $R_i^y(P, g^y, b^y) \equiv 0$.

Definition 1 (Strict Subgame-Perfect Equilibria) *Associated with strategy s^*, let $P^y(s^{*y})$ be* *a set of participants at Stage y, $g^y(s^{*y})$ be the outcome at Stage y. A strategy s^* is a Strict Subgame-Perfect Equilibrium iff it is a subgame-perfect equilibrium such that, for all $y \leq \bar{y}$ and $i \in P^y(s^{*y})$,*

$$R_i^y(P^y(s^{*y}), g^y(s^{*y}), b^{*y}) \geq R_i^y(P^y(s^{*y}) \setminus \{i\}, \tilde{g}^y, b^{*y}), \tag{7}$$

for **every** *subgame-perfect equilibrium cost-sharing vector $\tilde{g}^y$ that is attained at Stage $y.2$ whose set of participants is $P^y(s^{*y}) \setminus \{i\}$, and, for all $i \notin P^y(s^{*y})$,*

$$R_i^y(P^y(s^{*y}), g^y(s^{*y}), b^{*y}) \geq R_i^y(P^y(s^{*y}) \cup \{i\}, \tilde{g}^y, b^{*y}), \tag{8}$$

for **every** *subgame-perfect equilibrium cost-sharing vector $\tilde{g}^y$ that is attained at Stage $y.2$ whose set of participants is $P^y(s^{*y}) \cup \{i\}$.*

According to Muto (1986), the notion of strict subgame-perfect equilibrium is grounded upon the idea that every agent will not deviate from the equilibrium at each of her decision points even if she supposes the equilibrium is most favorable to her.[11] In Example 2, if agent i deviates from $\tilde{s}$ at Stage 1.1 by choosing OUT, the best scenario for her is the public good produced by $N \setminus \{i\}$. This can occur at a subgame-perfect equilibrium since $u > c(1)/2$. Therefore, (7) is not satisfied, so that $\tilde{s}$ is not a strict subgame-perfect equilibrium.[12] In general, we can show the following:

[11]By the assumption of history-independence, the strategy of the subsequent stages is unchanged. One can change the last part of the RHS's to $b^{*y} \setminus \tilde{b}_i^y$, but the best response of agent i against $(b_j^{*y})_{j \neq i}$ is b_i^{*y}.

[12]Okada (1993) proposed the notion of *payoff-dominance* under an assumption of identical agents, which basically requires the agents to adopt a strategy that is Pareto efficient among the participants. The application of his payoff dominance to the current environment also eliminates an equilibrium illustrated at Example 2.

Proposition 2 *Any strategy constructed at Steps 2 and 3 of Proposition 1 is a strict subgame-perfect equilibrium. Conversely, at any strict subgame-perfect equilibrium that supports an efficient allocation, the sets of participants at Stage y^* and Stage $\hat{y}$ satisfy (4) and (5) of Proposition 1, respectively.*

6 Coalition-Proofness

When there is no coercive power for participation for a public good provision, we often observe negotiations among members towards the mutual benefits in participation and in cost-share decisions.[13] As a reduced form of such a negotiation process, we now examine the implications of group self-enforcing behavior, based on the theoretical formulations in the literature.

The notion of group self-enforcing behavior is particularly important in understanding collective decision making. Self-enforceability with respect to individual deviation is insufficient in a situation where coalitions of agents can arrange plausible and mutually beneficial deviations from Nash agreements. Bernheim, Peleg and Whinston (1987) propose the notion of a *coalition-proof Nash equilibrium.* They also consider games in extensive form, and propose the notion of a *perfectly coalition-proof Nash equilibrium* (PCPNE). Their concept is based on a stronger concept of a *strong perfect equilibrium* by Rubinstein (1980), which requires that there is no mutually profitable deviation at any subgame. Here, we adopt Rubinstein's equilibrium concept into a unit-by-unit par-

[13] A real-world example includes the case of Russia's approval to Kyoto Protocol. Russia's participation was vital for all signatories, since the country is the world's second largest source of greenhouse gases. For example, the European Union (EU) proposed that Russia ratify the Protocol in exchange for the EU's support for Russia's entry into the World Trade Organization, an issue which the Russian administration had long been trying to achieve. Such negotiations were essential for Russia's approval of the treaty.

ticipation game, which is an extension of Shinohara (2005). Let $s_T^y \equiv (s_i^y)_{i\in T}$ and $\tilde{b}_T^y \equiv (\tilde{s}_T^{\hat{y}})_{y<\hat{y}\leq\bar{y}}$ (with $b_T^{\bar{y}} \equiv \emptyset$).

Definition 2 (Strong Perfect Equilibria) *A strategy profile s^* is a strong perfect equilibrium if there exists no y ($1 \leq y \leq \bar{y}$), $T \subseteq N$, $\tilde{s}_T$, and $P \supseteq T$ such that:*

$$\begin{aligned}
&\sum_{i\in T} R_i^y(P^y(s_{N\backslash T}^{*y}, \tilde{s}_T^y), g^y(s_{N\backslash T}^{*y}, \tilde{s}_T^y), b_{N\backslash T}^{*y}, \tilde{b}_T^y) > \sum_{i\in T} R_i^y(P^y(s^{*y}), g^y(s^{*y}), b^{*y}) \text{ or} \\
&\sum_{i\in T} R_i^y(P, \kappa^y(s_{N\backslash T}^{*y}, \tilde{s}_T^y|P), b_{N\backslash T}^{*y}, \tilde{b}_T^y) > \sum_{i\in T} R_i^y(P, \kappa^y(s^{*y}|P), b^{*y}).
\end{aligned} \tag{9}$$

The definition of a strong perfect equilibrium ensures that, in addition to the requirement of subgame-perfection that the action has to be dynamically consistent, no proper subgroup of agents should be able to make a mutually advantageous deviation from the agreement in any subgame. Here, a deviation is defined in terms of a sum of utilities among agents in a coalition, which is stronger than the requirements by Aumann (1959) and Rubinstein (1980). A strong perfect equilibrium strategy, if it exists, is robust against any deviation *even when monetary transfers among agents in coalitions are possible.* This is a demanding concept, and many games that are of interest do not have a strong perfect equilibrium.[14] Another modification applied to this context is the form of deviation at the second stage. Here, we restrict a class of potential deviations to subsets of participants, those who attended the meeting for the contribution towards the public good provision. Whereas it is a natural restriction consistent with our notion of participation, this

[14]Bernheim, Peleg and Whinston (1987) pointed out that the definition of the strong perfect equilibrium does not guarantee *self-enforcement* on the class of deviation. The class of games where such equilibrium exists is limited (see Aumann (1959) and Rubinstein (1980)).

point will be discussed later.

We now prove the following:

Proposition 3 *There exists a strong perfect equilibrium supporting an efficient allocation. Moreover, no subgame-perfect equilibrium that supports an inefficient allocation constitutes a strong perfect equilibrium.*

Proof We show first that there is a strategy profile s^* that constitutes a strong perfect equilibrium. We show next that no inefficient subgame-perfect equilibrium is a strong perfect equilibrium, by constructing a deviation that is similar to s^*.

Step 1 Let Q^{y^*} be a set of participants such that:

$$Q^{y^*} \in \arg\min_{P^{y^*}} \theta_{P^{y^*}}(y^*) \text{ s.t. (4).} \tag{10}$$

For this Q^{y^*}, construct a cost-share vector $(g_i^{y^*})_{i \in Q^{y^*}}$ as in Step 2 of Proposition 1. For $\hat{y}$ such that $0 < \hat{y} < y^*$, given $g^{\hat{y}+1,y^*}$ constructed inductively, let $Q^{\hat{y}}$ be a set of participants such that:

$$Q^{\hat{y}} \in \arg\min_{P^{\hat{y}}} \psi_{P^{\hat{y}}}(\hat{y}, y^*; g^{\hat{y}+1,y^*}) \text{ s.t. (5).} \tag{11}$$

For this $Q^{\hat{y}}$, construct a cost-share vector $(g_i^{\hat{y}})_{i \in Q^{\hat{y}}}$ as in Step 3 of Proposition 1. For these sets of participants, there is a strategy that is robust against any deviation.

The proof of this step is given in an Appendix.

Step 2 No inefficient subgame-perfect equilibrium is a strong perfect equilibrium.

Proof of Step 2. Consider an inefficient subgame-perfect equilibrium $\tilde{s}$, at which the game ends at Stage $\tilde{y}$. By Step 1 of Proposition 1, $\tilde{y} < y^* + 1$. Consider a subgame starting from Stage $\tilde{y}.1$, and consider the following deviation: construct a strategy profile for Stage $\hat{y}$, $y^* \geq \hat{y} \geq \tilde{y}$, in the same way as those of Step 1 above, by the set of agents $\cup_{y^* > \hat{y} \geq \tilde{y}} Q^{\hat{y}} \cup Q^{y^*}$ that satisfy (10) and (11). By construction, this is a profitable deviation (notice that $\phi_i(y, y^*; g^{y+1,y^*}) - g_i^y(s^{*y}) \geq 0$ for all i and all $y \leq y^*$). **(End of Proof of Step 2)**

By Steps 1 and 2, y^* units of the public good is produced at a strong perfect equilibrium. *Q.E.D.*

An efficient subgame-perfect equilibrium constructed at Step 1 of the proof is an extension of Shinohara (2005) in the context of binary provision. The sets of participants that satisfy (10) and (11) are the smallest sets among those which satisfy (4) and (5).

There are two remarks regarding Proposition 3. First, we discuss on our restriction of $P \supseteq T$ (limiting a class of potential deviations to subsets of participants) at Stage $y.2$. If $\psi_P(y, y^*; g^{y+1}, y^*) \geq c(y)$, this restriction is not necessary: even if a coalitional deviation occurs by a set of agents T including participants and non-participants, there is no $\tilde{s}_T$ such that $\sum_{i \in T} R_i^y(P, \kappa^y(s^{*y}_{N \setminus T}, \tilde{s}^y_T | P), b^{*y}_{N \setminus T}, \tilde{b}^y_T)$ $> \sum_{i \in T} R_i^y(P, \kappa^y(s^{*y}|P), b^{*y})$. When $\psi_P(y, y^*; g^{y+1}, y^*) < c(y)$, by setting $\gamma_i^y(P) = 0$ for all $i \in P$ (which is consistent with the equilibrium), no deviation occurs by $T \subseteq N$ such that $\sum_{i \in T} R_i^y(P, \kappa^y(s^{*y}_{N \setminus T}, \tilde{s}^y_T | P), b^{*y}_{N \setminus T}, \tilde{b}^y_T) > \sum_{i \in T} R_i^y(P^y(s^{*y}), g^y(s^{*y}), b^{*y})$: even though a deviation by participants and the non-participants is aimed at Stage $y.2$, the total benefit falls short of the equilibrium total benefit.[15] Therefore, a coalitional deviation does not affect participation and cost-share decisions along the equilibrium path.

Second, we can show that the strategies shown at Proposition 1 are *only* strategies that satisfy a weaker version of a strong perfect equilibrium in the following sense. For such s^*, there is no y $(1 \leq y \leq \bar{y})$, $T \subseteq N$, $\tilde{s}_T$, and $P \supseteq T$ such that:

$$(R_i^y(P^y(s_{N\backslash T}^{*y}, \tilde{s}_T^y), g^y(s_{N\backslash T}^{*y}, \tilde{s}_T^y), b^{*y}))_{i\in T} \geq (R_i^y(P^y(s^{*y}), g^y(s^{*y}), b^{*y}))_{i\in T} \text{ or}$$

$$(R_i^y(P, \kappa^y(s_{N\backslash T}^{*y}, \tilde{s}_T^y|P), b^{*y}))_{i\in T} \geq (R_i^y(P, \kappa^y(s^{*y}|P), b^{*y}))_{i\in T},$$

where the inequality $(\alpha_i)_{i\in T} \geq (\beta_i)_{i\in T}$ shows $\alpha_i \geq \beta_i$ for all $i \in T$ with at least one strict inequality. When the deviations are allowed with multiple stages, or when monetary transfers among agents in coalitions are possible, there may be a potential for coalitional deviations.[16] The strategy constructed by the sets of participants that satisfy (10) and (11) can eliminate such potentials.

Two corollaries arise. The first corollary concerns the PCPNE by Bernheim, Peleg and Whinston (1987). Similar to the strong perfect equilibrium, a PCPNE prevents a mutually advantageous deviation by a coalition in any subgame.[17] However, unlike the strong perfect equilibrium, the

[15]By setting $\gamma_i^y(P) = 0$ for all $i \in P$ (which is consistent with the equilibrium, since $\psi_P(y, y^*; g^{y+1}, y^*) < c(y)$), the total benefit by $T \subseteq N$ from a deviation is at most $\sum_{i\in T} u_i(y) - \left(c(y) - \sum_{j\in P\backslash T} \gamma_j^y(P)\right) + \sum_{\hat{y}=y+1}^{y^*} \sum_{i\in T} \left(u_i(\hat{y}) - g_i^{\hat{y}}(s^{*\hat{y}})\right) = \psi_T(y, y^*; g^{y+1}, y^*) - c(y)$. On the other hand, $\sum_{i\in T} R_i^y(P^y(s^{*y}), g^y(s^{*y}), b^{*y}) = \psi_T(y, y^*; g^{y+1}, y^*) - \sum_{i\in T} g_i^y(s^{*y}) \geq \psi_T(y, y^*; g^{y+1}, y^*) - c(y)$. Therefore, the claim holds.

See also footnote 16 below.

[16]However, one can show that a deviation at Stages y_1 and y_2 $(y_1 < y_2)$ by a coalition T is profitable only if, among T, some gainers at Stage y_1 compromise at Stage y_2 to incur some loss that is less than their gains at previous stages. When the commitment is infeasible, such agents may not agree with the deviation *ex post* at Stage y_2, so that such deviations show a lack of credibility. Similar comment applies for monetary transfers.

[17]See Bernheim, Peleg and Whinston (1987, p.11) for a formal definition. The definition of a PCPNE is defined

class of deviations under consideration are those that satisfy similar criteria: the deviations have to be dynamically consistent and self-enforcing. Any strong perfect equilibrium is a PCPNE. In general, neither the *existence* nor the *efficiency* of a PCPNE is guaranteed. However, Step 2 of Proposition 3 shows that *no subgame-perfect equilibrium that supports an inefficient allocation constitutes a PCPNE*. This is because the deviation constructed at Step 2 is self-enforcing (there is no further deviation against such deviation, as shown at Step 1).

Another corollary is a relationship between a strict subgame-perfect equilibrium and a strong perfect equilibrium. Although they are conceptually not related, in the context of the unit-by-unit participation game, every strong perfect equilibrium is a strict subgame-perfect equilibrium (that is Pareto efficient). We conclude the following:

Corollary 1 *(i) There exists a perfectly coalition-proof Nash equilibrium supporting a Pareto efficient allocation. Moreover, no subgame-perfect equilibrium that supports an inefficient allocation constitutes a perfectly coalition-proof Nash equilibrium.*

(ii) Every strong perfect equilibrium is a strict subgame-perfect equilibrium.

7 Conclusion

Using the conventional model of the public good provision, we constructed a mechanism where efficiency is compatible with agents' voluntary participation. The intuition is that every participant

inductively with respect to the number of agents and the number of stages. For a one-stage game with one agent, PCPNE is a strategy that maximizes the agent's utility. A strategy is *perfectly self-enforcing* if the strategy of any proper subset constitutes a PCPNE at any proper subgame, given the strategy of the others. A strategy is a PCPNE if it is Pareto efficient among perfectly self-enforcing strategies.

is ‘pivotal’, in that her incentive is compatible with the efficient provision of the public good. Using the notions of Rubinstein’s (1980) strong perfect equilibrium and Bernheim, Peleg and Whinston’s (1987) Perfect Coalition-Proof Nash equilibrium, we showed that only efficient equilibria are robust against refinements.

It has long been thought that it is difficult to provide a non-excludable public good since every agent has an incentive to free-ride. This is an important unsolved problem, and we provide a positive answer to cope with this issue. Furthermore, the unit-by-unit method may have a practical implication when we face a real provision of a public good, including the resolution of global environmental problems discussed in the Introduction. We hope that this paper will be a beginning for further studies of non-excludable public good provision.

Acknowledgment

The previous version of this paper was presented at the 39th Annual Meeting of the Canadian Economics Association, McMaster University, Canada (May 2005), the Asian Decentralization Conference 2005, Seoul National University, Korea (May 2005), Public Economic Theory 2005, Marseille, Universite de la Mediterranee, France (June 2005), the 61th Congress of the International Institute of Public Finance, ICC Jeju, Korea (August 2005), the 62th Congress of the Japanese Institute of Public Finance, Hitotsubashi University, Japan (October 2005), Professor Fukukane Nikaido Memorial Conference, Hitotsubashi University, Japan (March 2006), the Spring Meeting of Japanese Economics Association, Fukushima University, Japan (June 2006), the Eighth International Meeting of the Society for Social Choice and Welfare, Crowne Plaza Hotel, Turkey (July 2006), and seminars at University of Rochester, Tokyo University, Hitotsubashi University, Osaka University and Yokohama National University. We are grateful for comments and discussions by people including Koichi Tadenuma, Tatsuyoshi Saijo, Hiroo Sasaki, Akira Okada, Robert Moir, Stuart Mestelman, Steeve Mongrain, Yoshitomo Ogawa, Mikio Nakayama and Hiroshi Ohta. The usual disclaimer applies. Financial assistance from the Japan Economic Research Foundation is gratefully acknowledged. Nishimura additionally acknowledges the financial support from JSPS (Japan Society for the Promotion of Science) Postdoctoral Fellowships for Research Abroad.

Appendix

Proof of Proposition 1. We show the statement by proving that there is a subgame-perfect equilibrium s^* that is efficient.

Step 1 The game ends by Stage $y^* + 1$ at s^*.

Proof of Step 1. Suppose that the game continues up to Stage $y \geq y^* + 1$ and ends at Stage $y+1$ at s^*. Since all agents in $P^y(s^{*y})$ prefer to continue the game at Stage y, it has to be the case that $u_i(y) - g_i^y \geq 0$ for all $i \in P^y(s^{*y})$, which implies that $\theta_{P^y(s^{*y})}(y) - c(y) \geq 0$ by construction. However, by Assumption 2, $\theta_P(y) \leq \theta_N(y) < c(y)$ for all $P \subseteq N$ and all $y > y^*$. This is a contradiction. **(End of Proof of Step 1)**[18]

Step 2 Suppose that the game continues up to Stage $y^* - 1$. Then a set P^{y^*} at (4) and a cost-share vector $(g_i^{y^*})_{i \in P^{y^*}}$ constructed in the text constitute the strategy profile at Stage y^*.

Proof of Step 2. By Assumption 2, there is at least one set of participants satisfying (4). We can find a cost-share vector $(g_i^{y^*})_{i \in P^{y^*}}$ that satisfies $\sum_{i \in P^{y^*}} g_i^{y^*} = c(y^*)$ and $0 < g_i^{y^*} \leq u_i(y^*)$ for all $i \in P^{y^*}$.

We now show that the strategy constructed at the text constitutes a subgame-perfect equilibrium. If $i \in P^{y^*}$ chooses IN at Stage $y^*.1$, she obtains the marginal benefit $u_i(y^*) - g_i^{y^*}$ when $\gamma_i = g_i^{y^*}$ for all $i \in P^{y^*}$. This is the highest marginal benefit that she can obtain, since the marginal

[18] Assumption 1 is implicitly used in this part by supposing that the game finishes in finite stages.

benefit is 0 if $\gamma_i < g_i^{y^*}$, and if $\gamma_i > g_i^{y^*}$, she pays too much. Suppose that $i \in P^{y^*}$ chooses OUT at Stage $y^*.1$. Since $\theta_{P^{y^*}\setminus\{i\}}(y^*) < c(y^*)$, for any cost-share vector $(\tilde{g}_j^{y^*})_{j\in P^{y^*}\setminus\{i\}}$ that satisfies $\sum_{j\in P^{y^*}\setminus\{i\}} \tilde{g}_j^{y^*} = c(y^*)$, there arises agent $j \in P^{y^*} \setminus \{i\}$ such that $u_j(y^*) - \tilde{g}_j^{y^*} < 0$. Since we showed at Step 1 that the game ends by Stage y^*+1, such agent j will not choose her action to bear the cost-share of $\tilde{g}_j^{y^*}$. Therefore, the set $P^{y^*} \setminus \{i\}$ chooses not to provide the y^*'th unit of the public good at any equilibrium, so that agent i receives 0. Therefore, no agent in P^{y^*} can be strictly better off by choosing OUT or choosing IN and not cooperating in the cost-share decision, given the above strategy.

Let $i \notin P^{y^*}$. If agent i chooses IN, then she receives either the marginal benefit of $u_i(y^*) - \tilde{g}_i^{y^*}$ (where $\tilde{g}_i^{y^*} > 0$ is determined according to the strategy profile at Stage $y^*.2$), or she receives 0 (in the case where the total contribution falls short of $c(y^*)$). On the other hand, i obtains $u_i(y^*) > 0$ if i selects OUT. Since $\tilde{g}_i^{y^*} > 0$, no agent in $N \setminus P^{y^*}$ strictly prefers IN to OUT, given that P^{y^*} is the set of participants at Stage y^*. **(End of Proof of Step 2)**

The strategies at Stage $\hat{y}$ $(0 < \hat{y} < y^*)$ are constructed by backward induction, given the sequentially-rational behavior at the following stages:

Step 3 Consider Stage $\hat{y}$, $0 < \hat{y} < y^*$. Suppose that, associated with s^*, for all $y \in \{\hat{y}+1, ..., y^*\}$, the strategy s^{*y} at Stage y is given in a way such that the y'th unit of the public good is produced. Then a set $P^{\hat{y}}$ at (5) and a cost-share vector $(g_i^{y^*})_{i\in P^{\hat{y}}}$ constructed in the text constitute the strategy profile at Stage $\hat{y}$.

Proof of Step 3. For $\hat{y} = y^* - k$ $(0 < k < y^*)$, we construct inductively a set P^{y^*-k} at (5) and a cost-share vector $g^{y^*-k} \equiv (g_i^{y^*-k}(s^{*y^*-k}))_{i \in N}$ as claimed in the text. First, by the construction of $(g_i^{y^*})_{i \in P^{y^*}}$ at Step 2, with $g_i^{y^*} = 0$ for all $i \notin P^{y^*}$, $u_i(y^*) \geq g_i^{y^*}$ for all $i \in N$. Notice that, if $\phi_i(y^* - (k-1), y^*; g^{y^*-k+2,y^*}) \geq g_i^{y^*-(k-1)}$, then $\phi_i(y^* - k, y^*; g^{y^*-k+1,y^*}) = u_i(y^* - k) + \phi_i(y^* - (k-1), y^*; g^{y^*-k+2,y^*}) - g_i^{y^*-(k-1)} > 0$ for all i (where we adopt a convention that $\phi_i(y^*, y^*; g^{y^*+1,y^*}) \equiv u_i(y^*)$). Notice also that $\psi_N(y^* - k, y^*; g^{y^*-k+1,y^*}) = \theta_N(y^* - k) + \sum_{y=y^*-k+1}^{y^*}(\theta_N(y) - c(y)) > c(y^* - k)$ by Assumption 2. Then, for $\hat{y} = y^* - k$, we can find inductively a set $P^{\hat{y}}$ at (5) and assign a cost-share vector $(g_i^{\hat{y}})_{i \in P^{\hat{y}}}$ such that $\sum_{i \in P^{\hat{y}}} g_i^{\hat{y}} = c(\hat{y})$ and $0 < g_i^{\hat{y}} \leq \phi_i(\hat{y}, y^*; g^{\hat{y}+1,y^*})$ for all $i \in P^{\hat{y}}$, with $g_i^{\hat{y}} = 0$ for all $i \notin P^{\hat{y}}$.

Analogous to Step 2, given the strategy of the others, no agent in $P^{\hat{y}}$ can be strictly better off by choosing OUT or choosing IN and not cooperating in the cost-share decision, and no agent in $N \setminus P^{\hat{y}}$ strictly prefers IN to OUT. **(End of Proof of Step 3)**

By Assumption 2, an outcome of the game is efficient if the participants choose to continue up to Stage y^* and the game ends at Stage $y^* + 1$. The above argument shows that we can construct a strategy that is consistent with the agents' backward induction behavior where the game ends at Stage $y^* + 1$. *Q.E.D.*

Proof of Proposition 2. Let s^* be a strategy constructed at Steps 2 and 3 of Proposition 1. We have shown in the proof of Proposition 1 that, for all $y \leq \bar{y}$ and $i \in P^y(s^{*y})$, the set $P^y(s^{*y}) \setminus \{i\}$ chooses not to provide the y'th unit of the public good at every subgame-perfect equilibrium.

Therefore, $R_i^y(P^y(s^{*y}) \setminus \{i\}, \tilde{g}^y, b^{*y}) = 0$. On the other hand, $R_i^y(P^y(s^{*y}), g^y(s^{*y}), b^{*y}) \geq 0$, so that (7) is satisfied. Similarly, (8) has to be satisfied.

To show the second part of the proposition, let s^* be a strict subgame-perfect equilibrium at which an efficient allocation is attained. First, consider a subgame beginning from Stage $y^*.1$. Let $P^{y^*}(s^{*y^*})$ be a set of participants that is attained in Stage $y^*.1$ at s^*. Note that $\theta_{P^{y^*}(s^{*y^*})}(y^*) \geq c(y^*)$. Suppose, on the contrary, that $P^{y^*}(s^{*y^*})$ satisfies $\theta_{P^{y^*}(s^{*y^*})\setminus\{i\}}(y^*) \geq c(y^*)$ for some $i \in P^{y^*}(s^{*y^*})$. In this subgame, agent i obtains a payoff of $u_i(y^*) - g_i^{y^*}$ at Stage y^* at the equilibrium. On the other hand, if agent i deviates from IN to OUT at Stage y^*, then, applying Step 2 of Proposition 1, there is a strategy $\tilde{g}^{y^*}$ at which the y^*'th unit of the public good is provided, and agent i receives a payoff of $u_i(y^*)$. Since $g_i^{y^*} > 0$, agent i may receive the higher payoff if she deviates. This is a contradiction.

Let $y \in \{1, \ldots, y^* - 1\}$ and consider a subgame that starts from Stage $y.1$. Let $P^y(s^{*y})$ be a set of participants that is attained at s^* in Stage $y.1$. Since s^* supports an efficient allocation, $P^y(s^{*y})$ satisfies $\psi_{P^y(s^{*y})}(y, y^*; g^{y+1,y^*}) \geq c(y)$. Let us suppose that $\psi_{P^y(s^{*y})\setminus\{i\}}(y, y^*; g^{y+1,y^*}) \geq c(y)$ for some $i \in P^y(s^{*y})$. Note that agent i obtains the same payoff between Stage $(y+1).1$ and Stage $y^*.2$. Hence, it is sufficient to focus on the payoffs that is obtained at Stage y. Agent i receive a payoff $u_i(y) - g_i^y$ at s^* in Stage y, while, as in Stage y^*, she can obtain a payoff $u_i(y)$ at the maximum when i selects OUT. Therefore, if she deviates, then she may increase her payoff, which contradicts the idea that s^* is a strict subgame-perfect equilibrium. *Q.E.D.*

Proof of Step 1 of Proposition 3. Associated with a strategy s^*, let $P^y(s^{*y}) = Q^y$ for all $y \le y^*$ and $P^y(s^{*y}) = \emptyset$ for all $y > y^*$. As to Stage $y.2$ ($y \le y^*$), as in Steps 2 and 3 of Proposition 1, construct inductively cost-share vectors $(g_i^y)_{i \in Q^y}$ ($y \le y^*$) that satisfy $\sum_{i \in Q^y} g_i^y = c(y)$ and $0 < g_i^y \le \phi_i(y, y^*; g^{y+1,y^*})$ for all $i \in Q^y$ (where we adopt a convention that $\phi_i(y^*, y^*; g^{y^*+1,y^*}) \equiv u_i(y^*)$). Suppose also:

$$\begin{aligned} \gamma_i^y(P) &= g_i^y \text{ for all } y \le y^*,\ i \in Q^y \text{ and } P \subseteq N \text{ such that } i \in P \\ &= \epsilon_i^y < u_i(y) \text{ for all } y > y^*,\ i \in N \text{ and } P \subseteq N \text{ such that } i \in P. \end{aligned} \tag{12}$$

We first show that the strategy thus constructed constitutes a subgame-perfect equilibrium. From Proposition 1, this strategy is immune to any individual deviation at 'on-path' cases, i.e., at any Stage $y.2$ when the set of participants at that stage is Q^y. Consider now 'off-path' cases, i.e., Stage $y.2$ when the set of participants P is not equal to Q^y. Suppose that P satisfies $\psi_P(y, y^*; g^{y+1,y^*}) < c(y)$ for some $y \le y^*$ (where we adopt a convention that $\psi_P(y^*, y^*; g^{y^*+1,y^*}) \equiv \theta_P(y^*)$). Since $\psi_{P \cap Q^y}(y, y^*; g^{y+1}, y^*) - \sum_{i \in P \cap Q^y} \gamma_i^y(P) \ge 0$, then $\psi_{P \setminus Q^y}(y, y^*; g^{y+1}, y^*) - (c(y) - \sum_{i \in P \cap Q^y} \gamma_i^y(P)) < 0$. Therefore, for any s^y, if a y'th unit is produced, there arises $i \in P \setminus Q^y$ such that $\phi_i(y, y^*; g^{y+1,y^*}) - \kappa_i^y(s^y|P) < 0$. No such strategy constitutes a best response. Every agent $i \in P \setminus Q^y$ can avoid such a situation by choosing $\gamma_i^y(P)$ that satisfies $\phi_i(y, y^*; g^{y+1}, y^*) \ge \gamma_i^y(P)$. Accompanied with such $\gamma_i^y(P)$'s for $i \in P \setminus Q^y$ in addition to $\gamma_i^y(P)$, $i \in P \cap Q^y$ in (12), one can conclude $\sum_{i \in P} \gamma_i^y(P) \le \sum_{i \in P} \phi_i(y, y^*; g^{y+1}, y^*) < c(y)$, so that the game ends. This constitutes a subgame-perfect equilibrium. Suppose that P satisfies $\psi_P(y, y^*; g^{y,y^*}) \ge c(y)$. Suppose also that $\psi_{P\setminus Q^y}(y, y^*; g^{y+1}, y^*) < \sum_{i\in Q^y\setminus P} g_i^y$. Then $\psi_P(y, y^*; g^{y+1}, y^*) = \psi_{P\cap Q^y}(y, y^*; g^{y+1}, y^*) + \psi_{P\setminus Q^y}(y, y^*; g^{y+1}, y^*) < \psi_{P\cap Q^y}(y, y^*; g^{y+1}, y^*)+\sum_{i\in Q^y\setminus P} g_i^y$. Since $\sum_{i\in Q^y\setminus P} g_i^y \leq \psi_{Q^y\setminus P}(y, y^*; g^{y+1}, y^*)$, this implies that $\psi_P(y, y^*; g^{y+1}, y^*) < \psi_{Q^y}(y, y^*; g^{y+1}, y^*)$. Since $\psi_P(y, y^*; g^{y+1}, y^*) \geq c(y)$, one can find $P^y \subseteq P$ that satisfies (5). This contradicts the definition of Q^y. Therefore, if $\psi_P(y, y^*; g^{y,y^*}) \geq c(y)$, then $\psi_{P\setminus Q^y}(y, y^*; g^{y+1}, y^*) \geq \sum_{i\in Q^y\setminus P} g_i^y$ has to be the case. In this case, there exists a list of strategies $(\gamma_i^y(P))_{i\in P\setminus Q^y}$ such that $\sum_{i\in P\setminus Q^y} \gamma_i^y(P) = \sum_{i\in Q^y\setminus P} g_i^y$ and $\phi_i(y, y^*; g^{y+1}, y^*) \geq \gamma_i^y(P)$ for all $i \in P \setminus Q^y$. Such strategies, accompanied with $(\gamma_i^y(P))_{i\in P\cap Q^y}$ in (12), constitute a subgame-perfect equilibrium strategy where the y'th unit of the public good is produced. Let s^* be such a strategy.

We now show that s^* is robust against any deviation. Suppose T deviates from $s_T^* \equiv (s_T^{*y})_{y\leq\bar{y}}$ to $\tilde{s}_T \equiv (\tilde{s}_T^y)_{y\leq\bar{y}}$, and the game ends at Stage $\tilde{y}$ at the strategy profile $(s_{N\setminus T}^*, \tilde{s}_T)$. No coalitional deviation at Stage $\tilde{y} > y^* + 1$ is profitable (Step 1 of Proposition 1), so consider the case of $\tilde{y} \leq y^* + 1$. Notice that, for any deviation $\tilde{s}_T$, if the y'th unit of public good is produced, then $g_i^y(s_{N\setminus T}^{*y}, \tilde{s}_T^y) = g_i^y$ if $i \in Q^y \setminus T$ and $g_i^y(s_{N\setminus T}^{*y}, \tilde{s}_T^y) = 0$ if $i \in (N \setminus Q^y) \setminus T$. Therefore, $\sum_{j\in N\setminus T} g_j^y(s_{N\setminus T}^{*y}, \tilde{s}_T^y) = \sum_{j\in N\setminus T} g_j^y(s^{*y})$ for all $y \leq y^*$. Since $\sum_{i\in N} g_i^y(s^{*y}) = c(y)$ for all $y \leq y^*$, so, if T deviates at Stage y in order to produce the y'th unit, it has to bear a portion of the marginal cost by $c(y) - \sum_{i\in N\setminus T} g_i^y(s_{N\setminus T}^{*y}, \tilde{s}_T^y) = \sum_{i\in T} g_i^y(s^{*y})$. Then, at any Stage $y.1$, by deviation, T can at most reallocate the total benefit of $\sum_{i\in T}(u_i(y) - g_i^y(s^{*y}))$, which is, by construction, equal to that by choosing s_T^*. Moreover, if $\tilde{y} < y^* + 1$, T forgoes the benefit of $\sum_{\hat{y}=\tilde{y}}^{y^*} \sum_{i\in T} \left(u_i(\hat{y}) - g_i^{\hat{y}}(s^{*\hat{y}})\right) = \psi_T(\tilde{y}, y^*; g^{y+1,y^*}) - \sum_{i\in T} g_i^{\tilde{y}}$, which is non-negative by construction. In summary, $\sum_{i\in T} R_i^y(P^y(s^{*y}_{N\setminus T}, \tilde{s}^y_T), g^y(s^{*y}_{N\setminus T}, \tilde{s}^y_T), b^{*y}_{N\setminus T}, \tilde{b}^y_T) = \sum_{\hat{y}=y}^{\tilde{y}-1} \sum_{i\in T} \Big(u_i(\hat{y}) - g_i^{\hat{y}}(s^{*\hat{y}})\Big)$ $\leq \sum_{\hat{y}=y}^{y^*} \sum_{i\in T} \Big(u_i(\hat{y}) - g_i^{\hat{y}}(s^{*\hat{y}})\Big) = \sum_{i\in T} R_i^y(P^y(s^{*y}), g^y(s^{*y}), b^{*y})$.

On the other hand, consider a deviation at Stage $y.2$, $y \leq y^*$, where $P \subseteq N$ is a set of the participants. Suppose that $\psi_P(y, y^*; g^{y+1}, y^*) \geq c(y)$. At s^* constructed above, the y'th unit is produced. Suppose that T chooses a strategy $\tilde{s}_T$ to produce the y'th unit of the public good. In order to produce the y'th unit, T has to bear a portion of the marginal cost by $c(y) - \sum_{j\in P\setminus T} \kappa_j^y(s^{*y}_{N\setminus T}, \tilde{s}^y_T|P) = c(y) - \sum_{j\in P\setminus T} \kappa_j^y(s^*|P) = \sum_{i\in T} \kappa_i^y(s^*|P)$. Then, at any Stage $y.2$, by deviation, T can at most reallocate the total benefit of $\sum_{i\in T}(u_i(y) - \kappa_i^y(s^*|P))$. As to the payoffs following Stage $y+1$, the reasoning of the above paragraph applies. In summary, $\sum_{i\in T} R_i^y(P, \kappa^y(s^{*y}_{N\setminus T}, \tilde{s}^y_T|P), b^{*y}_{N\setminus T}, \tilde{b}^y_T)$ $\leq \sum_{i\in T}(u_i(y) - \kappa_i^y(s^*|P)) + \sum_{\hat{y}=y+1}^{y^*} \sum_{i\in T} \Big(u_i(\hat{y}) - g_i^{\hat{y}}(s^{*\hat{y}})\Big) = \sum_{i\in T} R_i^y(P, \kappa^y(s^{*y}|P), b^{*y})$. If T chooses $\tilde{s}_T$ to end the game at Stage $y.2$, then $\sum_{i\in T} R_i^y(P, \kappa^y(s^{*y}_{N\setminus T}, \tilde{s}^y_T|P), b^{*y}_{N\setminus T}, \tilde{b}^y_T) = 0$. Since s^* satisfies $\sum_{i\in T} R_i^y(P, \kappa^y(s^{*y}|P), b^{*y}) \geq 0$, the deviation is not profitable.

Suppose that $\psi_P(y, y^*; g^{y+1}, y^*) < c(y)$. At s^* constructed above, the y'th unit is not produced, so that $R_i^y(P, \kappa^y(s^{*y}|P), b^{*y}) = 0$ for all i. Consider a deviation by $T \subseteq P$. Similar to the above case, in order to produce the y'th unit, T has to bear a portion of the marginal cost by $c(y) - \sum_{j\in P\setminus T} \gamma_j^y(P)$. On the other hand, by construction, $\gamma_j^y(P) \leq \phi_j(y, y^*; g^{y+1}, y^*)$ for all $j \in P \setminus T$. As before, the total benefit following Stage $y+1$ is at most $\sum_{\hat{y}=y+1}^{y^*} \sum_{i\in T} \Big(u_i(\hat{y}) - g_i^{\hat{y}}(s^{*\hat{y}})\Big)$. In summary, the total benefit of T by deviation is at most $\sum_{i\in T} u_i(y) + \sum_{j\in P\setminus T} \gamma_j^y(P) - c(y) + \sum_{\hat{y}=y+1}^{y^*} \sum_{i\in T} \Big(u_i(\hat{y}) - g_i^{\hat{y}}(s^{*\hat{y}})\Big) \leq \psi_T(y, y^*; g^{y+1}, y^*) + \psi_{P\setminus T}(y, y^*; g^{y+1}, y^*) - c(y) < 0$.

Therefore, we have shown that no deviation is profitable. *Q.E.D.*

References

Aumann, R., 1959. Acceptable points in general cooperative n-person games. In: *Contributions to the theory of games IV*, by H. W. Kuhn and R. D. Luce, Eds., Princeton University Press: Princeton, 287-324.

Bernheim, D., Peleg, B., Whinston, M., 1987. Coalition-proof Nash equilibrium I: concepts. *Journal of Economic Theory* **42**, 1-12.

Cavaliere, A., 2001. Coordination and the provision of discrete public goods by correlated equilibria. *Journal of Public Economic Theory* **3**, 235-255.

Dixit, A., Olson, M., 2000. Does voluntary participation undermine the Coase theorem? *Journal of Public Economics* **76**, 309-335.

Groves, T., Ledyard, J., 1977. Optimal allocation of public goods: a solution to the free rider problem. *Econometrica* **45**, 783-809.

Muto, S., 1986. An information good market with symmetric externalities. *Econometrica* **54**, 295-312.

Okada, A., 1993. The possibility of cooperation in an n-person prisoners' dilemma with institutional arrangements. *Public Choice* **77**, 629-656.

Palfrey, T., Rosenthal, H., 1984. Participation and the provision of discrete public goods: a strategic analysis. *Journal of Public Economics* **24**, 171-193.

Ray, D., Vohra, R., 1997. Equilibrium binding agreements. *Journal of Economic Theory* **73**, 30-78.

Rubinstein, A., 1980. Strong perfect equilibrium in supergames. *International Journal of Game Theory* **9**, 1-12.

Saijo, T., Yamato, T., 1999. A voluntary participation game with a non-excludable public good. *Journal of Economic Theory* **84**, 227-242.

Shinohara, R., 2004. Voluntary participation in a mechanism implementing a public project. Discussion Paper Series #2004-14, Graduate School of Economics, Hitotsubashi University.

Shinohara, R., 2005. A participation game in a mechanism implementing a proportional cost-sharing rule. Presented at the Fall Meeting of Japanese Economic Association Conference (Okayama University, 2004).

Walker, M., 1981. A simple incentive compatible scheme for attaining Lindahl allocations. *Econometrica* **49**, 65-71.