Centre de Referència en Economia Analítica

Barcelona Economics Working Paper Series

Working Paper nº 148

Efficient and Stable Collective Choices under Crowding Preferences

Jordi Massó and Antonio Nicolò

November, 2004

Efficient and Stable Collective Choices under Crowding Preferences[*]

Jordi Massó[†] Antonio Nicolò[‡]

November 2004

Barcleona Economics WP nº 148

Abstract

We consider a set of agents who have to choose one alternative among a finite set of social alternatives. A final allocation is a pair given by the selected alternative and the group of its users. Agents have crowding preferences over allocations: between any pair of allocations with the same alternative, they prefer the allocation with the largest number of users. We require that a decision be efficient and stable (which guarantees free participation in the group of users and free exit from it). We propose a two-stage sequential mechanism

[*]We have received very useful comments from Pablo Amorós, Carmen Beviá, Sandro Brusco, Luis Corchón, Humberto Llavador, Bernardo Moreno, David Pérez-Castrillo, Socorro Puy, and Marc Vorsatz. The work of Massó is partially supported by the Spanish Ministry of Science and Technology through grant BEC2002-02130, and by Generalitat de Catalunya, through grant 2001SGR-00162 and the Barcelona Economics Program (CREA). Both authors acknowledge the financial support of the grant HI2001-0039 under the Acción Integrada (Azione Integrata) España-Italia from the Italian Ministry of University and Scientific Research (MIUR) and the Spanish Ministry of Science and Technology.

[†]Departament d'Economia i d'Història Econòmica and CODE, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain. E-mail: jordi.masso@uab.es.

[‡]Corresponding author. Dipartimento di Scienze Economiche, Università di Padova, via del Santo 33, 35123 Padova, Italy. Fax: (39)0498274211. E-mail: antonio.nicolo@unipd.it.

whose unique subgame perfect equilibrium outcome is an efficient and stable allocation which also satisfies a maximal participation property. The social choice function implemented by the proposed mechanism is also anonymous and group stable.

Keywords: Public Goods, Crowding Preferences, Subgame Perfect Implementation.

JEL Classification Numbers: D62, D71, H41.

1 Introduction

In many collective choice problems, after the social alternative (or public good) has been chosen, agents may decide whether or not to use it. If the size of the final set of users affects the welfare of each member, then the decision process has to take into account how many agents will eventually become users. In this paper we study the case when agents' preferences are positively affected by the size of the set of users, and participation is not compulsory. There are many examples of such problems. Members of a club choose the amount of some non-rival public good to be provided to themselves and the cost of its provision is usually equally shared among the set of its final users. This choice affects the composition (and the size) of the club, since some members may choose to leave the club if the level provided and its corresponding cost are unacceptable to them. Similarly, a local community which decides to provide a public facility (a swimming pool, a common garden, etc.) cannot set aside considerations regarding how many community members support this decision if those who are not in favor of it have the right not to pay for the facility. Many other problems do not directly involve money but can be similarly modeled. For instance members of a political party or a union decide which political line to follow and this decision affects their choice regarding their membership. The size of the organization matters for all of its members, since it determines how effective the organization is in pursing its objectives. A group of nations decides which common technological standard to adopt. Each country may prefer a different standard, but once a standard is adopted, social and individual welfare are increasing in the number of nations which agree to adopt it.

All these problems have two common features, other than the fact that agents care about how many other agents use the public good. The final allocation to be selected has to satisfy two properties: efficiency and stability. While the first requirement is well-known and typical in most of the public decision process, the latter deserves to be briefly mentioned. Stability requires that no agent can be forced to be a user and that no agent who wants to be a user could be excluded. Stability may be a necessary requirement due to institutional constraints (for instance, no nation can be forced to adopt any technological standard, or, according to the law, agents cannot be discriminated), but it is also a desirable property on the basis of normative principles like freedom (free participation) and equal treatment of equals (no discriminatory exclusion).[1]

The aim of this paper is to implement an efficient and stable social choice function when agents' crowding preferences are private information.

Our analysis starts by showing that, for any crowding preference profile, the set of efficient and stable allocations is non-empty. However, we can easily establish a negative result: no efficient and stable social choice function is Nash implementable (and therefore neither strategy-proof) because it is not Maskin monotonic. This result is related to previous results in Jackson and Nicolò (2004) who study similar social choice problems in a context where agents have single-peaked preferences over an infinite and linearly ordered set of alternatives. They show that, in general, strategy-proof and efficient social choice functions must fix the group of users and not allow it to vary with agents' preferences. Namely, when crowding effects are present strategy-proofness and efficiency impose that the group of users coincide with the entire society. Therefore, stability is incompatible with strategy-proofness and efficiency. But this

[1] See also Bogomolnaia and Nicolò (2004) for a brief discussion of the normative content of a slightly different definition of stability in the context of multiple provision of public goods.

result suggests that the trade off between informational constraints and normative properties of social choice functions could be overcome if we separate the decision of which alternative has to be chosen from the selection of the group of its users. We therefore investigate if an efficient and stable social choice function is subgame perfect Nash implementable. We first show that one of the sufficient conditions of subgame perfect Nash implementation in Moore and Repullo (1988) and Abreu and Sen (1990) does not hold in our framework. In particular, any efficient and stable social choice function does not satisfy the no veto power condition (that together with Condition α and that the number of agents is larger or equal than three guarantees that a social choice function is subgame perfect Nash implementable). This is because stability gives to any agent the power (by not being a final user) to veto an allocation which is unanimously considered by all remaining agents as being the best one. We then present the implementation result which also holds for the case of two agents. The proposed two-stage game depends on an exogenously given order on the set of agents and on a selection rule choosing an alternative from every subset of alternatives. Roughly, it is as follows. In the first stage of the game agents sequentially (iteratively and publicly), following the given order, propose a level of the public good and a natural number between 1 and the number of agents (interpreted as the number of users); among the proposed levels, one with the maximal number of users is chosen in accordance with the selection rule. In the second stage agents sequentially (and publicly), following the same given order, decide whether or not to use the level of the public good chosen at the first stage.

The game is relatively simple: it is finite, bounded, and the needed out-of-equilibrium penalties do not have to be large. Interestingly, the unique subgame perfect Nash equilibrium outcome of the game does not depend on the order according to which agents make their decisions; hence, the implemented social choice function is anonymous. The mechanism selects among the set of efficient and stable allocations the alternative which maximizes the number of its users (if there are many, it selects the one chosen by a given selection rule). We justify this maximality property on a purely normative ground, since it allows to minimize the number of agents with the minimum level of welfare.

Finally, our paper is also related to Bag and Winter (1999), in which the authors propose a sequential iterated mechanism to uniquely implement a core allocation for an economy with an excludable public good. In their model a level of a public good is produced using a technology and the contributions of a private good made by the final set of users. However, our setting is different from theirs at least with respect to the following features. First, in our setting exclusion is voluntary (our stability notion reflects that). Second, their setting is cardinal (preferences are quasi-linear in the private good) while our ordinal setting not only admits a larger class of preferences but also admits problems in which the choice of a social alternative does not generate costs. Third, in their setting efficiency implies no exclusion, and thus, in the equilibrium outcome of their game all agents consume the public good; in contrast, in our setting efficiency may require that only a subset of agents is the final set of users of the public good.

The paper proceeds as follows. In section 2, we give preliminary notation and definitions, describe the preference domain, establish the existence of efficient and stable allocations, and provide a negative result for Nash implementation. In section 3, we describe the extensive-form game and state our main result. In section 4, we offer some examples that illustrate the role of some features of the extensive-game form, discuss on the non-neutrality of our mechanism, and give the relationship between the set of efficient and stable allocations and the set of group stable allocations. An Appendix at the end of the paper contains the proofs omitted in the text.

2 Preliminaries

Let $N = \{1, ..., n\}$ be the set of *agents* and X be the finite set of levels of a public good (or social alternatives). We assume that $n, \#X \geq 2$. Subsets of N are denoted by S and T, elements of N by i and j, and elements of X by x and y. An *allocation* is a pair $(x, S) \in A \equiv X \times 2^N$, where $x \in X$ is the level of the public good and $S \in 2^N$ is the subset of its users. Agents have *preferences* over the set of allocations. The *preference relation* of agent $i \in N$ over the set of allocations A, denoted by R_i, is a complete, reflexive and transitive binary relation. As usual, let P_i and I_i denote the strict and indifference preference relations induced by R_i, respectively. We assume that preference relations satisfy the following properties:

(An) Anonymity: For all $x \in X$ and $S, T \in 2^N$ such that $i \in S \cap T$ and $\#S = \#T$, $(x, S)\, I_i\, (x, T)$.

(Crow) Crowding: For all $x \in X$ and $S, T \in 2^N$ such that $i \in S \cap T$ and $\#S > \#T$, $(x, S)\, P_i\, (x, T)$.

(Apa) Apathy: For all $x, y \in X$ and $S \in 2^N$ such that $i \notin S$, $(x, S)\, I_i\, (y, \emptyset)$.

(Strict) Strictness: For all $x, y \in X$ and $S, T \in 2^N$ such that $i \in S$ if (1) $x \neq y$ or (2) $x = y$ and $\#S \neq \#T$ hold, then either $(x, S)\, P_i\, (y, T)$ or $(y, T)\, P_i\, (x, S)$.

Anonymity requires that agent i only cares about the number of users but not on their identities. Crowding implies that agent i strictly prefers to use the public good with larger groups. Apathy says that agent i does not care about the level of the public good if he does not use it.[2] Finally, Strictness requires that agent i is never indifferent between two different allocations with the properties that i is a user of at least one of them and the two allocations differ either on the level of the public good and/or on the size of its users.

A preference relation R_i satisfying these four properties is called a *crowding preference relation* and $\mathcal{R}_i$ denotes the set of all such preference relations for agent i. Notice

[2] Note that when the public good to be chosen has some type of externality, even those members who are not direct users may have strict preferences over which alternative has to be selected. In these cases (Apa) turns to be a too restrictive assumption. Nevertheless in many interesting contexts, like the provision of club goods, it seems a natural assumption.

that all four conditions are agent specific and therefore $\mathcal{R}_i \neq \mathcal{R}_j$ for different agents i and j. Observe that the set of crowding preferences for agent i admits preferences with very different trade-offs between the selected level of the public good and the size of its users; for instance, a crowding preference R_i might well order $(x, \{i\}) P_i (y, N)$.

A *profile* $R = (R_1, ..., R_n)$ is a n-tuple of crowding preference relations. Let $\mathcal{R} = \mathcal{R}_1 \times ... \times \mathcal{R}_n$ be the set of profiles. To emphasize the role of agent i's preference relation a profile R is represented by (R_i, R_{-i}).

We say that an allocation (y, T) *Pareto dominates* the allocation (x, S), denoted by $(y, T) PD(x, S)$, if $(y, T)\, R_i\, (x, S)$ for all $i \in N$ and $(y, T)\, P_j\, (x, S)$ for at least one $j \in N$.

Definition 1 An allocation (x, S) is *efficient under* R if it is not Pareto dominated by any other allocation.

Definition 2 An allocation (x, S) is *stable under* R if for all $i \in N$:
(INTERNAL STABILITY) $i \in S$ implies $(x, S)\, P_i\, (x, S \backslash \{i\})$.
(EXTERNAL STABILITY) $i \notin S$ implies $(x, S)\, P_i\, (x, S \cup \{i\})$.

Observe that (APA) implies that if (x, S) is internally stable then, $i \in S$ implies $(x, S) P_i (x, \{\emptyset\})$. Given a profile $R \in \mathcal{R}$, let $Z(R)$ denote the set of efficient and stable allocations under R. Proposition 1 below establishes the fact that for all $R \in \mathcal{R}$ the set of efficient and stable allocations under R is non-empty. But first, we show two preliminary results concerning efficient and stable allocations. Lemma 1 says that for each level of the public good x we can find a (maximal) set of users S_x for which the allocation (x, S_x) is stable.

Lemma 1 *Let $R \in \mathcal{R}$ be given. For each $x \in X$ there exists a unique $S_x \in 2^N$ such that $(x, S_x) \in A$ is stable under R and for any $T \in 2^N$ such that $(x, T) \in A$ is stable under R, $\#S_x \geq \#T$.*

Proof Let $R \in \mathcal{R}$ and $x \in X$ be given. For $0 \leq k \leq n$ define the set $N^k(x) = \left\{ i \in N \mid \text{there exists } S \in 2^N \text{ such that } (x, S) P_i (x, \emptyset) \text{ and } \#S = k \right\}$. Observe first that

$\#N^0(x) = 0$ and, by (CROW), $N^0(x) \subset N^1(x) \subset ... \subset N^n(x)$. Take the maximal k such that $\#N^k(x) = k$ and set $S_x \equiv N^k(x)$. By construction, $(x, S_x) \in A$ is stable under R and if $(x, T) \in A$ is stable under R then $\#S_x \geq \#T$. ■

We call the stable allocation (x, S_x) identified in Lemma 1 the *stable and efficient allocation relative to* x, and refer to S_x as the *maximal stable set of users of* x. In fact such allocation can be Pareto dominated only by another (stable) allocation (y, T) with $y \neq x$.

Lemma 2 *Let $R \in \mathcal{R}$ be given. If $(x, S) \in A$ is stable but not efficient under R, then there exists another $(y, T) \in A$ stable under R such that (y, T) Pareto dominates (x, S).*

Proof Let $R \in \mathcal{R}$ be given. Assume that $(x, S) \in A$ is stable but not efficient under R. Then, there exists $(y, S') \in A$ such that $(y, S')\, R_i (x, S)$ for all $i \in N$ and $(y, S')\, P_j (x, S)$ for some $j \in N$. Since (x, S) is stable under R, by (APA), $(x, S)\, P_i (x, \emptyset)$ for all $i \in S$. Since $(y, S')\, PD\, (x, S)$ then $S' \supseteq S$ and, by (APA), $(y, S')\, P_j (y, \emptyset)$ for all $j \in S$. Suppose that there exists $i \in S'$ such that $(y, \emptyset) P_i\, (y, S')$, then $i \notin S$, but then $(x, S)\, P_i (y, S')$ which is a contradiction. Finally, let $T \in 2^N$ be such that there does not exist any $i \notin T$ for whom $(y, T \cup \{i\})\, P_i\, (y, \emptyset)$. By (CROW), $(y, T)\, R_i\, (y, S')$ for all $i \in N$ and (y, T) is stable under R. By transitivity of R_i, $(y, T) R_i (x, S)$ for all $i \in N$ and $(y, T) P_j (x, S)$ for some $j \in N$. Hence, $(y, T) PD (x, S)$. ■

Lemmata 1 and 2 have two important consequences. First, to know whether or not a stable allocation is efficient it is enough to check that is not Pareto dominated by any other *stable* allocation. Second, given that the set of stable allocations is not empty, the set of stable and efficient allocations is non-empty. We state this second consequence as Proposition 1 below.

Proposition 1 *For all $R \in \mathcal{R}$, $Z(R) \neq \emptyset$.*

Proof Let $R \in \mathcal{R}$ be given. Consider any stable allocation (x, S_x) under R, whose existence is established by Lemma 1. If (x, S_x) is efficient under R, Proposition 1 follows; otherwise, by Lemma 2, there exists a stable allocation (z, S_z) under R which Pareto dominates (x, S_x). Since X is finite and the Pareto dominance relation is transitive, there must exist a stable and efficient allocation (y, S_y) under R. ■

Among the set of efficient and stable allocations we will be specially interested on those that have the largest set of users. Given $R \in \mathcal{R}$, define

$$MP(R) = \{(x, S) \in Z(R) \mid \#S \geq \#T \text{ for all } (y, T) \in Z(R)\}.$$

Observe that since $Z(R)$ is non-empty and finite, $MP(R) \neq \emptyset$ for all $R \in \mathcal{R}$. We will refer to the set $MP(R)$ as the *maximal participation set.* In our setting the minimum level of welfare that any agent i can get is the level that i obtains in any allocation (x, S) where $i \notin S$. In fact, stability guarantees that each agent can always refuse to use the public good and, by (Apa), all allocations where agent i is not a user are indifferent for him. Maximality hence guarantees that the final allocation minimizes the number of agents with the minimum level of welfare. Therefore, it is a normative property inspired by a rawlsian maxmin principle.

A *social choice function* is a mapping $\varphi : \mathcal{R} \to X \times 2^N$ selecting an allocation for each preference profile. A social choice function is efficient and stable if, for each $R \in \mathcal{R}$, the allocation $\varphi(R)$ is efficient and stable under R.

Information about individual preferences is often not available to the decision-maker. In addition, the institution under which the social decision has to be taken may give to each agent the right to claim as one's own any crowding preference (even if it is known that this is not the case). Therefore, if we want the choice of the allocation to be dependent on the preference profile (in the appropriate way to insure efficiency and stability), we have to design a mechanism to implement an efficient and stable social choice function. But it is easy to prove that no efficient and stable social choice function is Nash implementable in the set of profiles of crowding preference relations. Before stating this result we need some additional notation and definitions.

A *mechanism* (or game form) is a pair (M, Φ) where $M = M_1 \times ... \times M_n$ is a Cartesian product of message spaces (one for each agent) and $\Phi : M \to A$ is an outcome function. Thus, each player i submits a message $m_i \in M_i$ and, given $(m_1, ..., m_n) \in M$, the allocation $\Phi(m_1, ..., m_n)$ is selected. A social choice function $\varphi : \mathcal{R} \to A$ is *Nash implementable* if there exists a mechanism (M, Φ) such that for all $R \in \mathcal{R}$, $\varphi(R) = \Phi(m_1^*, ..., m_n^*)$ for all Nash equilibria $(m_1^*, ..., m_n^*) \in M$ of the induced normal form game $(N, (M, \Phi), R)$. A social choice function $\varphi : \mathcal{R} \to A$ is *Maskin monotonic* if for any $R \in \mathcal{R}$, $R' \in \mathcal{R}$, and $a = \varphi(R)$ such that $a \neq \varphi(R')$ there exist $i \in N$ and $b \in A$ such that aR_ib and $bP_i'a$. Maskin monotonicity is a necessary condition for a social choice function to be Nash implementable.[3]

Proposition 2 *No efficient and stable social choice function $\varphi : \mathcal{R} \to X \times 2^N$ is Nash implementable.*

Proof Let $\varphi : \mathcal{R} \to X \times 2^N$ be an efficient and stable social choice function. Take $x, y \in X$ arbitrary and select any profile $R \in \mathcal{R}$ of crowding preference relations with the following properties: (1) for all $i \in N$ and $S \in 2^N$ such that $S \neq N$, $(z, N)P_i(z', S)$ for all $z, z' \in X$; (2) $(x, N)P_1(z, N)$ for all $z \neq x$; and (3) for all $i \neq 1$, $(y, N)P_i(z, N)$ for all $z \neq y$. By efficiency and stability, $\varphi(R) = (\hat{z}, N)$ for some $\hat{z} \in X$. Without loss of generality, assume that $\hat{z} \neq x$. Consider now the crowding preference relation $R_1' \in \mathcal{R}_1$ with the following properties: (1) for all $S, S' \in 2^N$ such that $1 \in S \cap S'$, $(y, S)P_1'(y', S')$ if and only if $(y, S)P_1(y', S')$ and (2) $(x, N)P_1'(x, \emptyset)P_1'(z, N)$ for all $z \neq x$. By stability, if $\varphi(R_1', R_{-1}) = (z, S)$ with $z \neq x$ then $1 \notin S$. Therefore, by efficiency, $\varphi(R_1', R_{-1}) = (x, N)$. Hence, $\varphi(R) = (\hat{z}, N)$, $\varphi(R') = (x, N) \neq (\hat{z}, N)$, and for all $j \neq 1$, $R_j = R_j'$. Thus, Maskin monotonicity is violated since (x, N) is the best alternative for agent 1 according to R_1 and R_1'. Thus, the efficient and stable social choice function φ is not Nash implementable. ■

Remark 1 Jackson and Nicolò (2004) showed that, in the continuous version of our model, there are no strategy-proof, efficient, internally stable, and outsider independent social choice functions on the domain of crowding and single-peaked preference

[3] See, for instance, Maskin (1999)'s original paper or Jackson (2001)'s survey on implementation theory.

relations.[4] Since negative implementation results on smaller domains are stronger, observe that the preference profile $R \in \mathcal{R}$ and the preference relation $R_1' \in \mathcal{R}_1$ used in the proof of Proposition 2 might be single-peaked. Hence, the proof of Proposition 2 shows that any efficient and stable social choice function defined on the domain of crowding *and* single-peaked preference relations is not fully Nash implementable.

3 The Implementation

Given the impossibility to implement any efficient and stable social choice function as Nash equilibria of a game in normal form, we now address the natural question whether it is possible to implement some of them as Subgame Perfect Nash Equilibria (SPNE) of a game in extensive form. However, we will not be able to apply directly general results of the implementation theory because efficient and stable social choice functions do not satisfy one of the sufficient conditions for SPNE implementation in both Moore and Repullo (1988) and Abreu and Sen (1990). In our setting a social choice function $\varphi : \mathcal{R} \to X \times 2^N$ satisfies the *no veto power* condition if, whenever some allocation $(x, S) \in X \times 2^N$ is top-ranked for at least $n-1$ agents at profile $R \in \mathcal{R}$ then $\varphi(R) = (x, S)$. Example 1 below shows that the no veto power condition is incompatible with internal stability. Free participation, in fact, must be guaranteed even if all the other agents have a common preferred allocation, which might require that the set of users be the full set of agents.

Example 1 Let $X = \{x, y\}$. Consider any $N = \{1, ..., n\}$ and let $R \in \mathcal{R}$ be any

[4] A social choice function $\varphi : \mathcal{R} \to X \times 2^N$ is *outsider independent* if for all $i \in N$, $R \in \mathcal{R}$ and $R_i' \in \mathcal{R}_i$, if $i \notin S \cup S'$ where $(x, S) = \varphi(R)$ and $(x', S') = \varphi(R_i', R_{-i})$, then $\varphi(R) = \varphi(R_i', R_{-i})$.

Assume X is endowed with a linear order $\leq$. A preference relation $R_i \in \mathcal{R}_i$ is *single-peaked* if there exists $p(R_i) \in X$ such that for all $x, y \in X$

$$y < x \leq p(R_i) \text{ or } p(R_i) \leq x < y \text{ implies } (x, S)\, P_i\, (y, S),$$

for all $S \in 2^N$ such that $i \in S$.

crowding preference profile such that for all $i \neq 1$, $(x, N)\, P_i\, (z, S)$ for all $(z, S) \neq (x, N)$, and $(x, \emptyset)\, P_1(x, N)$. Let $\varphi : \mathcal{R} \to X \times 2^N$ be a stable social choice function. The no veto power condition requires that $\varphi(R) = (x, N)$. But, since the allocation (x, N) is not stable under R, the stability of φ implies that $\varphi(R) \neq (x, N)$.

The structure of the problem (the social choice has two components: the level of the public good and the set of its users) as well as previous results in similar frameworks (see Bogomolnaia and Nicolò (2004) and Jackson and Nicolò (2004)) suggest that in order to achieve efficiency and stability the selection of the alternative to be chosen and the group of its users must be separated. Therefore a two-stage mechanism seems to be a natural way to implement an efficient and stable social choice function. But before proceeding any further, there is another aspect that deserves to be briefly mentioned. Mechanisms constructed to prove general SPNE implementation results are unbounded and infinite. They contain, for instance, integer subgames (without Nash equilibria) or large out-of-equilibrium penalties. In contrast, our proposed mechanism has the following simple features: each player has a finite set of choices and strategies, out-of-equilibrium penalties may be (infinitely) small, and all subgames have Nash equilibria.

Since the maximal participation set $MP(R)$ might have several allocations, to define our two-stage game that implements in SPNE a social choice function selecting, for each preference profile $R \in \mathcal{R}$, an allocation in the set $MP(R)$, we need a selection rule on the subsets of X. Let $H : 2^X \to X$ be any selection rule (i.e., $H(\mathcal{X}) \in \mathcal{X}$ for all $\mathcal{X} \in 2^X \backslash \{\emptyset\}$ and $H(\emptyset) \in X$) with the following *independence of irrelevant alternatives* property: If $x \in \mathcal{X} \subsetneq \mathcal{Y}$ and $H(\mathcal{X}) \neq x$ then $H(\mathcal{Y}) \neq x$. For instance, if the set of alternatives X has a linear order, the selection rule could choose from each set $\mathcal{X} \subseteq X$ its smallest alternative. Now, given H, define the social choice function $\varphi_H : \mathcal{R} \to X \times 2^N$ as follows: for each $R \in \mathcal{R}$, let $\varphi_H(R) = (x, S)$, where $(x, S) \in MP(R)$ and $x = H(\{y \in X \mid \text{there exists } T \in 2^N \text{ such that } (y, T) \in MP(R)\})$.

Let $\sigma : \{1, ..., n\} \to N$ be a one-to-one mapping representing an exogenously given order of agents; namely, $\sigma(t) = i$ means that agent i is in the t^{th} position according to the ordering σ. Let Σ be the set of all $n!$ possible orderings and denote by $Pre(i,\sigma)$ the set of predecessors of agent i according to σ. Namely,

$$Pre(i,\sigma) = \left\{ j \in N \mid \sigma^{-1}(j) < \sigma^{-1}(i) \right\}.$$

To iterate a given order $\sigma : \{1,...,n\} \to N$, extend σ to $\hat{\sigma} : \mathbb{N} \to N$ as follows: given $n \in \mathbb{N}$, the number of agents, each integer $m \in \mathbb{N}$ can uniquely be written as $m = tn + r$ for some $t \in \mathbb{N} \cup \{0\}$ and $1 \leq r \leq n$. Define this number r as $r \equiv m[\mathrm{mod}\, n]$. Then, set $\hat{\sigma}(m) = \sigma(m[\mathrm{mod}\, n])$.

3.1 The Extensive-Game Form $\Gamma^{\sigma,H}$

Let $\sigma : \{1,...,n\} \to N$ and $H : 2^X \to X$ be given.

- Stage 1:

  - Step 1: agent $i = \sigma(1)$ proposes either $p_i = (x_i, k_i) \in X \times \{1,...,n\} \equiv \overline{A}$ or does not propose anything (identified as the proposal $p_i = (NP, 0)$).

    Assume that m proposals $p_{\hat{\sigma}(1)}, ..., p_{\hat{\sigma}(m)}$ have already been made. Define $A_m = \{p_{\hat{\sigma}(q)} \mid p_{\hat{\sigma}(q)} = (x_{\hat{\sigma}(q)}, k_{\hat{\sigma}(q)}) \in \overline{A}$ for some $1 \leq q \leq m\}$ and let $\bar{k}_m$ be the maximum among $\{k_{\hat{\sigma}(1)}, ..., k_{\hat{\sigma}(m)}\}$ (set $\bar{k}_m = 0$ if $A_m = \emptyset$).

  - Step $m+1$: agent $i = \hat{\sigma}(m+1)$ proposes either $p_i = (x_i, k_i) \in \overline{A} \backslash A_m$ such that $k_i \geq \bar{k}_m$ or does not propose anything ($p_i = (NP, 0)$).

    If after the first n steps all agents proposed $(NP, 0)$ then the game ends with the outcome $(H(\emptyset), \emptyset)$. Otherwise, let $m > 1$ be the first step such that $p_{\hat{\sigma}(m)} = (x_{\hat{\sigma}(m)}, k_{\hat{\sigma}(m)}) \in \overline{A}$ and $p_{\hat{\sigma}(m+1)} = ... = p_{\hat{\sigma}(m+n)} = (NP, 0)$. Given $p_{\hat{\sigma}(1)}, ..., p_{\hat{\sigma}(m)}$, define

$$\hat{x} = H(\{x \in X \mid \exists 1 \leq q \leq m \text{ s.t. } p_{\hat{\sigma}(q)} = (x, k) \text{ and } k = k_{\hat{\sigma}(m)}\}).$$

Set $\hat{k} = k_{\hat{\sigma}(m)}$ and $\hat{\imath} = \hat{\sigma}(q)$ where q is such that $p_{\hat{\sigma}(q)} = (\hat{x}, \hat{k})$. Then, the outcome of Stage 1 is $(\hat{x}, \hat{k}, \hat{\imath}) \in \overline{A} \times N$; namely, a proposal $(\hat{x}, \hat{k})$ and the agent $\hat{\imath}$ who made it.

Each proposer has to burden an ε-cost if none of her proposals is the selected one at Stage 1, $(\hat{x}, \hat{k})$.[5]

- Stage 2: Each agent j, knowing the outcome $(\hat{x}, \hat{k}, \hat{\imath}) \in \overline{A} \times N$ of Stage 1 and the decision of j's predecessors, announces sequentially (following the order σ) whether he wants to use (denoted by u) or not to use (denoted by nu) the public good at level $\hat{x}$.

  The final set of users of $\hat{x}$ is the set of agents who have announced to be willing to be a user, only if this set contains at least $\hat{k}$ agents; otherwise, no agent uses $\hat{x}$. Agent $\hat{\imath}$, who made the proposal $(\hat{x}, \hat{k})$ in Stage 1, has to burden an additional ε-cost if he is not a user of $\hat{x}$; i.e., either $(\hat{x}, \emptyset)$ is selected, and/or i announced nu.

3.2 Strategies

A *consumption strategy* of agent i in Stage 1 is a choice of a feasible proposal at each of i's information sets. We assume that agents only use stationary consumption strategies in the sense that, among the set of pairs previously proposed (if any), their decisions only depend upon those proposals with a maximum number of users.[6] Thus,

[5] We do not put any restrictions on these ε-costs. In particular, and to be consistent with our ordinal setting, they can be non-transferable. But, if we embed the ordinal setting into a cardinal one, these ε-costs can be interpreted as monetary fines (potentially, infinitely small). These ε-costs are only used in the proof of our main result to take away from agents the incentives (which exist due to indifferences) of making a proposal that has no effect to themselves (because, independently of whether or not this proposal is made, the proposer will not use the finally chosen alternative), yet the proposal has influence on the outcome of Stage 1.

[6] The other proposals with an smaller number of users have already been excluded as possible outcomes of Stage 1, regardless of the given selection rule. Observe that we could restrict a bit further

the sets of choices of agent i in Stage 1, denoted by $F_i(\cdot)$, are the following. At the information set $(NP, 0)$ where no predecessor has made a proposal yet, $F_i(NP, 0) = \overline{A} \cup \{(NP, 0)\}$. To denote the information sets in which at least one predecessor has already made a proposal, define $\mathcal{C}_k = \{(x', k) \in \overline{A} \mid x' \in X\}$; then at the information set $C_k \subseteq \mathcal{C}_k$,

$$F_i(C_k) = \left\{(x', k') \in \overline{A} \backslash C_k \mid k' \geq k\right\} \cup \{(NP, 0)\}.$$

A consumption strategy for agent i in Stage 1 is a rule $f_i(\cdot)$ that selects, for each i's information set, a feasible proposal: $f_i(NP, 0) \in F_i(NP, 0)$ and for all $1 \leq k \leq n$ and all $C_k \subseteq \mathcal{C}_k$, $f_i(C_k) \in F_i(C_k)$. Let F_i be the set of consumption strategies of agent i in Stage 1 and let f_i be a generic element of this set.

Assume that the outcome of Stage 1 is $(\hat{x}, \hat{k}, \hat{\imath}) \in \overline{A} \times N$.[7] In Stage 2, and after knowing $(\hat{x}, \hat{k}, \hat{\imath})$, agents decide sequentially whether or not they would like to use the public good at level $\hat{x}$ with at least $\hat{k}$ users. Given $\sigma \in \Sigma$, the set of *participation strategies* of agent i at the subgame starting at $(\hat{x}, \hat{k}, \hat{\imath}) \in \overline{A} \times N$, $\Gamma^{\sigma}(\hat{x}, \hat{k}, \hat{\imath})$, is the set of functions

$$B_i^{\sigma}(\hat{x}, \hat{k}, \hat{\imath}) = \left\{b_i[\hat{x}, \hat{k}, \hat{\imath})] : 2^{Pre(i,\sigma)} \to \{u, nu\}\right\},$$

where $b_i^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]\,(S)$ specifies whether or not agent i is willing to use the public good at level $\hat{x}$, given that the set of agents in $S \in 2^{Pre(i,\sigma)}$ have already announced that they are willing to do so, $\hat{\imath}$ made the proposal $(\hat{x}, \hat{k})$, and $\hat{k}$ users are necessary. Let $B_i^{\sigma} = \bigcup_{(\hat{x},\hat{k},\hat{\imath}) \in \overline{A} \times N} B_i^{\sigma}(\hat{x}, \hat{k}, \hat{\imath})$ denote the set of participation strategies of agent i in Stage 2 and let b_i^{σ} be a generic element of this set. Given $\sigma \in \Sigma$, let $G_i^{\sigma} = F_i \times B_i^{\sigma}$ denote the set of strategies of agent i. A strategy profile $g^{\sigma} = (f, b^{\sigma}) \in F \times B^{\sigma}$ is an n-tuple of strategies, where $F = F_1 \times ... \times F_n$ and $B^{\sigma} = B_1^{\sigma} \times ... \times B_n^{\sigma}$. Let $G^{\sigma} = G_1^{\sigma} \times ... \times G_n^{\sigma}$ be the set of strategy profiles.

the stationarity of the strategies by applying the selection rule to the set of proposed alternatives with a maximum number of users, but then they would depend on the specific selection rule.

[7] Note that if after the first n steps all agents proposed $(NP, 0)$ the game does not move to Stage 2 and ends with the outcome $(H(\emptyset), \emptyset)$.

3.3 Outcome Functions

Given an order $\sigma \in \Sigma$ and a consumption strategy profile $f \in F$, let $p_{\hat{\sigma}(m)}(f)$ be the proposal made by agent $\hat{\sigma}(m)$ according to f at Step m of Stage 1. Denote the path generated by f by $path^{\sigma}(f) = \{p_{\hat{\sigma}(1)}(f), ..., p_{\hat{\sigma}(M)}(f)\}$, where M is the last step of Stage 1. Given a selection rule $H : 2^X \to X$, the *outcome* of Stage 1 generated by f is

$$o_1^{\sigma,H}(f) = \begin{cases} (H(\emptyset), \emptyset) & \text{if } path^{\sigma}(f) = \{\underbrace{(NP, 0), ..., (NP, 0)}_{n-times}\} \\ (\hat{x}, \hat{k}, \hat{\imath}) & \text{otherwise,} \end{cases}$$

where $(\hat{x}, \hat{k}, \hat{\imath})$ is defined in the obvious (but tedious) way. Given a consumption strategy profile $f \in F$, we define the indicator function of agent $i \in N$, $\varepsilon_1^{\sigma,H}(f)_i$, where 1 means that agent i has made some proposal and none of them has been selected. Namely,

$$\varepsilon_1^{\sigma,H}(f)_i = \begin{cases} 1 & \text{if } \exists\ 1 \leq m \leq M \text{ s.t. } \hat{\sigma}(m) = i,\ p_{\hat{\sigma}(m)}(f) \neq (NP, 0), \text{ and } i \neq \hat{\imath} \\ 0 & \text{otherwise.} \end{cases}$$

Given an order $\sigma \in \Sigma$, an outcome of Stage 1 in which at least a proposal has been made $(\hat{x}, \hat{k}, \hat{\imath}) \in \overline{A} \times N$, and a participation strategy profile $b^{\sigma} \in B^{\sigma}$ define recursively (in the obvious and tedious way) the indicator function of the decision of agent j along the play of the subgame starting at $(\hat{x}, \hat{k}, \hat{\imath})$ generated by the profile b^{σ} as

$$\pi_j(b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]) = \begin{cases} 1 & \text{if agent } j \text{ announced } u \\ 0 & \text{if agent } j \text{ announced } nu. \end{cases}$$

Let $S(b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]) \equiv \{j \in N \mid \pi_j(b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]) = 1\}$ be the set of agents that announced their willingness to be a user along the play generated by the participation strategy profile $b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]$. Then, the outcome of Stage 2 starting at $(\hat{x}, \hat{k}, \hat{\imath})$ generated by $b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]$ is

$$o_2^{\sigma}(b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]) = \begin{cases} (\hat{x}, S(b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}])) & \text{if } \#S(b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]) \geq \hat{k} \\ (\hat{x}, \emptyset) & \text{otherwise;} \end{cases}$$

that is, the set of final users is the set of agents who announced u, $S(b^\sigma[\hat{x},\hat{k},\hat{\imath}])$, as long as its cardinality is larger or equal than $\hat{k}$; otherwise, no agent becomes a user. Moreover, let $\varepsilon_2^\sigma(b^\sigma[\hat{x},\hat{k},\hat{\imath}])$ indicate whether or not agent $\hat{\imath}$ (who proposed $(\hat{x},\hat{k})$) is a final user of the public good; namely,

$$\varepsilon_2^\sigma(b^\sigma[\hat{x},\hat{k},\hat{\imath}]) = \begin{cases} 1 & \text{if either } o_2^\sigma(b^\sigma[\hat{x},\hat{k},\hat{\imath}]) = (\hat{x},\emptyset) \text{ or } \pi_{\hat{\imath}}(b^\sigma[\hat{x},\hat{k},\hat{\imath}]) = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Finally, define the outcome function $o^{\sigma,H} : G^\sigma \to A$ of the overall extensive-game form $\Gamma^{\sigma,H}$ as follows. For each $(f,b^\sigma) = ((f_1,b_1^\sigma),...,(f_n,b_n^\sigma)) \in G^\sigma$,

$$o^{\sigma,H}(f,b^\sigma) = \begin{cases} o_2^\sigma(b^\sigma[o_1^{\sigma,H}(f)]) & \text{if } o_1^{\sigma,H}(f) \neq (H(\emptyset),\emptyset) \\ (H(\emptyset),\emptyset) & \text{otherwise.} \end{cases}$$

Additionally, to keep track of who has to burden the ε-cost, given an strategy profile $(f,b^\sigma) \in G^\sigma$, define: for each $i \in N$,

$$\varepsilon_i(f,b^\sigma) = \begin{cases} \varepsilon_2^\sigma(b^\sigma[o_1^{\sigma,H}(f)]) & \text{if } i = \hat{\imath}, \text{ where } \hat{\imath} \text{ is s.t. } o_1^{\sigma,H}(f) = (\hat{x},\hat{k},\hat{\imath}) \\ \varepsilon_1^{\sigma,H}(f)_i & \text{otherwise.} \end{cases}$$

3.4 The Implementation Result

Given a preference profile $R \in \mathcal{R}$ we define, for each ordering $\sigma \in \Sigma$ and selection rule $H : 2^X \to X$, the game in extensive form

$$\Gamma^{\sigma,H}(R) = \left(N, G^\sigma, o^{\sigma,H}, R\right).$$

The main result of the paper states that the extensive-game form $\Gamma^{\sigma,H} = (N, G^\sigma, o^{\sigma,H})$ implements in SPNE the social choice function $\varphi_H : \mathcal{R} \to A$. Formally,

Theorem 1 *Let $R \in \mathcal{R}$, $\sigma \in \Sigma$, and $H : 2^X \to X$ be given. The allocation $\varphi_H(R)$ is the unique SPNE outcome of $\Gamma^{\sigma,H}(R)$.*

Proof See the Appendix at the end of the paper.

Observe that Theorem 1 implies that the unique SPNE outcome of $\Gamma^{\sigma,H}(R)$ does not depend on σ. Moreover, Lemma 3 below states that no agent has to burden an ε-cost in equilibrium.

Lemma 3 *Let $R \in \mathcal{R}$, $\sigma \in \Sigma$, and H be given. Assume (f, b^{σ}) is a SPNE of $\Gamma^{\sigma,H}(R)$. Then, for all $i \in N$, $\varepsilon_i(f, b^{\sigma}) = 0$.*

Proof See the Appendix at the end of the paper.

4 Final Remarks

4.1 Extensive-Game Form

Our mechanism is less simple than we would like. First, in Stage 1 the order in which agents make proposals has to be iterated until all n agents do not make new proposals (if j reacts to i's proposal, i should still be able to counteract). Second, proposers have to be burdened with a cost (which may be very "small") in the case that none of their proposals has been selected at Stage 1, or the proposer $\hat{\imath}$ of the chosen proposal at Stage 1 is either not a final user and/or the number of those who declared their willingness to be users in Stage 2 is smaller than the integer $\hat{k}$ proposed by $\hat{\imath}$ in Stage 1.[8] In the following examples we show that these features are indispensable. In each example we consider the extensive-game form described in Section 3, except that we remove from the original extensive-game form one of the above features.

Example 2 (The order σ of proposals in Stage 1 is not iterated) Let $X = \{x, y, z\}$, $N = \{1, 2\}$, and consider the following selection rule: $H(\{x, y, z\}) = H(\{x, y\}) = H(\{x, z\}) = x$ and $H(\{y, z\}) = y$. Take any $R \in \mathcal{R}$ such that

$$(z, \{1\})\, P_1\, (x, \{1\})\, P_1\, (y, \{\emptyset\})\, P_1\, (y, \{1, 2\})$$

and

$$(y, \{2\})\, P_2\, (z, \{\emptyset\})\, I_2\, (x, \{\emptyset\})\, P_2\, (z, \{1, 2\})\, P_2\, (x, \{1, 2\})\, .$$

Observe that $Z(R) = \{(z, \{1\}), (y, \{2\})\}$. Fix $\sigma(1) = 1$ and $\sigma(2) = 2$. It is easy to check that the unique SPNE outcome of the game without iterating σ in Stage 1 is

[8] The idea of using either small penalties or awards in implementation theory is not new (see Abreu and Mastushima (1994) for penalties and Benoit and Ok (2004) and Sanver (2004) for awards).

the inefficient allocation $(x, \{1\})$. Fix now $\sigma'(1) = 2$ and $\sigma'(2) = 1$. Then, the unique SPNE outcome of the game is the allocation $(z, \{1\})$. Hence, without the iteration of the order in which proposals are made in Stage 1 the SPNE outcome might depend on the exogenously given order, and more importantly, it might be inefficient.

Example 3 (To make a proposal is never costly) Consider the same X, N, H, and R of Example 2. Now, the inefficient allocation $(x, \{1\})$ is a SPNE outcome of the game since there exists a SPNE in which agent 2 first announces $(y, 1)$ and then agent 1 announces $(x, 1)$.

Example 4 (The proposer that is not a final user does not have to burden a cost) Let $X = \{x, y, z\}$ and $N = \{1, 2, 3\}$. Consider the preference profile $R = (R_1, R_2, R_3) \in \mathcal{R}$ where agents 1 and 2 have the same preference relations as in Example 2 and let R_3 be such that

$$(x, \{\emptyset\})\, P_3\, (x, \{N\})\, P_3\, (y, \{N\})\, P_3\, (z, \{N\})\,.$$

Consider the selection rule H of Example 2. Suppose that $\sigma(1) = 3$. There is a SPNE in which agent 3 proposes $(x, 1)$ in Stage 1, no other agent proposes anything else (since $H(\mathcal{X}) = x$ if $x \in \mathcal{X}$). Therefore, the final SPNE outcome is the inefficient allocation $(x, \{1\})$.

Example 5 (Proposer $\hat{\imath}$ does not have to burden a cost when the number of agents willing to use $\hat{x}$ is smaller than $\hat{k}$) Consider the same X, N, H, and R of Example 4. There is a SPNE in which agent 3 proposes $(x, 3)$ and the final SPNE outcome is the inefficient and non-externally stable allocation $(x, \{\emptyset\})$.

4.2 Neutrality

The social choice function $\varphi_H : \mathcal{R} \to A$ implemented in SPNE by our mechanism is anonymous but not neutral. The equilibrium outcome of the game depends on the selection rule H used to select a single alternative for each possible set of alternatives. It is natural to ask whether it is possible to implement the social choice correspondence

$\psi : \mathcal{R} \twoheadrightarrow A$, where for each $R \in \mathcal{R}$, $\psi(R) = MP(R)$. The answer is positive and easy for the case $n \geq 3$. Let $\mathcal{H}$ be the set of all possible selection rules. Add a preliminary stage in the extensive-game form in which all agents simultaneously announce some $H \in \mathcal{H}$. Given $R \in \mathcal{R}$, if at least $n-1$ agents announce the same H then they play the game $\Gamma^{\sigma,H}(R)$, otherwise the game $\Gamma^{\sigma,H'}(R)$ is played with a prespecified selection rule H'. It is straightforward to check that, for all $R \in \mathcal{R}$, the set of SPNE outcomes of this enlarged game coincides with the maximal participation set $MP(R)$.

4.3 Group Stability

Our notion of stability refers to individual decisions. According to our definition a stable allocation is, in fact, a Nash equilibrium outcome of the game played once the public alternative is already selected (see Berga, Bergantiños, Massó, and Neme (2003) for more on this interpretation). We now want to establish the relationship between the set of efficient and stable allocations and the set of group stable allocations. We first state the definition of group stability.

Definition 3 An allocation (x, S) is *group stable under* R if:

(Internal Group Stability) there does not exist $T \subseteq S$ such that, for all $i \in T$, $(x, S \backslash T) P_i (x, S)$;

(External Group Stability) there does not exists any $T \subseteq N \backslash S$ such that, for all $i \in T$, $(x, S \cup T) P_i (x, S)$.

Lemma 4 *Let $R \in \mathcal{R}$ be given. An allocation (x, S) is group stable under R if and only if it is individually stable under R and efficient relative to x.*

Proof Let (x, S_x) be an allocation with the largest stable group. Assume that there exists T such that $T \subseteq N \backslash S_x$ such that, for all $i \in T$, $(x, S_x \cup T) P_i (x, S_x)$. Let $T' \supseteq T$ be the group of agents with maximal cardinality such that, for all $i \in T'$, $(x, S_x \cup T') P_i (x, S_x) I_i (x, \emptyset)$. By (Crow), this contradicts that S_x was the largest stable group since, for all $i \in S_x$, $(x, S_x \cup T')P_i(x, S_x)P_i(x, \emptyset)$. Group internal stability follows by (Crow) and because, by internal stability, $(x, S_x)P_i(x, \emptyset)$ for all $i \in S_x$.

Let (x, S) be a group stable allocation under R and let (x, S_x) be the stable allocation under R and efficient relative to x. Suppose that $\#S_x > \#S$ and define $T = S_x \backslash S$. Observe that, for all $i \in T$, $(x, S \cup T)P_i(x, S)$, contradicting external group stability. ■

References

[1] Abreu, D. and H. Matshushima. "Exact implementation," *Journal of Economic Theory* 64, 1-19 (1994).

[2] Abreu, D. and A. Sen. "Subgame perfect implementation: A necessary and almost sufficient condition," *Journal of Economic Theory* 50, 285-299 (1990).

[3] Bag, P.K. and E. Winter. "Simple subscription mechanisms for excludable public goods," *Journal of Economic Theory* 87, 71-94 (1999).

[4] Benoit, J.P. and E. Ok. "Nash Implementation without no-veto power," mimeo (2004).

[5] Berga, D., G. Bergantiños, J. Massó, and A. Neme. "On exiting after voting," mimeo (2004).

[6] Bogomolnaia, A. and A. Nicolò. "Stable assignment of public facilities under congestion," forthcoming in *Journal of Public Economic Theory* (2004).

[7] Jackson, M. "A crash course in implementation theory," *Social Choice and Welfare* 18, 655-708 (2001).

[8] Jackson, M. and A. Nicolò. "The strategy-proof provision of public goods under congestion and crowding preferences," *Journal of Economic Theory* 115, 278-308 (2004).

[9] Maskin, E. "Nash equilibrium and welfare optimality," *Review of Economic Studies* 66, 23-38 (1999).

[10] Moore, J. and R. Repullo. "Subgame perfect implementation," *Econometrica* 56, 1191-1220 (1988).

[11] Sanver, R. "Nash implementing non-monotonic social choice rules by awards," mimeo (2004).

5 Appendix

5.1 Proof of Theorem 1

We proceed by backwards induction. First, we prove that for any outcome $(\hat{x}, \hat{k}, \hat{\imath})$ of Stage 1, the subgame $\Gamma^{\sigma}(\hat{x}, \hat{k}, \hat{\imath})$ has a unique SPNE outcome (Proposition 3). Then, we show that the outcome of any SPNE of the game $\Gamma^{\sigma,H}(R)$ satisfies the desirable properties (Proposition 4). Finally, we demonstrate that the SPNE outcome of the game $\Gamma^{\sigma,H}(R)$ is unique and coincides with $\varphi_H(R)$ (Proposition 5).

Proposition 3 *Let $R \in \mathcal{R}$ and $\sigma \in \Sigma$ be given. For each outcome of Stage 1 $(\hat{x}, \hat{k}, \hat{\imath}) \in \overline{A} \times N$ the subgame $\Gamma^{\sigma}(\hat{x}, \hat{k}, \hat{\imath})$ has a unique SPNE outcome. Moreover for every SPNE strategy $b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]$ of $\Gamma^{\sigma}(\hat{x}, \hat{k}, \hat{\imath})$,*

$$o_2^{\sigma}(b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]) = \begin{cases} (\hat{x}, S_{\hat{x}}) & \text{if } \#S_{\hat{x}} \geq \hat{k} \\ (\hat{x}, \emptyset) & \text{otherwise,} \end{cases}$$

where $S_{\hat{x}}$ is the maximal stable set of users of $\hat{x}$.

Proof Consider agent $\sigma(n)$ and let $T \in 2^{\mathrm{Pre}(\sigma(n),\sigma)}$ be an arbitrary information set of agent $\sigma(n)$. By (Strict) agent $\sigma(n)$ orders strictly the two allocations $(\hat{x}, \emptyset)$ and $(\hat{x}, T \cup \{\sigma(n)\})$. We distinguish between two cases.

<u>Case 1</u>: $\#(T \cup \{\sigma(n)\}) \geq \hat{k}$.

If $(\hat{x}, \emptyset)\, P_{\sigma(n)}\, (\hat{x}, T \cup \{\sigma(n)\})$ then, for any SPNE participation strategy $b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]$, we have $b^{\sigma}_{\sigma(n)}[\hat{x}, \hat{k}, \hat{\imath}](T) = nu$ and the SPNE outcome of the subgame starting at T is $(\hat{x}, \emptyset)$ if $\#T < \hat{k}$ or $(\hat{x}, T)$ if $\#T \geq \hat{k}$.

If $(\hat{x}, T \cup \{\sigma(n)\})\, P_{\sigma(n)}\, (\hat{x}, \emptyset)$ then, for any SPNE participation strategy $b^{\sigma}[\hat{x}, \hat{k}, \hat{\imath}]$, we have $b^{\sigma}_{\sigma(n)}[\hat{x}, \hat{k}, \hat{\imath}](T) = u$ and the SPNE outcome of the subgame starting at T is $(\hat{x}, T \cup \{\sigma(n)\})$.

<u>Case 2</u>: $\#\{T \cup \{\sigma(n)\}\} < \hat{k}$.

Then any SPNE outcome of the subgame starting at T is $(\hat{x}, \emptyset)$. However, consider the participation strategy $\tilde{b}^{\sigma}_{\sigma(n)}[\hat{x}, \hat{k}, \hat{\imath}]$ of agent $\sigma(n)$ that coincides with $b^{\sigma}_{\sigma(n)}[\hat{x}, \hat{k}, \hat{\imath}]$

but at all information sets T' where $\#T' + 1 < \hat{k}$ it is defined by

$$\tilde{b}^{\sigma}_{\sigma(n)}[\hat{x},\hat{k},\hat{\imath}](T') = \begin{cases} u & \text{if } \sigma(n) \in S_{\hat{x}} \\ nu & \text{otherwise.} \end{cases}$$

Thus, by the backwards induction principle, we can replace the information set T of $\sigma(n)$ by the unique outcome previously identified, generated also by the participation strategy profile $\tilde{b}^{\sigma}[\hat{x},\hat{k},\hat{\imath}] = (\tilde{b}^{\sigma}_{\sigma(n)}[\hat{x},\hat{k},\hat{\imath}], b^{\sigma}_{-\sigma(n)}[\hat{x},\hat{k},\hat{\imath}])$. Following the induction argument we obtain the uniqueness of the outcome and a SPNE strategy $\tilde{b}^{\sigma}(\hat{x},\hat{k},\hat{\imath})$. Moreover, it is straightforward to check that for all $i \in N$, $\pi_i(\tilde{b}^{\sigma}[\hat{x},\hat{k},\hat{\imath}]) = \pi_i(b^{\sigma}[\hat{x},\hat{k},\hat{\imath}]) = 1$ if and only if $i \in S_{\hat{x}}$. Therefore $o_2^{\sigma}(b^{\sigma}[\hat{x},\hat{k},\hat{\imath}])$ has the desired property. ■

Corollary 1 *Let $R \in \mathcal{R}$ and $\sigma, \sigma' \in \Sigma$ be given. Let $(\hat{x},\hat{k},\hat{\imath}) \in \overline{A} \times N$ be the outcome of Stage 1. Assume $b^{\sigma}[\hat{x},\hat{k},\hat{\imath}]$ is a SPNE of the subgame $\Gamma^{\sigma}(\hat{x},\hat{k},\hat{\imath})$ and $b^{\sigma'}[\hat{x},\hat{k},\hat{\imath}]$ is a SPNE of the subgame $\Gamma^{\sigma'}(\hat{x},\hat{k},\hat{\imath})$. Then, $o_2^{\sigma}(b^{\sigma}[\hat{x},\hat{k},\hat{\imath}]) = o_2^{\sigma'}(b^{\sigma'}[\hat{x},\hat{k},\hat{\imath}])$.*

Proposition 4 *Let $R \in \mathcal{R}$, $\sigma \in \Sigma$, and $H : 2^X \to X$ be given. Assume (f, b^{σ}) is a SPNE of $\Gamma^{\sigma,H}(R)$. Then,*
(1) $o^{\sigma,H}(f, b^{\sigma})$ is stable under R.

(2) $o^{\sigma,H}(f, b^{\sigma})$ is efficient under R.

(3) $o^{\sigma,H}(f, b^{\sigma})$ belongs to the maximal participation set $MP(R)$.

Proof Since $\Gamma^{\sigma,H}(R)$ is a finite extensive form game with perfect information it has at least a SPNE in pure strategies. Let (f, b^{σ}) be a SPNE of $\Gamma^{\sigma,H}(R)$ and let $o^{\sigma,H}(f, b^{\sigma}) = (x, S) \in A$ be its outcome. We first establish the following two claims.

Claim 1 If $i \notin S$ then $\varepsilon_i(f, b^{\sigma}) = 0$.

Proof of Claim 1 Assume $i \notin S$ and $\varepsilon_i(f, b^{\sigma}) \neq 0$ holds. This means that agent i made a proposal in $\overline{A}$ at Stage 1. Consider the strategy $(\tilde{f}_i, \tilde{b}^{\sigma}_i)$ where $\tilde{f}_i(\cdot) = (NP, 0)$ and for all $(\hat{x},\hat{k},\hat{\imath})$, $\tilde{b}^{\sigma}_i[\hat{x},\hat{k},\hat{\jmath}](T) = nu$ for all $T \in 2^{Pre(i,\sigma)}$. Using this strategy i does not consume the public good neither he has to burden any cost. Hence, (f_i, b^{σ}_i) is not one of his best replies to $(f_{-i}, b^{\sigma}_{-i})$.

Claim 2 Let $(\hat{x}, \hat{k}, \hat{\imath})$ be an outcome of Stage 1 and assume there exists $i \in N$ such that $\pi_i(b^\sigma[\hat{x}, \hat{k}, \hat{\imath}]) = 0$. Consider the strategy $\tilde{b}_i^\sigma$ that coincides with b_i^σ except that $\pi_i(\tilde{b}_i^\sigma[\hat{x}, \hat{k}, \hat{\imath}], b_{-i}^\sigma[\hat{x}, \hat{k}, \hat{\imath}]) = 1$; then, for all $j \in N$ such that $\pi_j(b^\sigma[\hat{x}, \hat{k}, \hat{\imath}]) = 1$ we have $\pi_j(\tilde{b}_i^\sigma[\hat{x}, \hat{k}, \hat{\imath}], b_{-i}^\sigma[\hat{x}, \hat{k}, \hat{\imath}]) = 1$.

Proof of Claim 2 It follows immediately from the following observation: since $b^\sigma[\hat{x}, \hat{k}, \hat{\imath}]$ is a SPNE of the subgame starting at $(\hat{x}, \hat{k}, \hat{\imath})$, by (Crow), for all $j \in N$, $b_j^\sigma[\hat{x}, \hat{k}, \hat{\imath}]\,(T) = u$ implies $b_j^\sigma[\hat{x}, \hat{k}, \hat{\imath}]\,(T') = u$ for all $T, T' \in 2^{\mathrm{Pre}(\sigma,j)}$ with $\#T' \geq \#T+1$.

(1) $o^{\sigma,H}(f, b^\sigma)$ is stable under R.

If $i \in S$ and $(x, \emptyset)P_i(x, S)$ then i is not best-replying since the strategy $\tilde{b}_i^\sigma$ that coincides with b_i^σ except that $\pi_i(\tilde{b}_i^\sigma[\hat{x}, \hat{k}, \hat{\imath}], b_{-i}^\sigma[\hat{x}, \hat{k}, \hat{\imath}]) = 0$ has the property that $o^{\sigma,H}(f, (b_{-i}^\sigma, \tilde{b}_i^\sigma)) = (x, T)$ with $i \notin T$. Hence (x, S) is internally stable. To show that (x, S) is externally stable, assume $i \notin S$ and

$$(x, S \cup \{i\})P_i(x, \emptyset). \tag{1}$$

We distinguish between two different cases:

Case 1.1: $S = \emptyset$.

(1.1.1) $o_1^{\sigma,H}(f) = (H(\emptyset), \emptyset)$. Consider $\tilde{f}_i(NP, 0) = (x, 1)$. If $o_1^{\sigma,H}(f_{-i}, \tilde{f}_i) = (x, 1, i)$, that is, nobody else made a proposal after $(x, 1)$, then by (1) and Proposition 3, $(x, S_x) = o^{\sigma,H}(f_{-i}, \tilde{f}_i, b^\sigma)P_i o^{\sigma,H}(f, b^\sigma) = (x, \emptyset)$, contradicting that g^σ is a SPNE of $\Gamma^{\sigma,H}(R)$. If $o_1^{\sigma,H}(f_{-i}, \tilde{f}_i) = (y, k, j) \neq (x, 1, i)$, with $k \geq 1$, that is, agent j made the definitive proposal of Stage 1 triggered by $i's$ deviation. Observe that by Claim 2 $\pi_j(b^\sigma(y, k, j)) = 1$. Let m be the step at which agent j proposed (y, k) after $i's$ deviation; namely, $j = \hat{\sigma}(m)$, $p_{\hat{\sigma}(m)}(f_{-i}, \tilde{f}_i) = (y, k)$ and, $p_{\hat{\sigma}(m+1)}(f_{-i}, \tilde{f}_i) = \ldots = p_{\hat{\sigma}(m+n)}(f_{-i}, \tilde{f}_i) = (NP, 0)$. Consider j's deviation such that $\tilde{f}_j(NP, 0) = (y, k)$ and for all k', $\tilde{f}_j(C_{k'}) = f_j(C_{k'})$ for all $C_{k'} \in \mathcal{C}_{k'}$. Let $\tilde{m}$ be the step at which agent j proposed (y, k) in the original equilibrium path; namely, $j = \hat{\sigma}(\tilde{m})$ and $p_{\hat{\sigma}(\tilde{m})}(f_{-j}, \tilde{f}_j) = (y, k)$. Let $l = \hat{\sigma}(\tilde{m} + 1) = \hat{\sigma}(m + 1)$ (i.e., $\sigma(\sigma^{-1}(j) + 1) = l$). Consider the information sets of agent l at the step $m + 1$ in $path^\sigma(f_{-i}, \tilde{f}_i)$, $C_k = \{(y, k)\} \cup \overline{C}$, where $\overline{C}$ is a (potentially empty) subset of $\overline{A}$ and at the step $\tilde{m} + 1$ in $path^\sigma(f_{-j}, \tilde{f}_j)$, $\tilde{C}_k =$

$\{(y,k)\}$. Observe that although $F_l(\tilde{C}_k) \supseteq F_l(C_k), \overline{C} = F_l(\tilde{C}_k) \backslash F_l(C_k)$, and thus, by IIA, if $p_{\hat{\sigma}(m+1)}(f_{-i}, \tilde{f}_i) = (NP, 0)$ then $p_{\hat{\sigma}(\tilde{m}+1)}(f_{-j}, \tilde{f}_j) = (NP, 0)$. Iterating this argument, we deduce that $p_{\hat{\sigma}(\tilde{m}+1)}(f_{-j}, \tilde{f}_j) = ... = p_{\hat{\sigma}(\tilde{m}+n)}(f_{-j}, \tilde{f}_j) = (NP, 0)$. Hence, $o_1^{\sigma,H}(f_{-j}, \tilde{f}_j) = (y,k,j)$ and $o^{\sigma,H}(f_{-j}, \tilde{f}_j, b^\sigma) P_j o^{\sigma,H}(f, b^\sigma)$, contradicting that g^σ is a SPNE of $\Gamma^{\sigma,H}(R)$.

(1.1.2) $o_2^g(b^\sigma(x,k,i)) = (x,\emptyset)$ because the outcome of Stage 1, (x,k,i), has the property that $k > \Sigma_{j\in N}\pi_j(b^\sigma(x,k,i))$. Then, by Claim 1, g^σ is not a SPNE of $\Gamma^{\sigma,H}(R)$.

Case 1.2: $S \neq \emptyset$.

Let $(x,k,j) = o_1^{\sigma,H}(f)$. Then, $\#S \geq k$. But then, by Proposition 3, $S = S_x$. Thus, (x,S) is stable.

(2) $o^{\sigma,H}(f, b^\sigma) = (x,S)$ is efficient under R.

Assume (x,S) is stable but not efficient under R. By Lemma 2, there exists a stable allocation (y,S_y) such that $(y,S_y)PD(x,S)$. First, note that $S_y \supseteq S$. By (Apa), S_y is not empty. We distinguish between two different cases:

Case 2.1: $S = \emptyset$.

(2.1.1) $o_1^{\sigma,H}(f) = (H(\emptyset),\emptyset) = (x,S)$. There exists $i \in S_y$ with the deviation $\tilde{g}_i = (\tilde{f}_i, b_i^\sigma)$ equal to $g_i = (f_i, b_i^\sigma)$ except that $\tilde{f}_i(NP,0) = (y,\#S_y)$. If $(y,\#S_y,i)$ is the outcome of Stage 1 then, by Proposition 3, $o^{\sigma,H}(g_{-i},\tilde{g}_i) = (y,S_y)$, contradicting that g is a SPNE of $\Gamma^{\sigma,H}(R)$. If $o_1^{\sigma,H}(f_{-i},\tilde{f}_i) = (z,k,j) \neq (y,\#S_y,i)$, with $k \geq \#S_y$, then applying a similar argument than we use in case (1.1.1) to prove external stability (and noting that $(z,S_z)R_j(y,S_y)P_j(x,S)$), we conclude that (z,k) was a profitable deviation for agent j in the original path g.

(2.1.2) $o_2^g(b^\sigma(x,k,j)) = (x,\emptyset)$ because the outcome of Stage 1, (x,k,j), has the property that $k > \Sigma_{j'\in N}\pi_{j'}(b^\sigma(x,k,j))$. Then, by Claim 1, g^σ is not a SPNE of $\Gamma^{\sigma,H}(R)$.

Case 2.2: $S \neq \emptyset$.

Let $o_1^{\sigma,H}(f) = (x,k,i)$. Observe that, by Proposition 3, $\#S \geq k$ and, by Claim 1, $i \in S$. Hence, there exists an information set $C_{k'}$ such that $f_i(C_{k'}) = (x,k)$. Consider the deviation $\tilde{f}_i$ equal to f_i except that $\tilde{f}_i(C_{k'}) = (y,\#S_y)$. If $o_1^{\sigma,H}(f_{-i},\tilde{f}_i) = (y,\#S_y,i)$

then i was not best replying since, by Proposition 3, $o^{\sigma,H}(g_{-i}, \tilde{g}_i) = (y, S_y)$. Assume $o_1^{\sigma,H}(f_{-i}, \tilde{f}_i) = (z, t, j)$. By CLAIM 2, $\pi_j(b^\sigma(z, t, j)) = 1$, and therefore, by Proposition 3, $o_2^g(b^\sigma(z, t, j)) = (z, S_z)$. Then, applying a similar argument than we use in case (1.1.1) to prove external stability, we conclude that (z, t) was a profitable deviation for agent j in the original path g.

(3) $o^{\sigma,H}(f, b^\sigma) \in MP(R)$.

Let (x, S) be stable and efficient under R and assume there exists (y, S_y) maximal, stable and efficient under R such that $\#S_y > \#S$. Consider $i \in S_y \backslash S$ and a deviation $\tilde{f}_i$ such that in the information set C_k at the equilibrium path, $\tilde{f}_i(C_k) = (y, \#S_y)$. Either the outcome of Stage 1 is $(y, \#S_y)$, in which case g was not SPNE, or $o_1^{\sigma,H}(f_{-i}, \tilde{f}_i) = (z, t, j)$ with $t \geq \#S_y$. By Proposition 3 and CLAIM 2, $o_2^g(b^\sigma(z, t, j)) = (z, S_z)$ and there exists $l \in S_z \backslash S_x$ who could have proposed (z, t) in the equilibrium path of g^σ. ■

Proposition 5 *Let $R \in \mathcal{R}$, $\sigma \in \Sigma$, and $H : 2^X \to X$ be given. Assume (f, b^σ) is a SPNE of $\Gamma^{\sigma,H}(R)$. Then, $o^{\sigma,H}(f, b^\sigma) = \varphi_H(R)$.*

Proof Let $H(MP(R)) = x$. Assume $S_x = \emptyset$. Then the outcome of the game $\Gamma^{\sigma,H}(R)$ is unique and equal to $(H(\emptyset), \emptyset)$. In fact, by Proposition 3, no agent in equilibrium will announce u in any subgame $\Gamma^\sigma(\hat{x}, \hat{k}, \hat{\imath})$, since any proposer of a pair in $\overline{A}$ will have to burden a cost; hence, in equilibrium no proposal in $\overline{A}$ is made in Stage 1. Observe that $\varphi_H(R) = (x, \emptyset)$. Assume $S_x \neq \emptyset$ and let (y, S_y) be an equilibrium outcome of $\Gamma^{\sigma,H}(R)$ such that $(x, S_x) \neq (y, S_y)$. By Proposition 4, $y \in MP(R)$ and $\#S_x = \#S_y$. If $x = y$ then, by Lemma 1, $S_x = S_y$. Hence, $x \neq y$. Since (x, S_x) is efficient under R there exists i such that $(x, S_x)\, P_i(y, S_y)$. Let $g^\sigma = (f, b^\sigma)$ be a SPNE strategy that generates (y, S_y) and consider the deviation $\tilde{f}_i$ consisting of proposing $(x, \#S_x)$ just after y has been proposed (together with some integer smaller or equal than $\#S_y$). Then, by IIA of H, $(x, \#S_x, i)$ is the outcome of Stage 1. Hence, (x, S_x) is the outcome of $(g_{-i}^\sigma, \tilde{g}_i^\sigma)$, contradicting that g^σ was a SPNE of $\Gamma^{\sigma,H}(R)$. Since $\Gamma^{\sigma,H}(R)$

is a finite extensive form game with perfect information it has at least a SPNE in pure strategies. Therefore $\varphi_H(R) = (H(MP(R)), S_{H(MP(R))}) = (x, S_x)$. ■

5.2 Proof of Lemma 3

Assume $\varepsilon_i(f, b^\sigma) > 0$ for some $i \in N$. Then, by CLAIM 1 in the proof of Proposition 4, there are at least two agents who made a proposal at Stage 1. Let $o_1^{\sigma,H}(f) = (\hat{x}, \hat{k}, \hat{\imath})$.

CLAIM 3 For all $j \in N$, $f_j(C_{\hat{k}}) = (NP, 0)$ for any $C_{\hat{k}} \in \mathcal{C}_{\hat{k}}$ such that $(\hat{x}, \hat{k}) \in C_{\hat{k}}$.

PROOF OF CLAIM 3 Suppose otherwise, and let $f_j(C_{\hat{k}}) = (y, k')$. Since $o_1^{\sigma,H}(f) = (\hat{x}, \hat{k}, \hat{\imath})$, $k' = k$. Consider the consumption strategy $\tilde{f}_j$ that coincides with f_j in all information sets except those $C_{\hat{k}} \in \mathcal{C}_{\hat{k}}$ such that $(\hat{x}, \hat{k}) \in C_{\hat{k}}$, in which case, $\tilde{f}_j(C_{\hat{k}}) = (NP, 0)$. Then, either $o_1^{\sigma,H}(\tilde{f}_j, f_{-j}) = (\hat{x}, \hat{k}, \hat{\imath})$, in which case f_j is not a best reply, or else $o_1^{\sigma,H}(\tilde{f}_j, f_{-j}) = (\hat{z}, \hat{q}, \hat{l})$. But then, $o_2^g(b^\sigma[o_1^{\sigma,H}(\tilde{f}_j, f_{-j})]) P_{\hat{l}} o_2^g(b^\sigma[o_1^{\sigma,H}(f)])$ which means that agent $\hat{l}$ has a profitable deviation from f. This proves the Claim.

By CLAIM 3 there exists $j \in N$ such that $f_j(NP, 0) = (z, \tilde{k}) \neq (\hat{x}, \hat{k})$. Moreover, by CLAIM 2 in the proof of Proposition 4, $\pi_j(b^\sigma[\hat{x}, \hat{k}, \hat{\imath}]) = 1$. Consider another consumption strategy $\tilde{f}_j$ that coincides with f_j in all information sets except in $(NP, 0)$ where $\tilde{f}_j(NP, 0) = (\hat{x}, \hat{k})$. By IIA and CLAIM 3, $o_1^{\sigma,H}(\tilde{f}_j, f_{-j}) = (\hat{x}, \hat{k}, j)$ and agent j does not have to burden a cost since his proposal has been selected. ■