

# Efficient and Stable Collective Choices under Gregarious Preferences\*

Jordi Massó<sup>†</sup>      Antonio Nicolò<sup>‡</sup>

February 2007

## Abstract

We consider collective choice problems where a set of agents have to choose an alternative from a finite set and agents may or may not become users of the chosen alternative. An allocation is a pair given by the chosen alternative and the set of its users. Agents have gregarious preferences over allocations: given an allocation, they prefer that the set of users becomes larger. We require that the final allocation be efficient and stable (no agent can be forced to be a user

---

\*We are grateful to two anonymous referees for their thoughtful comments and suggestions that greatly improved the paper. We have also received very useful comments from Pablo Amorós, Carmen Beviá, Sandro Brusco, Luis Corchón, Humberto Llavador, Bernardo Moreno, David Pérez-Castrillo, Socorro Puy, Joaquim Silvestre, Marc Vorsatz, and Ashel Wolinsky. The work of Massó is partially supported by the Spanish Ministry of Education and Science through grant SEJ2005-01481/ECON and FEDER and through project CONSOLIDER-INGENIO 2010 (CDS2006-00016), and by the Generalitat de Catalunya through grant SGR2005-00454 and through the Barcelona Economics Program (XREA). The work of Nicolò is partially supported by the Italian Ministry of University and Research through grant 2005137858.

<sup>†</sup>Departament d'Economia i d'Història Econòmica and CODE, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain. Fax: (34) 935812461. E-mail: jordi.massó@uab.es.

<sup>‡</sup>Corresponding author. Dipartimento di Scienze Economiche, Università di Padova, via del Santo 33, 35123 Padova, Italy. Fax: 00390498274211. E-mail: antonio.nicolo@unipd.it.

and no agent who wants to be a user can be excluded). We propose a two-stage sequential mechanism whose unique subgame perfect equilibrium outcome is an efficient and stable allocation which also satisfies a maximal participation property.

*Keywords:* Public Goods, Gregarious Preferences, Subgame Perfect Implementation.

*JEL Classification Numbers:* D62, D71, H41.

## 1 Introduction

In many collective choice problems, after the social alternative (or level of a public good) has been chosen, agents may decide whether or not to use it. If the final set of users affects the welfare of each member, then the decision process has to take into account the set of agents that will eventually become users. In this paper we study the case when participation is not compulsory and agents' preferences are gregarious in the sense that they are (set-inclusion) monotonic with respect to the set of users. 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 to not pay for the facility. Many other problems do not directly involve money but can be similarly modeled. For instance, 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 the common feature that agents also care about the set of users of the public good. For this reason, and since participation is not compulsory, the final allocation has to satisfy not only efficiency but also stability. While the first requirement is well-known and desirable in many collective choice problems, the latter deserves to be briefly explained. 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 technological or institutional constraints: for instance, no nation can be forced to adopt any technological standard, or according to the law agents can not be forced to contribute to the provision of a public facility, but at the same time it is impossible or too costly to exclude someone from its consumption.<sup>1</sup>

The aim of this paper is to implement an efficient and stable social choice function when agents' gregarious preferences are private information. We focus on the case where agents have to select a *single* alternative. Namely, only one level of the public good can be provided and agents have to decide whether or not to be users of it, but a subset of agents *cannot* secede and choose (and use) another alternative.

Our analysis starts by showing that, for any gregarious 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 (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 independently of agents' preferences. Namely, when gregarious 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 result suggests that the trade-off between informational constraints and normative properties of social choice functions

---

<sup>1</sup>See also Bogomolnaia and Nicolò (2005) for a brief discussion of the normative content of a slightly different definition of stability in the context of multiple provision of public goods.

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 present a mechanism with two features: first, it is invariant with respect to the names of the agents and second, it is simple enough to be implemented in real collective decision problems. Roughly, the proposed two-stage game is as follows. In Stage 1 agents sequentially (iteratively and publicly), following an exogenously 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 (if there are several, one of them is chosen in accordance with a pre-specified selection rule). In Stage 2 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 penalties out-of-equilibrium play 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 take decisions along the game. We think that this is an important property should the mechanism be used. The mechanism selects among the set of efficient and stable allocations an alternative which maximizes the number of its users. 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 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 be 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 mechanism and state our main result. In Section 4, we offer some examples that illustrate the role of some features of the mechanism, discuss its non-neutrality, 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 formal definitions of strategies and outcome functions of the game in extensive form and collects the proofs omitted in the text.

## 2 Preliminaries

Let  $N = \{1, \dots, n\}$  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:

(GREG) GREGARIOUSNESS: For all  $x \in X$  and  $S, T \in 2^N$  such that  $i \in T \subset S$ ,

$(x, S) P_i(x, T)$ .

(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 not  $(x, S) I_i(y, T)$ .

GREGARIOUSNESS implies that adding new members to a group of users increases users' welfare. Notice that GREGARIOUSNESS does not impose any condition when comparing two allocations with different levels of public good. In particular, GREGARIOUSNESS admits the possibility that agent  $i \in S \cap T$  strictly prefers  $(x, S)$  to  $(x', T)$ , for  $x \neq x'$ , even if  $S \subset T$ . Moreover, GREGARIOUSNESS does not imply anonymity: agent  $i \in S \cap T$  may prefer  $(x, S)$  to  $(x, T)$ , even if  $\#S < \#T$  when there exists some  $j \in S$  but  $j \notin T$ . Hence GREGARIOUSNESS restricts agents' preference to be positively affected by the enlargement of the set of users, but it also admits that agents may have non-anonymous preferences over the set of users (*i.e.*, agents are not only interested in the cardinality of the set of users but also in their identities).<sup>2</sup>

APATHY says that agent  $i$  does not care about the level of the public good if he does not use it.<sup>3</sup> 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 three properties is called a *gregarious preference relation* and  $\mathcal{R}_i$  denotes the set of all such preference relations for agent  $i$ . Notice that all conditions are agent specific and therefore  $\mathcal{R}_i \neq \mathcal{R}_j$  for different

---

<sup>2</sup>A preference relation  $R_i$  is *anonymous* if for all  $x \in X$  and  $S, T \in 2^N$  such that  $i \in S \cap T$  and  $\#S = \#T$ ,  $(x, S) I_i(x, T)$ .

<sup>3</sup>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 too restrictive. Nevertheless in many interesting contexts, like the provision of club goods, it seems a natural assumption.

agents  $i$  and  $j$ . We want to stress that the set of gregarious preference relations for agent  $i$  admits preferences with very different trade-offs between the selected level of the public good and the set of its users; for instance, gregarious preference relations  $R_i$  and  $R'_i$  might well order (1)  $(x, \{i\})P_i(y, N)$  for  $x \neq y$ , and (2)  $(x, \emptyset)P'_i(x, N)$ . The reason is that gregarious effects might be relatively small for agent  $i$  when compared with  $i$ 's evaluation of the alternatives. For instance, in (1)  $y$  may be considered as being a very bad alternative compared with  $x$ ; and in (2),  $i$  considers  $x$  undesirable even if all agents use it.

A *profile*  $R = (R_1, \dots, R_n)$  is a  $n$ -tuple of gregarious preference relations. Let  $\mathcal{R} = \mathcal{R}_1 \times \dots \times \mathcal{R}_n$  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  $(x, S)P_i(x, S \setminus \{i\})$ .

(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  $R$  while  $S_x$  is set-inclusion maximal among all*

$T \in 2^N$  where  $(x, T) \in A$  is stable under  $R$ . Moreover  $(x, S_x)$  Pareto dominates all allocations  $(x, T)$  with  $T \neq S_x$  that are stable under  $R$ .

**Proof** Let  $R \in \mathcal{R}$  and  $x \in X$  be given. For each  $1 \leq k \leq n$ , define

$$N^k(x) = \{S \in 2^N \mid \#S = k \text{ and } (x, S)P_i(x, \emptyset) \text{ for all } i \in S\}. \quad (1)$$

Namely,  $N^k(x)$  is the family of sets of users with cardinality  $k$  that satisfy internal stability at  $x$  under  $R$ .

First, suppose

$$N^k(x) = \emptyset \text{ for all } k = 1, \dots, n. \quad (2)$$

Then, set  $S_x = \emptyset$ . The allocation  $(x, \emptyset)$  satisfies internal stability trivially. Since  $N^1(x) = \emptyset$ ,  $(x, \emptyset)$  is externally stable and therefore stable under  $R$ . By (2), there does not exist  $T \neq \emptyset$  such that  $(x, T)$  is stable under  $R$ . Hence, the conclusions of Lemma 1 follow in this case.

Second, suppose there exists  $1 \leq k \leq n$  such that  $N^k(x) \neq \emptyset$ . Let  $1 \leq K \leq n$  be such that  $N^K(x) \neq \emptyset$  and for all  $k > K$  (if any),  $N^k(x) = \emptyset$ .

CLAIM 1:  $\#N^K(x) = 1$ .

PROOF OF CLAIM 1: Assume  $\#N^K(x) > 1$ . Let  $S, S' \in N^K(x)$  with  $S \neq S'$ . By (GREG), the definition of  $N^K(x)$ , and the facts that  $S \subsetneq S \cup S'$  and  $S' \subsetneq S \cup S'$  hold, it follows that  $(x, S \cup S')P_i(x, S)P_i(x, \emptyset)$  if  $i \in S$ , and  $(x, S \cup S')P_i(x, S')P_i(x, \emptyset)$  if  $i \in S'$ . Hence,  $S \cup S' \in N^k(x)$  for some  $k > K$ , contradicting the definition of  $K$ . This ends the proof of CLAIM 1.

Let  $S_x$  be the unique set of users in  $N^K(x)$ .

CLAIM 2:  $(x, S_x)$  is stable under  $R$ .

PROOF OF CLAIM 2: Internal stability of  $(x, S_x)$  directly follows from (1). If  $(x, S_x)$  is not externally stable, then there exists  $i \notin S_x$  such that  $(x, S_x \cup \{i\})P_i(x, S_x)$ . By (APA),  $(x, S_x)I_i(x, \emptyset)$ . By (GREG),  $(x, S_x \cup \{i\})P_j(x, S_x)$  for all  $j \in S_x$ . Hence  $S_x \cup \{i\} \in N^{K+1}(x)$ , contradicting the definition of  $K$ . This ends the proof of CLAIM 2.

CLAIM 3: If  $(x, T)$  is internally stable under  $R$ , then  $S_x \supseteq T$ .

PROOF OF CLAIM 3: Assume that  $(x, T)$  is internally stable and there exists

$$i \in T \setminus S_x. \quad (3)$$

Consider  $S_x \cup T$ . By definition of  $S_x$ ,  $\#S_x = K$ . Thus, by (3),  $\#(S_x \cup T) > K$ . By definition of  $K$ ,  $S_x \setminus T \neq \emptyset$ ; otherwise,  $T \in N^k(x)$  for some  $k > K$ . By (GREG) and internal stability of  $(x, T)$  and  $(x, S_x)$ ,  $(x, S_x \cup T)P_i(x, S_x)P_i(x, \emptyset)$  for all  $i \in S_x$  and  $(x, S_x \cup T)P_i(x, T)P_i(x, \emptyset)$  for all  $i \in T$ . Hence,  $S_x \cup T \in N^k(x)$  for some  $k > K$ , which is a contradiction. This ends the proof of CLAIM 3.

Assume  $(x, T) \in A$  is stable under  $R$ . Then,  $(x, T)$  is internally stable and, by CLAIM 3,  $S_x \supseteq T$ . Thus,  $S_x$  is set-inclusion maximal among all  $T \in 2^N$  where  $(x, T) \in A$  is stable under  $R$ .

CLAIM 4: Let  $(x, T)$  be stable under  $R$  and  $T \neq S_x$ . Then,  $(x, S_x)$  Pareto dominates  $(x, T)$ .

PROOF OF CLAIM 4: Let  $(x, T)$  be stable under  $R$  and  $T \neq S_x$ . By CLAIM 3,  $T \not\subseteq S_x$ . Thus,  $S_x \neq \emptyset$ . Hence, by (GREG),  $(x, S_x)P_i(x, T)$  for all  $i \in S_x$  and  $(x, S_x)I_i(x, T)$  for all  $i \notin S_x$ . Thus,  $(x, S_x)PD(x, T)$ . This ends the proof of CLAIM 4 and Lemma 1. ■

Observe that gregarious preference profiles admit the possibility that the maximal stable set of users of an alternative  $x$  is the empty set. 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 at  $x$* . As a direct consequence of Lemma 1, Lemma 2 below states that if  $(x, S_x)$  is Pareto dominated, then it is Pareto dominated by some stable  $(y, S_y)$ .

**Lemma 2** *Let  $R \in \mathcal{R}$  and  $x \in X$  be given. Assume  $(x, S_x)$  is not efficient under  $R$ . Then there exists  $y \in X$  such that the stable allocation  $(y, S_y)$  Pareto dominates  $(x, S_x)$ .*

**Proof** Let  $R \in \mathcal{R}$  and  $x \in X$  be given. Assume that  $(x, S_x) \in A$  is not efficient under  $R$ . There exists  $(y, T) \in A$  such that  $(y, T)PD(x, S_x)$ . Next, we show that

$(y, T)$  satisfies internal stability. Assume otherwise; namely, there exists  $i \in T$  such that

$$(y, T \setminus \{i\})P_i(y, T). \quad (4)$$

By (APA),  $(y, T \setminus \{i\})I_i(y, \emptyset)$ . If  $i \in S_x$  then, by internal stability of the allocation  $(x, S_x)$ ,  $(x, S_x)P_i(x, S_x \setminus \{i\})$  and by (APA),  $(x, S_x \setminus \{i\})I_i(y, \emptyset)$ . Hence, by (4) and transitivity,  $(x, S_x)P_i(y, T)$ . Thus,  $(y, T)$  cannot Pareto dominate  $(x, S_x)$ . If  $i \notin S_x$  then, by (APA),  $(x, S_x)I_i(y, \emptyset)$ . By (4),  $(x, S_x)P_i(y, T)$  which contradicts that  $(y, T)PD(x, S_x)$ . Hence,  $(y, T)$  is internally stable. By CLAIM 3 in the proof of Lemma 1 and definition of  $S_y$ , either  $T = S_y$  or  $T \subsetneq S_y$ . If  $T = S_y$ , then  $(y, S_y)$  is stable under  $R$  by Lemma 1 and  $(y, S_y)PD(x, S_x)$ . If  $T \subsetneq S_y$ , by (GREG),  $(y, S_y)PD(y, T)$ . By transitivity of the Pareto dominance relation  $(y, S_y)PD(x, S_x)$ , and the proof is completed. ■

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 it is not Pareto dominated by any other *stable* allocation. Second, given that the set of stable allocations is non-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  $R$ . ■

A *social choice function* is a mapping  $\varphi : \mathcal{R} \rightarrow X \times 2^N$  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 preference relations 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 gregarious preference relations (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 gregarious preference relations. Before stating this result we need some additional notation and definitions.

A *mechanism* (or game form) is a pair  $(M, \Phi)$  where  $M = M_1 \times \cdots \times M_n$  is a Cartesian product of message spaces (one for each agent) and  $\Phi : M \rightarrow A$  is an outcome function. Thus, each player  $i$  submits a message  $m_i \in M_i$  and, given  $(m_1, \dots, m_n) \in M$ , the allocation  $\Phi(m_1, \dots, m_n) \in A$  is selected. A social choice function  $\varphi : \mathcal{R} \rightarrow A$  is *Nash implementable* if there exists a mechanism  $(M, \Phi)$  such that for all  $R \in \mathcal{R}$ ,  $\varphi(R) = \Phi(m_1^*, \dots, m_n^*)$  for all Nash equilibria  $(m_1^*, \dots, m_n^*) \in M$  of the induced normal form game  $(N, (M, \Phi), R)$ . A social choice function  $\varphi : \mathcal{R} \rightarrow A$  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  $aR_i b$  and  $bP'_i a$ . Maskin monotonicity is a necessary condition for a social choice function to be Nash implementable.<sup>4</sup>

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

**Proof** Let  $\varphi : \mathcal{R} \rightarrow X \times 2^N$  be an efficient and stable social choice function. Take  $x, y \in X$  arbitrary and select any profile  $R \in \mathcal{R}$  of gregarious preference relations with the following properties: (1)  $(x, N)P_1(y, N)P_1(z, \{\emptyset\})P_1(z, N)$  for all  $z \neq x, y$  (if any); (2) for all  $i \neq 1$ ,  $(y, N)P_i(x, N)P_i(z, \{\emptyset\})P_i(z, N)$  for all  $z \neq x, y$  (if any); and (3) for all  $i \in N$  and  $S \in 2^N$  such that  $S \neq N$  and  $i \in S$ ,

$$(z, N)P_i(z', S) \text{ for all } z, z' \in X. \quad (5)$$

---

<sup>4</sup>See, for instance, Maskin (1999)'s original paper or Jackson (2001)'s survey on implementation theory.

That is, no agent wants to be a user of an alternative  $z \neq x, y$ , agent 1 prefers alternative  $x$  to  $y$ , all other agents prefer alternative  $y$  to  $x$ , but all agents prefer to be users of their second best alternative with the entire society than be users of their preferred alternative with a smaller group of users. By efficiency, either  $\varphi(R) = (x, N)$  or  $\varphi(R) = (y, N)$ .

Assume  $\varphi(R) = (y, N)$ . The case  $\varphi(R) = (x, N)$  proceeds similarly. Consider now the gregarious preference relation  $R'_1 \in \mathcal{R}_1$  with the following properties: (1) for all  $y' \in X$  and 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)  $(x, N)P'_1(z, \emptyset)P'_1(z, N)$  for all  $z \neq x$ . Thus, agent 1 does not want to use the public good at any level  $z \neq x$ , regardless of the set of its users. Hence, by stability, if  $\varphi(R'_1, R_{-1}) = (z, S)$  with  $z \neq x$  then  $1 \notin S$ . Therefore, by efficiency and (5),  $\varphi(R'_1, R_{-1}) = (x, N)$ . 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  $\varphi$  is not Nash implementable.

■

Before finishing this section, three remarks are in order.

**Remark 1** A natural question is to ask if the non-resolute allocation rule  $Z$  which, at each preference profile  $R$ , selects the set of stable and efficient allocations  $Z(R)$  is Maskin monotonic.<sup>5</sup> The answer is negative and follows from the observations that in the proof of Proposition 2,  $Z(R) = \{(y, N), (x, N)\}$ ,  $Z(R'_1, R_{-1}) = \{(x, N)\}$  (i.e.,  $(y, N) \in Z(R)$  and  $(y, N) \notin Z(R'_1, R_{-1})$ ), and for all  $(z, S) \in A$  such that  $(y, N)R_1(z, S)$  it holds that, by condition (1) in the definition of  $R'_1$ ,  $(y, N)R'_1(z, S)$ .

**Remark 2** 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 anonymous domain of gregarious and single-

---

<sup>5</sup>A social choice correspondence  $\psi : \mathcal{R} \rightarrow A$  is *Maskin monotonic* if for all  $R, R' \in \mathcal{R}$  and all  $a \in \psi(R)$  such that  $a \notin \psi(R')$  there exist  $i \in N$  and  $b \in A$  such that  $aR_i b$  and  $bP'_i a$ .

peaked preference relations.<sup>6</sup> 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 are anonymous and might be single-peaked. Hence, the proof of Proposition 2 shows that any efficient and stable social choice function defined on the anonymous domain of gregarious *and* single-peaked preference relations is not Nash implementable.

**Remark 3** All the results contained in this section, namely Lemmata 1 and 2 and Propositions 1 and 2, have been proved without using (STRICT), and therefore they hold without assuming that preferences satisfy this property. In particular, in the proof of the impossibility result of Proposition 2 we use profiles  $R$  and  $(R'_1, R_{-1})$  with strict preferences. If we enlarge the domain of profiles by admitting also preference relations satisfying (GREG) and (APA) but not necessarily (STRICT), the impossibility result still holds since the profiles  $R$  and  $(R'_1, R_{-1})$  remain in the enlarged domain. However, (STRICT) is a necessary property in order to obtain positive subgame-perfect implementation results, as Example 5 in Subsection 4.2 will show.

## 3 The Implementation

### 3.1 SPNE Implementation: Preliminaries

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 of whether or not it is possible to implement some of them as Subgame Perfect Nash

---

<sup>6</sup>A social choice function  $\varphi : \mathcal{R} \rightarrow X \times 2^N$  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  $<$ . 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  $i \in S$ .

Equilibria (SPNE) of a game in extensive form. Our aim is not only to prove that an efficient and stable social choice function can be implemented in our preference domain, but also to show that it can be implemented by means of a simple mechanism. Hence, we do not look at those general games proposed by Moore and Repullo (1988), Abreu and Sen (1990), and Vartiainen (2005 and 2006), who find general results for SPNE implementation, but we directly propose a mechanism that can be easily implemented in real collective decision problems within our framework.

However, it is interesting to note that the no veto power property (that together with Condition  $\alpha$  and  $n \geq 3$  is one of the sufficient conditions for SPNE implementation in Moore and Repullo (1988) and Abreu and Sen (1990)) does not hold in our setting. A social choice function  $\varphi : \mathcal{R} \rightarrow X \times 2^N$  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  $\varphi(R) = (x, S)$ . It is easy to see 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.

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 Jackson and Nicolò (2004) and Bogomolnaia and Nicolò (2005)) 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. Observe that 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.

## 3.2 Maximal Participation

Among the set of efficient and stable allocations we will be specially interested in 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 efficient and stable set with maximal participation. In our setting there is a minimum level of welfare that any agent  $i$  can secure; this 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.

Since the maximal participation set  $MP(R)$  might have several allocations we need a selection rule on  $A$ . Since  $(x, S), (y, T) \in MP(R)$  and  $(x, S) \neq (y, T)$  imply  $x \neq y$ , it is enough to have a rule to select an alternative from each subset of alternatives. Let  $\succ$  be a linear order (complete, transitive and antisymmetric binary relation) on  $X$ . Given a non-empty subset  $C \in 2^X \setminus \{\emptyset\}$ , we denote by  $\succ(C)$  the maximizer of  $\succ$  on  $C$ ; namely,  $\succ(C) = x$ , where  $x \in C$  and for all  $y \in C$ ,  $y \neq x$ ,  $x \succ y$ . We set  $\succ(\emptyset) = \succ(X)$ .

## 3.3 The Mechanism and the Main Result

The mechanism we propose is a two-stage game. In the first stage agents make a proposal sequentially, according to the order  $1, \dots, n$ , and publicly.<sup>7</sup> A proposal consists of an alternative and a positive integer (smaller or equal to  $n$ ) that indicates the cardinality of the group of users according to the proposer's view. Agents can

---

<sup>7</sup>We will later argue that the unique SPNE outcome of the game is invariant with respect to the order in which players move along the game.

make a new proposal only if they announce either a new alternative with an equal or greater cardinality of the group of users or the same alternative with a strictly greater cardinality. They can also pass the turn and not make any proposal. The order is iterated until either all agents have decided to not make a new proposal, or there are no more feasible proposals left.<sup>8</sup> Each proposer has to burden a cost if none of her proposals has been selected. The alternative with the largest group of users is selected (in case there are more than one, the chosen alternative is the maximizer of  $\succ$  on this set). In the second stage, agents sequentially announce (knowing the outcome of the first stage) whether they want to be users of the selected alternative or not. If the number of agents who accept to be users is equal or higher than the number announced by the agent who made the proposal in the first stage, then the final allocation is the selected alternative and the set of agents who accepted to be users in the second stage. Otherwise, the final allocation is  $(\succ(\emptyset), \emptyset)$ . If the agent who made the accepted proposal at Stage 1 is not a final user, he burdens a (larger) cost. We now define the mechanism formally.

- Stage 1:

- Step 1: agent 1 proposes either  $p_1 = (x_1, k_1) \in X \times \{1, \dots, n\} \equiv \bar{A}$  or does not propose anything (identified as the proposal  $p_1 = (NP, 0)$ ).

Assume that  $m$  proposals  $p_1, \dots, p_m$  have already been made. Define

$$A(p_1, \dots, p_m) = \{p_q \mid p_q = (x_q, k_q) \in \bar{A} \text{ for some } 1 \leq q \leq m\}.$$

- Step  $m+1$ : agent  $i = (m+1) \bmod n$  proposes either  $p_{m+1} = (x_{m+1}, k_{m+1}) \in \bar{A} \setminus A(p_1, \dots, p_m)$  such that  $k_{m+1} \geq \max\{k_1, \dots, k_m\}$  or does not propose any-

---

<sup>8</sup>To iterate the order  $1, \dots, n$  we proceed as follows. It is easy to check that each positive 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 as  $r \equiv m \bmod n$ . Then, agent  $r$  (where  $r$  is between 1 and  $n$ ) is the agent who has to make a proposal in step  $m$ .

thing  $(p_{m+1} = (NP, 0))$ .<sup>9</sup>

If after the first  $n$  steps all agents proposed  $(NP, 0)$  then the game ends with the outcome  $(\succ (\emptyset), \emptyset)$ . Otherwise, let  $m > 1$  be the step where  $p_m = (x_m, k_m) \in \bar{A}$  and  $p_{m+1} = \dots = p_{m+n-1} = (NP, 0)$ . Then, Stage 1 finishes at step  $m + n - 1$ .

Given  $p_1, \dots, p_m$ , define

$$\hat{x} = \succ (\{x \in X \mid \exists 1 \leq q \leq m \text{ s.t. } p_q = (x, k) \text{ and } k = k_m\}).$$

Set  $\hat{k} = k_m$  and  $\hat{i} = q \pmod{n}$  where  $1 \leq q \leq m$  is such that  $p_q = (\hat{x}, \hat{k})$ . Then, the outcome of Stage 1 is  $(\hat{x}, \hat{k}, \hat{i}) \in \bar{A} \times N$ ; namely, a proposal  $(\hat{x}, \hat{k})$  and its proposer  $\hat{i}$ .

Each proposer has to burden an  $\varepsilon'$ -cost if none of her proposals is the selected one at Stage 1,  $(\hat{x}, \hat{k})$ .<sup>10</sup>

- Stage 2: Each agent  $j$ , knowing the outcome  $(\hat{x}, \hat{k}, \hat{i}) \in \bar{A} \times N$  of Stage 1 and the decision of  $j$ 's predecessors in Stage 2, announces sequentially (following the same order  $1, \dots, n$ ) whether he/she 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  $\hat{i}$ , who made the proposal  $(\hat{x}, \hat{k})$  in Stage 1, has to burden a  $\varepsilon''$ -cost, worse than the  $\varepsilon'$ -cost of Stage 1, if he is not a user of  $\hat{x}$ ; *i.e.*, either  $(\hat{x}, \emptyset)$  is selected, and/or  $\hat{i}$  announced  $nu$ .

Before stating the main result of the paper, three general comments on the mechanism are appropriate. First, the equilibrium outcome of the game is invariant with

---

<sup>9</sup>Observe that, at Step  $m + 1$ , if  $(x, k)$  has already been proposed at earlier steps and  $k = k_m$ , the current proposer can propose  $(z, k)$  if it has not been proposed yet and  $z \neq x$ , but he/she can not propose again  $(x, k)$ .

<sup>10</sup>See Subsection 4.1 for a discussion on the  $\varepsilon$ -costs.

respect to the order in which players move along the game (see Corollary 2 in Subsection 5.4). Second, stability imposes a strong restriction on the set of feasible mechanisms. In ours, Stage 2 makes sure that the outcome is a stable allocation. Stage 1, then, has to be designed in order to satisfy maximal participation and efficiency in the selection of the alternative. Third, the use of a linear order  $\succ$  as a tie-breaking rule is necessary to guarantee that the implemented (in SPNE) social choice function  $\varphi : \mathcal{R} \rightarrow A$  treats agents symmetrically, but it makes the social choice function not neutral. Neutral social choice functions could be easily constructed (as well the mechanisms to implement them). However a unique selection should be obtained by breaking the ties in the maximal participation set by means of an asymmetric treatment of the agents. Example 1 clarifies these points.

**Example 1** Let  $X = \{x, y\}$ ,  $N = \{1, 2\}$  and consider any two preference profiles  $R, R' \in \mathcal{R}$  such that<sup>11</sup>

$$\begin{aligned}
& (x, \{1, 2\}) P_1 (x, \{1\}) P_1 (x, \{\emptyset\}) P_1 (y, \{1, 2\}) \\
& (y, \{1, 2\}) P_2 (y, \{2\}) P_2 (y, \{\emptyset\}) P_2 (x, \{1, 2\}) \\
& (y, \{1, 2\}) P'_1 (y, \{1\}) P'_1 (y, \{\emptyset\}) P'_1 (x, \{1, 2\}) \\
& (x, \{1, 2\}) P'_2 (x, \{2\}) P'_2 (x, \{\emptyset\}) P'_2 (y, \{1, 2\}).
\end{aligned}$$

Assume  $x \succ y$ . The unique SPNE of our mechanism is  $(x, \{1\})$  at profile  $R$  (in Stage 1, agent 1 proposes  $(x, 1)$  and agent 2 does not propose anything), and  $(x, \{2\})$  at profile  $R'$  (in Stage 1, agent 1 does not propose anything and agent 2 proposes  $(x, 1)$ ). Observe that (i) in both cases to not propose  $(y, 1)$  is part of the equilibrium since  $x \succ y$  and the proposer would burden a cost for proposing  $(y, 1)$ ; if  $y \succ x$  the unique SPNE outcome would be  $(y, \{2\})$  at profile  $R$  and  $(y, \{1\})$  at profile  $R'$ ; (ii) the SPNE outcome in all the cases would be the same if agents play in the inverted order 2, 1; (iii) the social choice function that chooses  $(x, \{1\})$  at profile  $R$  and  $(y, \{1\})$  at profile  $R'$  does not treat agents symmetrically. This neutral social choice function

---

<sup>11</sup>We only list the relevant parts of the preference relations.

could be implemented by a slight modification of the mechanism in which in Stage 1 each agent makes a unique announcement according to the order 1, 2, and agent 2's proposal, if any, has to contain a positive integer (the cardinality of the group of users) strictly larger than the previously announced. Note that the fact that the social choice function does not treat agents symmetrically, is reflected by the undesirable feature that the SPNE outcome of the mechanism now depends on the order in which agents play.  $\square$

The mechanism defines a finite extensive-game form with perfect information. Denote it by  $\Gamma$ . Now, given a preference profile  $R \in \mathcal{R}$ , let  $\Gamma(R)$  be the finite game in extensive form with perfect information (the extensive-game form played by agents that evaluate outcomes according to the preference profile  $R$ ). The main result of the paper states that the game in extensive form  $\Gamma(R)$  has a unique SPNE outcome which is a stable and efficient allocation in the set with maximal participation. Formally,

**Theorem 1** *Let  $R \in \mathcal{R}$  be given. The game in extensive form  $\Gamma(R)$  has a unique SPNE outcome that belongs to the efficient and stable set with maximal participation.*

**Proof** See Subsection 5.3 in the Appendix at the end of the paper.

Before moving to the next section, we want to emphasize that no agent has to burden any cost in equilibrium (see Lemma 6 in Subsection 5.4).

## 4 Final Remarks

### 4.1 Ordinal Penalties

The only restriction we impose on the  $\varepsilon$ -costs we use as out-of-equilibrium threats is that the  $\varepsilon''$ -cost burdened at Stage 2 by the proposer who is not a final user be worse than the  $\varepsilon'$ -cost burdened by the proposers if none of their proposals has been accepted (see Example 4 below). Apart from this restriction, no other one is imposed; in particular, to be consistent with our ordinal setting, they may be non-transferable.

But, if we embed the ordinal setting into a cardinal one, these costs can be interpreted as monetary fines (potentially, infinitely small). They are only used in the proof of our main result to take away from agents the unsubstantial 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.

In our proofs we assume that these costs never induce a preference reversal between two allocations and therefore, the negative result for Nash implementation still holds. We have chosen to avoid a formal treatment of these problems in the previous sections. This would require either to have a cardinal setting or to deal explicitly with these  $\varepsilon$ -costs in the ordinal setting. We provide here a sketch of how to deal with them when preferences are ordinal.

Let  $N = \{1, \dots, n\}$  be the set of agents,  $X$  be the finite set of levels of a public good (or social alternatives), and  $E = \{\emptyset, \varepsilon', \varepsilon''\}$  be a set of three levels of a private object, that can be interpreted as a fine. An allocation is a triple  $(x, S, e) \in \tilde{A} \equiv X \times 2^N \times E^n$ , where  $x \in X$  is the level of the public good,  $S \in 2^N$  is the subset of its users and  $e = (e_1, \dots, e_n) \in E^n$  is a  $n$ -dimensional vector where  $e_i = \varepsilon \in \{\varepsilon', \varepsilon''\}$  indicates that agent  $i$  has to burden the corresponding  $\varepsilon$ -cost and  $e_i = \emptyset$  indicates that he does not. The preference relation of agent  $i \in N$  over the set of allocations  $\tilde{A}$ , denoted by  $\tilde{R}_i$ , is a complete, reflexive and transitive binary relation, which satisfies the following additional restrictions.

(PC) PRIVATE COST: For all  $x \in X$ ,  $S \in 2^N$ , and  $e \in E^n$ ,

$$(x, S, (e_{-i}, \emptyset)) \tilde{P}_i (x, S, (e_{-i}, \varepsilon')) \tilde{P}_i (x, S, (e_{-i}, \varepsilon'')).$$

(NPR) NO PREFERENCE REVERSAL: For all  $x, y \in X$  and  $S, T \in 2^N$ , if  $(x, S, e) \tilde{P}_i (y, T, e)$  for some  $e \in E^n$ , then  $(x, S, e) \tilde{P}_i (y, T, e')$  for all  $e' \in E^n$ .

Assume that  $\tilde{R} = (\tilde{R}_1, \dots, \tilde{R}_n)$  satisfies (PC) and (NPR) for all  $i \in N$ . Then, we can define a restricted allocation set,  $A \equiv X \times 2^N$ , and focus, without loss of

generality, on the restricted preference profile  $R = (R_1, \dots, R_n) \in \mathcal{R}$ , where for each  $i \in N$ ,  $R_i$  is the preference relation on the set  $A$  induced (in the obvious way) by  $\tilde{R}_i$ .

Condition (NPR) is a reasonable assumption because the set of levels of the public good  $X$  is finite. From the implementation point of view it means that the mechanism designer, who knows the set of gregarious preference relations  $\mathcal{R}_i$  over  $A = X \times 2^N$  for each  $i \in N$ , is able to find two private objects  $\varepsilon'$  and  $\varepsilon''$  (which are not necessarily a monetary fine) guaranteeing that (NPR) holds for all agents' preference relations.

## 4.2 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 remaining  $n - 1$  agents do not make new proposals (if  $j$  reacts to  $i$ 's proposal, for  $i \neq j$ ,  $i$  should still be able to counteract). Second, proposers have to burden a  $\varepsilon'$ -cost in the case that none of their proposals has been selected at Stage 1, and the proposer  $\hat{i}$  of the chosen proposal at Stage 1 burdens a  $\varepsilon''$ -cost if he is not a final user.<sup>12</sup> 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 these features.

**Example 2** (The order of proposals in Stage 1 is not iterated) Let  $X = \{x, y, z\}$ ,  $N = \{1, 2\}$ , and consider the linear order  $x \succ y \succ z$ . 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  $Z(R) = \{(z, \{1\}), (y, \{2\})\}$ . It is easy to check that the unique SPNE outcome of our game is the allocation  $(z, \{1\})$  and that the same result holds if the

---

<sup>12</sup>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 (2006) for awards).

order in which players move is inverted.

Consider our mechanism, except that Stage 1 is played without iteration. The unique SPNE outcome is the inefficient allocation  $(x, \{1\})$ . Fix now the order 2, 1. Then, the unique SPNE outcome of the game without iterating the order 2, 1 in Stage 1 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.  $\square$

**Example 3** (To make a proposal is never costly) Let  $X = \{x, y, z\}$ ,  $N = \{1, 2\}$ , and consider again the linear order  $x \succ y \succ z$ . Take any  $R \in \mathcal{R}$  such that

$$(y, \{1\}) P_1 (z, \{\emptyset\}) I_1 (x, \{\emptyset\}) P_1 (z, \{1, 2\}) P_1 (x, \{1, 2\})$$

and

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

Observe that  $Z(R) = \{(y, \{1\}), (z, \{2\})\}$ . The allocation  $(z, \{2\})$  is a SPNE outcome of the game since there exists a SPNE in which agent 1 does not propose anything in Stage 1 and agent 2 announces  $(z, 1)$  at step 2 of Stage 1. However, now, the inefficient allocation  $(x, \{2\})$  is also a SPNE outcome of the game since there exists another SPNE in which agent 1 first announces  $(y, 1)$  and then agent 2 announces  $(x, 1)$ .  $\square$

**Example 4** (The proposer who is not a final user does not burden a  $\varepsilon''$ -cost or the  $\varepsilon''$ -cost is not worse than the  $\varepsilon'$ -cost) Let  $X = \{x, y, z\}$  and  $N = \{1, 2, 3\}$  and consider the linear order  $x \succ y \succ z$ . Consider the preference profile  $R = (R_1, R_2, R_3) \in \mathcal{R}$  where

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

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

and

$$(y, \{3\}) P_3 (x, \{3\}) P_3 (z, \{\emptyset\}) P_3 (z, \{N\}).$$

The SPNE of the game we propose is the allocation  $(y, \{3\})$ . To make the argument more transparent assume that the  $\varepsilon$ -costs are cardinal. Suppose  $\varepsilon'' = 0$ ; that is, the proposer at Stage 1 who is not a final user does not burden any cost. Then, there is a SPNE in which agent 1 proposes  $(x, 1)$  in Stage 1, no other agent proposes anything else (since  $\succ(C) = x$  if  $x \in C$ ). Therefore, the final SPNE outcome is the inefficient allocation  $(x, \{3\})$ .

Consider now the case  $0 < \varepsilon'' < \varepsilon'$ . Then, the SPNE outcome is the allocation  $(z, \{2\})$ . In fact, agent 2 can play a strategy with the following features:

$$f_2(p_1) = \begin{cases} (z, 1) & \text{if } p_1 = (NP, 0) \\ (NP, 0) & \text{otherwise,} \end{cases}$$

and

$$f_5(p_1, p_2, p_3, p_4) = \begin{cases} (x, 3) & \text{if } p_2 = (z, 1) \text{ and there exists } 2 < m \leq 4 \text{ such that} \\ & p_m \neq (NP, 0) \text{ and } (x, 3) \in \overline{A} \setminus \{A(p_1, p_2, p_3, p_4)\} \\ (NP, 0) & \text{otherwise.} \end{cases}$$

Note that to play  $p_5 = (x, 3)$ , under the specific circumstances, is a credible threat that prevents player 3 to propose  $(y, 1)$  at step 3 because  $\varepsilon'' < \varepsilon'$ : once agent 3 proposes  $(y, 1)$  agent 2 prefers to propose  $(x, 3)$  than proposing  $(NP, 0)$ , since he has to burden the  $\varepsilon''$ -cost instead of the  $\varepsilon'$ -cost. It is easy to see that the SPNE outcome in this case also depends on the exogenously given order; if agents play according to the order 3,2,1, the SPNE outcome is the allocation  $(y, \{3\})$ .  $\square$

Our last example shows why (STRICT) is necessary in order to obtain positive subgame-perfect implementation results.

**Example 5** (Agents' preference relations admit indifferences) Let  $X = \{x, y\}$ ,  $N = \{1, 2\}$ , and consider the linear order  $x \succ y$ . Take any  $R \in \mathcal{R}$  such that

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

and

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

The unique efficient and stable allocation is  $(x, \{1, 2\})$ , but the allocation  $(x, \{1\})$  is a SPNE outcome. In fact, there exists a SPNE strategy in which agent 1 proposes  $(x, 1)$  and at Stage 2, agent 2 chooses not to be a user in all information sets. Finally, note that an alternative to (STRICT), which restores the positive implementation result, is to assume that in case of indifference between the outcomes of two different actions, all agents always choose the action that induces the Pareto optimal outcome. But this is *ad hoc* since it has no individual strategic justification.  $\square$

### 4.3 Neutrality

The SPNE outcome of the game depends on the linear order  $\succ$  used to select a single alternative for each possible subset of alternatives. It is natural to ask whether it is possible to implement the social choice correspondence  $\psi : \mathcal{R} \rightarrow A$ , 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 linear orders. Add a preliminary stage in the extensive-game form in which all agents simultaneously announce some  $\succ \in \mathcal{H}$ . Given  $R \in \mathcal{R}$ , if at least  $n - 1$  agents announce the same  $\succ$  then they play the game  $\Gamma^\succ(R)$  with the linear order  $\succ$ ; otherwise, the game  $\Gamma^{\succ'}(R)$  is played with a pre-specified linear order  $\succ'$ . It is straightforward to check that, for all  $R \in \mathcal{R}$ , the set of SPNE outcomes of this enlarged game coincides with the set  $MP(R)$ .

### 4.4 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 (2006) for more on this interpretation). We now want to establish the relationship between the set of efficient and stable allocations under  $R$  and the set of group stable allocations under  $R$ . 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  $i \in T$ ,  $(x, S \setminus T)P_i(x, S)$ ;

(EXTERNAL GROUP STABILITY) There does not exist  $T \subseteq N \setminus S$  such that, for all  $i \in T$ ,  $(x, S \cup T)P_i(x, S)$ .

**Lemma 3** *Let  $R \in \mathcal{R}$  be given. An allocation  $(x, S)$  is group stable under  $R$  if and only if it is stable under  $R$  and efficient relative to  $x$ .*

**Proof** Let  $(x, S_x)$  be stable under  $R$  and efficient relative to  $x$ . Assume that there exists  $T \subseteq N \setminus S_x$  such that, for all  $i \in T$ ,  $(x, S_x \cup T)P_i(x, S_x)$ . Thus,  $T \neq \emptyset$ . By (GREG), for all  $i \in S_x$ ,  $(x, S_x \cup T)P_i(x, S_x)$ . Hence, for all  $i \in S_x \cup T$ ,  $(x, S_x \cup T)P_i(x, S_x)$ . This contradicts Lemma 1. By (GREG), and because, by internal stability,  $(x, S_x)P_i(x, \emptyset)$  for all  $i \in S_x$ , group internal stability of  $(x, S_x)$  follows.

Let  $(x, S)$  be group stable under  $R$ . By definition,  $(x, S)$  is stable under  $R$ . If  $S \neq S_x$ , by Lemma 1,  $\#S < \#S_x$ . For all  $i \in S_x \setminus S \equiv T$ ,  $(x, S \cup T)P_i(x, S)$ . Hence, external group stability is violated. Thus,  $S = S_x$  and  $(x, S)$  is stable and efficient relative to  $x$ . ■

## 5 Appendix

In the first two subsections of this Appendix we describe formally strategies and outcome functions of the game in extensive form  $\Gamma(R)$ . Subsection 5.3 contains the proof of Theorem 1 as well as its Corollary stating that the order in which agents take decisions along the game does not affect its SPNE outcome. Lemma 6, in Subsection 5.4, states that no agent has to burden any  $\varepsilon$ -cost in equilibrium.

### 5.1 Strategies

A *consumption strategy* of agent  $i$  in Stage 1 is a choice of a feasible proposal at each of his information sets. Since the order in which agents make proposals along the

game and proposals are public, each information set consists of a unique node that can be identified with the history of earlier proposals. For  $m \geq 1$ , let  $h^m = (p_1, \dots, p_m)$  be a history of  $m$  proposals and let  $H^m$  be the set of all possible histories of length  $m$ . Set  $h^0 = \emptyset$  and  $H^0 = \{h^0\}$ . Observe that agent  $i$  plays at step  $i$ , given a history of length  $i - 1$ , and agent  $i$  may have to play at steps  $k \cdot n + i$ , for  $1 \leq k \leq K$ ,<sup>13</sup> given a history of length  $k \cdot n + i - 1$ . For each  $m \geq 1$  and  $h^m = (p_1, \dots, p_m) \in H^m$ , define

$$P(h^m) = \{p_{m+1} = (x_{m+1}, k_{m+1}) \in \bar{A} \setminus A(h^m) \mid k_{m+1} \geq k_m\} \cup \{(NP, 0)\}$$

as the set of feasible choices at step  $m + 1$  available to agent  $(m + 1) \bmod n$  given the history  $h^m$  (see Subsection 3.3 for the definitions of  $\bar{A}$  and  $A(h^m)$ ). Hence, a consumption strategy  $f_i$  for agent  $i$  is a feasible choice for each possible history; namely,

$$f_i : \bigcup_{k=0}^K H^{k \cdot n + i - 1} \rightarrow \bar{A} \cup \{(NP, 0)\},$$

with the property that for all  $h^m \in \bigcup_{k=0}^K H^{k \cdot n + i - 1}$ ,  $f_i(h^m) \in P(h^m)$ . Let  $F_i$  be the set of consumption strategies of agent  $i$  in Stage 1.

Assume that the outcome of Stage 1 is  $(\hat{x}, \hat{k}, \hat{i}) \in \bar{A} \times N$ .<sup>14</sup> In Stage 2, and after knowing  $(\hat{x}, \hat{k}, \hat{i})$ , agents decide sequentially whether or not they would like to use the public good at level  $\hat{x}$  with at least  $\hat{k}$  users. The set of *participation strategies* of agent  $i$  at the subgame starting at  $(\hat{x}, \hat{k}, \hat{i}) \in \bar{A} \times N$ ,  $\Gamma(\hat{x}, \hat{k}, \hat{i})$ , is the set of functions

$$B_i(\hat{x}, \hat{k}, \hat{i}) = \left\{ b_i[\hat{x}, \hat{k}, \hat{i}] : 2^{\{j \in N \mid j < i\}} \rightarrow \{u, nu\} \right\},$$

where  $b_i[\hat{x}, \hat{k}, \hat{i}](S)$  specifies agent  $i$ 's willingness to use the public good at level  $\hat{x}$  (when  $\hat{i}$  made the proposal  $(\hat{x}, \hat{k})$  and  $\hat{k}$  users are necessary), given that the set  $S \in 2^{\{j \in N \mid j < i\}}$  of agents have already announced that they are willing to do so, and the set  $\{j \in N \mid j < i\} \setminus S$  of agents have announced that they are not. Let  $B_i =$

<sup>13</sup> $K \equiv n \cdot \#X$  is the maximal number of times that a player may have to play.

<sup>14</sup>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  $(\succ(\emptyset), \emptyset)$ .

$\bigcup_{(\hat{x}, \hat{k}, \hat{i}) \in \bar{A} \times N} B_i(\hat{x}, \hat{k}, \hat{i})$  denote the set of participation strategies of agent  $i$  in Stage 2 and let  $b_i$  be a generic element of this set. Let  $G_i = F_i \times B_i$  denote the set of strategies of agent  $i$ . A strategy profile  $g = (f, b) \in F \times B$  is a  $n$ -tuple of strategies, where  $F = F_1 \times \cdots \times F_n$  and  $B = B_1 \times \cdots \times B_n$ . Let  $G = G_1 \times \cdots \times G_n$  be the set of strategy profiles.

## 5.2 Outcome Functions

Given a consumption strategy profile  $f \in F$ , let  $p_m(f)$  be the proposal made by agent  $m \pmod n$  according to  $f$  at step  $m$  of Stage 1, and let  $M$  be the last step of Stage 1. Let  $h(f) = (p_1(f), \dots, p_M(f))$  be the history generated by  $f$ . The *outcome* of Stage 1 is

$$o_1(f) = \begin{cases} (\succ(\emptyset), \emptyset) & \text{if } h(f) = \underbrace{((NP, 0), \dots, (NP, 0))}_{n\text{-times}} \\ (\hat{x}, \hat{k}, \hat{i}) & \text{otherwise,} \end{cases}$$

where  $(\hat{x}, \hat{k}, \hat{i})$  is defined in the obvious (but tedious) way. Given a consumption strategy profile  $f \in F$  define, for each agent  $i \in N$ ,

$$\varepsilon_1(f)_i = \begin{cases} \varepsilon' & \text{if } \exists 1 \leq m \leq M \text{ s.t. } i = m \pmod n, p_m(f) \neq (NP, 0), \text{ and } i \neq \hat{i} \\ 0 & \text{otherwise;} \end{cases}$$

namely,  $\varepsilon_1(f)_i = \varepsilon'$  means that agent  $i$  has made some proposal and none of them has been selected, and therefore  $i$  has to burden the  $\varepsilon'$ -cost.

Let  $(\hat{x}, \hat{k}, \hat{i}) \in \bar{A} \times N$  be the outcome of Stage 1. Given the participation strategy profile  $b \in B$ , define recursively (in the obvious and tedious way) the indicator function of the decision of agent  $j$  in the subgame starting at  $(\hat{x}, \hat{k}, \hat{i})$  as

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

Let  $S(b[\hat{x}, \hat{k}, \hat{i}]) \equiv \{j \in N \mid \pi_j(b[\hat{x}, \hat{k}, \hat{i}]) = 1\}$  be the set of agents that announced their willingness to be a user along the play generated by the participation strategy profile

$b[\hat{x}, \hat{k}, \hat{i}]$ . Then, the outcome of Stage 2 starting at  $(\hat{x}, \hat{k}, \hat{i})$  generated by  $b[\hat{x}, \hat{k}, \hat{i}]$  is

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

that is, the set of final users is the set of agents who announced  $u$ ,  $S(b[\hat{x}, \hat{k}, \hat{i}])$ , as long as its cardinality is larger or equal than  $\hat{k}$ ; otherwise, no agent becomes a user.

Moreover, define

$$\varepsilon_2(b[\hat{x}, \hat{k}, \hat{i}]) = \begin{cases} \varepsilon'' & \text{if either } o_2(b[\hat{x}, \hat{k}, \hat{i}]) = (\hat{x}, \emptyset) \text{ or } \pi_{\hat{i}}(b[\hat{x}, \hat{k}, \hat{i}]) = 0 \\ 0 & \text{otherwise;} \end{cases}$$

namely,  $\varepsilon_2(b[\hat{x}, \hat{k}, \hat{i}]) = \varepsilon''$  indicates that agent  $\hat{i}$  (who proposed  $(\hat{x}, \hat{k})$ ) is not a final user of the public good, and therefore  $\hat{i}$  has to burden the  $\varepsilon''$ -cost.

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

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

Additionally, to keep track of who has to burden a  $\varepsilon$ -cost, given a strategy profile  $(f, b) \in G$ , define, for each  $i \in N$ ,

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

### 5.3 Proof of Theorem 1

Let  $R \in \mathcal{R}$  be given. The game  $\Gamma(R)$  is a finite game in extensive form with perfect information. Hence, it has at least a SPNE in pure strategies.

First, we prove that for any outcome  $(\hat{x}, \hat{k}, \hat{i})$  of Stage 1, the subgame  $\Gamma(\hat{x}, \hat{k}, \hat{i})$  has a unique SPNE outcome which can be easily identified (Lemma 4). Second, we prove Theorem 1 by arguing that the game  $\Gamma(R)$  has a unique SPNE outcome and by showing that it satisfies the desirable properties; *i.e.*, if  $(f, b)$  is a SPNE of  $\Gamma(R)$  then,  $o(f, b) \in MP(R)$ .

**Lemma 4** Let  $R \in \mathcal{R}$  be given. For each outcome  $(\hat{x}, \hat{k}, \hat{i}) \in \bar{A} \times N$  of Stage 1, the subgame  $\Gamma(\hat{x}, \hat{k}, \hat{i})$  has a unique SPNE outcome. Moreover, for every SPNE participation strategy profile  $b[\hat{x}, \hat{k}, \hat{i}]$  of  $\Gamma(\hat{x}, \hat{k}, \hat{i})$ ,

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

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

**Proof** Let  $b[\hat{x}, \hat{k}, \hat{i}]$  be a SPNE of  $\Gamma(\hat{x}, \hat{k}, \hat{i})$ . Consider agent  $n$  and let  $T \in 2^{\{j \in N | j < n\}}$  be an arbitrary information set of agent  $n$ . By (STRICT), agent  $n$  orders strictly the two allocations  $(\hat{x}, \emptyset)$  and  $(\hat{x}, T \cup \{n\})$ . We distinguish between two cases.

Case 1:  $\#(T \cup \{n\}) \geq \hat{k}$ .

If  $(\hat{x}, \emptyset) P_n(\hat{x}, T \cup \{n\})$  then  $b_n[\hat{x}, \hat{k}, \hat{i}](T) = nu$  and the SPNE outcome of the subgame starting at  $T$  is equal to  $(\hat{x}, \emptyset)$  if  $\#T < \hat{k}$  or equal to  $(\hat{x}, T)$  if  $\#T \geq \hat{k}$ .

If  $(\hat{x}, T \cup \{n\}) P_n(\hat{x}, \emptyset)$  then  $b_n[\hat{x}, \hat{k}, \hat{i}](T) = u$  and the SPNE outcome of the subgame starting at  $T$  is equal to  $(\hat{x}, T \cup \{n\})$ .

Case 2:  $\#\{T \cup \{n\}\} < \hat{k}$ .

Then, any SPNE outcome of the subgame starting at  $T$  is equal to  $(\hat{x}, \emptyset)$ .

Thus, by the backwards induction principle, we can replace the information set  $T$  of agent  $n$  by the unique outcome previously identified. Following the induction argument we obtain the uniqueness of the SPNE outcome of the subgame  $\Gamma(\hat{x}, \hat{k}, \hat{i})$ . Moreover, it is straightforward to check that for all  $i \in N$ ,  $\pi_i(b[\hat{x}, \hat{k}, \hat{i}]) = 1$  if and only if  $i \in S_{\hat{x}}$ . Therefore,  $o_2(b[\hat{x}, \hat{k}, \hat{i}])$  has the desired property.  $\blacksquare$

**Proof of Theorem 1** Let  $(f, b)$  be a SPNE of  $\Gamma(R)$  and let  $o(f, b) \equiv (x, S) \in A$  be its outcome. We first state two claims without proof.<sup>15</sup>

CLAIM 1 If  $i \notin S$  then  $\varepsilon_i(f, b) = 0$ .

<sup>15</sup>CLAIMS 1 and 2 are immediately implied by the fact that  $(f, b)$  is a SPNE of  $\Gamma(R)$ . In particular, they hold because each agent  $i$  has always available the strategy consisting of never making a proposal in Stage 1 and declaring  $nu$  in Stage 2, regardless of the other agents' behavior.

CLAIM 2 For all  $i \in N$ ,  $(x, S)R_i(x, \emptyset)$ .

Define  $ES(R) = \{(y, S_y) \in A \mid (y, S_y) \text{ is stable and efficient relative to } y\}$  and

$MES(R) = \{(y, S_y) \in ES(R) \mid \#S_y \geq \#S_z \text{ for all } z \in X \text{ such that } (z, S_z) \in ES(R)\}$ .

Observe that  $MP(R) \subseteq MES(R)$  for all  $R \in \mathcal{R}$ . Let  $L = \#S_y$ , where  $y$  is any alternative such that  $(y, S_y) \in MES(R)$ . Assume  $R$  is such that  $L = 0$ . Then, the set of efficient and stable allocations with maximal participation is the set  $MP(R) = MES(R) = \{(y', S') \in A \mid y' \in X \text{ and } S' = \emptyset\}$ . By CLAIMS 1 and 2,  $(x, S) = (\succ (\emptyset), \emptyset)$ . Thus,  $o(f, b) = (x, S) \in MP(R)$ . Thus, the statement of Theorem 1 holds when  $L = 0$ . In the rest of the proof we will assume that  $R$  is such that  $L \geq 1$ . We first show that  $(x, S) \in MES(R)$ . Assume otherwise; *i.e.*,

$$(x, S) \notin MES(R). \quad (6)$$

Let  $y = \succ (\{z \in X \mid (z, S_z) \in MES(R)\})$ . We distinguish between two cases.

Case 1: Suppose  $o_1(f) = (\succ (\emptyset), \emptyset)$ ; namely,  $h(f) = \underbrace{((NP, 0), \dots, (NP, 0))}_{n\text{-times}}$ . Since

$L \geq 1$ , there exists  $i \in S_y$  such that  $(y, S_y)P_i(\succ (\emptyset), \emptyset)$ . Consider agent  $i$ 's deviation  $\tilde{f}_i$  from  $f$  such that  $\tilde{f}_i$  is equal to  $f_i$  in all histories except in the history  $h^{i-1} = \underbrace{((NP, 0), \dots, (NP, 0))}_{(i-1)\text{-times}}$ , where  $\tilde{f}_i(h^{i-1}) = (y, \#S_y)$  and  $\tilde{f}_i(h^{m'-1}) = (NP, 0)$  for all

$m' > i$  with  $i = m'[\text{mod } n]$ . If  $o((\tilde{f}_i, f_{-i}), b) = (y, S_y)$  then  $\tilde{f}_i$  is a profitable deviation, contradicting that  $(f, b)$  is a SPNE of  $\Gamma(R)$ . Hence,  $o((\tilde{f}_i, f_{-i}), b) \neq (y, S_y)$ . By definition of  $y$  and Lemma 4,  $o((\tilde{f}_i, f_{-i}), b) = (z, \emptyset)$ . Thus, there exist agent  $j \neq i$  and step  $m > i$  with the properties that  $j = m[\text{mod } n]$ ,  $p_m(\tilde{f}_i, f_{-i}) \neq (NP, 0)$  and  $p_{m+1}(\tilde{f}_i, f_{-i}) = \dots = p_{m+n-1}(\tilde{f}_i, f_{-i}) = (NP, 0)$ . Hence,  $\varepsilon_j((\tilde{f}_i, f_{-i}), b) = \varepsilon''$ , since  $j$  made the final proposal and  $j$  is not a user in the allocation  $(z, \emptyset)$ . Thus,  $j$  is not best replying in the history  $h^{m-1} = (p_1(\tilde{f}_i, f_{-i}), \dots, p_{m-1}(\tilde{f}_i, f_{-i}))$ , contradicting that  $f$  is a SPNE of  $\Gamma(R)$ .

Case 2: Suppose  $o_1(f) = (x, q, j)$ . By Lemma 4 and CLAIM 1,  $o(f, b) = (x, S_x)$  with  $\#S_x \geq 1$ . Our contradiction hypothesis (6) implies  $\#S_x < \#S_y$ . Then, there

exists  $i \in S_y \setminus S_x$  such that  $(y, S_y)P_i(x, S_x)$ . Let  $m$  be the step such that  $p_{m+1}(f) = \dots = p_{m+n-1}(f) = (NP, 0)$  and let  $m'$  the last step in which agent  $i$  plays at Stage 1; *i.e.*,  $m \leq m' \leq m + n - 1$  and  $i = m'[\text{mod } n]$ . Consider agent  $i$ 's deviation  $\tilde{f}_i$  from  $f$  such that  $\tilde{f}_i$  is equal to  $f_i$  in all histories except in the history  $\tilde{h}^{m'-1}$ , where  $\tilde{f}_i(\tilde{h}^{m'-1}) = (y, \#S_y)$  and  $\tilde{f}_i(\tilde{h}^{m''-1}) = (NP; 0)$  for all  $m'' > m'$  with  $i = m''[\text{mod } n]$ . If  $o((\tilde{f}_i, f_{-i}), b) = (y, S_y)$  then  $\tilde{f}_i$  is a profitable deviation, contradicting that  $(f, b)$  is a SPNE of  $\Gamma(R)$ . If  $o((\tilde{f}_i, f_{-i}), b) \neq (y, S_y)$ , then by Lemma 4,  $o((\tilde{f}_i, f_{-i}), b) = (z, \emptyset)$  for some  $z \in X$ . Hence, there exists some agent  $j \neq i$  and step  $\bar{m} > m'$  with the properties that  $j = \bar{m}[\text{mod } n]$ ,  $p_{\bar{m}}(\tilde{f}_i, f_{-i}) \neq (NP, 0)$  and  $p_{\bar{m}}(\tilde{f}_i, f_{-i}) = \dots = p_{\bar{m}+n-1}(\tilde{f}_i, f_{-i}) = (NP, 0)$ . Thus,  $\varepsilon_j((\tilde{f}_i, f_{-i}), b) = \varepsilon''$ ; *i.e.*,  $j$  burdens the  $\varepsilon''$ -cost since he makes the final proposal at some step  $\bar{m} > m'$  and  $j$  is not a user in the allocation  $(z, \emptyset)$ . Since the  $\varepsilon''$ -cost is worse than the  $\varepsilon'$ -cost, even if agent  $j$  has made a proposal at some earlier step  $\hat{m} < m'$ , to make the winning proposal at step  $\bar{m}$  cannot be part of an equilibrium strategy. Hence,

$$o(f, b) \in MES(R). \quad (7)$$

It remains to be proven that  $o(f, b)$  is efficient under  $R$ , but before proceeding with its proof, and since preference relations satisfy (STRICT), we can apply the standard backwards induction argument to show that the game  $\Gamma(R)$  has a unique SPNE outcome.

**Lemma 5** *The game  $\Gamma(R)$  has a unique SPNE outcome.*

**Proof** Suppose there exist two different SPNE equilibrium outcomes  $(x, S_x) = o(f, b) \neq o(f', b') = (y, S_y)$ . Since  $\Gamma(R)$  is a finite game in extensive form with perfect information, the backwards induction principle and Lemma 4 imply that there exists  $i \in N$  and  $h^{m-1} = (p_1(f), \dots, p_{m-1}(f)) = (p_1(f'), \dots, p_{m-1}(f'))$ , where  $i = m[\text{mod } n]$ , such that

$$f_i(h^{m-1}) \neq f'_i(h^{m-1}) \quad (8)$$

and  $(x, S_x)I_i(y, S_y)$ . By (STRICT),  $i \notin S_x \cup S_y$ . By (8), either  $f_i(h^{m-1}) \neq (NP, 0)$  and  $\varepsilon_i(f, b) \neq 0$  or  $f'_i(h^{m-1}) \neq (NP, 0)$  and  $\varepsilon_i(f', b') \neq 0$ . This contradicts CLAIM 1. ■

CLAIM 3 below states that the unique SPNE outcome of  $\Gamma(R)$  can be obtained by an equilibrium participation strategy profile that generates in Stage 1 a very simple history.

CLAIM 3 *Let  $(f, b)$  be a SPNE of  $\Gamma(R)$  and let  $o(f, b) = (x, S_x)$  be its outcome. Then, there exists  $\hat{f} \in F$  such that  $(\hat{f}, b)$  is a SPNE of  $\Gamma(R)$  and*

$$h(\hat{f}) = \underbrace{((NP, 0), \dots, (NP, 0))}_{(i-1)\text{-times}}, \underbrace{(x, \#S_x)}_{\text{step } i}, \underbrace{((NP, 0), \dots, (NP, 0))}_{(n-1)\text{-times}},$$

where  $i \in S_x$  and for all  $j \in S_x \setminus \{i\}$ ,  $j > i$ .<sup>16</sup>

CLAIM 3 holds because if  $(\hat{f}, b)$  has a profitable deviation then  $(f, b)$  has also a profitable deviation. To proceed with the proof that the allocation  $o(f, b)$  is efficient under  $R$  describe the set  $MES(R)$  as  $\{(x_1, S_{x_1}), \dots, (x_Z, S_{x_Z})\}$ , where  $x_1 \succ \dots \succ x_Z$ . For all  $1 \leq q < q' \leq Z$ ,  $\#S_{x_q} = \#S_{x_{q'}}$  holds.

Assume  $(x_1, S_{x_1})$  is efficient under  $R$  and  $o(f, b) = (x_q, S_{x_q})$  holds for  $q > 1$ . Since  $(x_1, S_{x_1})$  is efficient under  $R$ , by (STRICT), there exists  $i \in S_{x_1} \setminus S_{x_q}$  such that  $(x_1, S_{x_1})P_i(x_q, S_{x_q})$ . Let  $m$  be the step such that  $p_{m+1}(f) = \dots = p_{m+n-1}(f) = (NP, 0)$  and let  $m \leq m' \leq m+n-1$  be such that  $i = m'[\text{mod } n]$ . Consider agent  $i$ 's deviation  $\tilde{f}_i$  from  $f$  such that  $\tilde{f}_i$  is equal to  $f_i$  in all histories except in the history  $h^{m'-1}$ , where  $\tilde{f}_i(h^{m'-1}) = (x_1, \#S_{x_1})$  and  $\tilde{f}_i(h^{m''-1}) = (NP, 0)$  for all  $m'' > m'$  with  $i = m''[\text{mod } n]$ . Using an argument similar to the one already used in Case 1, we obtain a contradiction with the fact that  $(f, b)$  is a SPNE of  $\Gamma(R)$ . Thus, either  $(x_1, S_{x_1})$  is efficient under  $R$  and  $o(f, b) = (x_1, S_{x_1})$ , in which case  $o(f, b) \in MP(R)$  and the statement of Theorem 1 follows, or  $(x_1, S_{x_1})$  is not efficient under  $R$ .

Assume  $(x_1, S_{x_1})$  is not efficient under  $R$ . Let  $T$  be the smallest integer such that  $1 \leq T \leq Z$  and  $(x_T, S_{x_T}) \in MES(R)$ ,  $(x_T, S_{x_T})PD(x_1, S_{x_1})$  and  $(x_T, S_{x_T})$  is efficient

---

<sup>16</sup>Obviously,  $o(\hat{f}, b) = (x, S_x)$ .

under  $R$ . By Lemma 2, there exists such  $T$ . Observe that, by its definition, (i) if  $(x_t, S_{x_t})$  is such that  $1 \leq t < T$  and  $(x_t, S_{x_t})PD(x_1, S_{x_1})$  then  $(x_T, S_{x_T})PD(x_t, S_{x_t})$ ; and (ii), by (STRICT), if  $(x_T, S_{x_T})PD(x_t, S_{x_t})$  then  $S_{x_t} = S_{x_T}$ . Hence, there exists a sequence  $(x_1, S_{x_1}), (x_{t_1}, S_{x_{t_1}}), \dots, (x_{t_s}, S_{x_{t_s}}), (x_T, S_{x_T})$  in  $MES(R)$  such that

$$x_1 \succ x_{t_1} \succ \dots \succ x_{t_s} \succ x_T \quad (9)$$

and

$$(x_T, S_{x_T})PD(x_{t_s}, S_{x_{t_s}})PD \dots PD(x_{t_1}, S_{x_{t_1}})PD(x_1, S_{x_1}). \quad (10)$$

Observe that (9) and (10) hold simultaneously without loss of generality, because if we have  $t$  and  $t'$  with the property that  $(x_T, S_{x_T})PD(x_t, S_{x_t})PD(x_{t'}, S_{x_{t'}})PD(x_1, S_{x_1})$  and  $x_1 \succ x_t \succ x_{t'} \succ x_T$  then, we could proceed by eliminating  $(x_t, S_{x_t})$  in the sequence. However, we will have to take care of this possibility at the beginning of Case II below. Let  $Q$  be the set of those indices; *i.e.*,  $Q = \{1, t_1, \dots, t_s, T\}$ . Again, observe that

$$S_{x_1} = S_{x_{t_1}} = \dots = S_{x_{t_s}} = S_{x_T}. \quad (11)$$

We now show that  $o(f, b) = (x_T, S_{x_T})$ . Assume not. By (7), there exists  $1 \leq q \leq Z$  such that  $o(f, b) = (x_q, S_{x_q}) \neq (x_T, S_{x_T})$ . By CLAIM 3, we can assume that

$$h(f) = \underbrace{((NP, 0), \dots, (NP, 0))}_{(i-1)\text{-times}}, \underbrace{(x_q, \#S_{x_q})}_{\text{step } i}, \underbrace{((NP, 0), \dots, (NP, 0))}_{(n-1)\text{-times}}, \quad (12)$$

where  $i \in S_{x_q}$  and for all  $j \in S_{x_q} \setminus \{i\}$ ,  $j > i$ . We distinguish between two cases.

Case I:  $q \in Q$ . Then, by (11),  $S_{x_q} = S_{x_T}$ . Consider agent  $i$ 's deviation  $\tilde{f}_i$  from  $f$  such that  $\tilde{f}_i$  is equal to  $f_i$  in all histories except in the history  $h^{i-1}$ , where  $\tilde{f}_i(h^{i-1}) = (x_T, \#S_{x_T})$ ; *i.e.*, agent  $i$  instead of proposing  $(x_q, \#S_{x_q})$  proposes  $(x_T, \#S_{x_T})$ . It is easy to see that  $o((\tilde{f}_i, f_{-i}), b) = (x_T, S_{x_T})$  holds because  $(x_T, S_{x_T})PD(x_q, S_{x_q})$  and  $o(f, b) = (x_q, S_{x_q})$  is a SPNE outcome of  $\Gamma(R)$ . Thus, the deviation  $\tilde{f}_i$  is profitable, contradicting that  $(f, b)$  is a SPNE of  $\Gamma(R)$ .

Case II:  $q \notin Q$ . If  $(x_q, S_{x_q})PD(x_1, S_{x_1})$  then  $(x_T, S_{x_T})PD(x_q, S_{x_q})$  and  $S_{x_q} = S_{x_T}$  (*i.e.*,  $(x_q, S_{x_q})$  was eliminated in the sequence going from  $(x_1, S_{x_1})$  to  $(x_T, S_{x_T})$  to make

sure that conditions (9) and (10) hold simultaneously). Since  $S_{x_q} = S_{x_T}$  we can proceed as in Case I above and obtain a contradiction with the fact that  $(f, b)$  is a SPNE of  $\Gamma(R)$ . Thus, assume  $(x_q, S_{x_q})$  does not Pareto dominate  $(x_1, S_{x_1})$ . Then, there exists  $j \in S_{x_1}$  such that  $(x_1, S_{x_1}) P_j(x_q, S_{x_q})$ . Consider agent  $j$ 's deviation  $\tilde{f}_j$  from  $f$  such that  $\tilde{f}_j$  is equal to  $f_j$  in all histories except in the history  $h^{j-1} = (p_1(f), \dots, p_{j-1}(f))$ , where  $\tilde{f}_j(h^{j-1}) = (x_1, \#S_{x_1})$ , and  $\tilde{f}_j(h^{m-1}) = (NP, 0)$  for all  $m > j$  with  $j = m \pmod{n}$ . Observe that, by (12) and the fact that  $x_1 \succ x_q$ , the proposal  $p_j = (x_1, \#S_{x_1})$  is feasible for agent  $j$  after the history  $h^{j-1}$ . By using the argument already used in Case 2 above and the fact that  $x_1 \succ x_t$  holds for all  $1 < t \leq Z$ , it follows that  $o((\tilde{f}_j, f_{-j}), b) = (x_1, S_{x_1})$ . Thus, the deviation  $\tilde{f}_j$  is profitable, which contradicts that  $(f, b)$  is a SPNE of  $\Gamma(R)$ .

Hence  $o(f, b) = (x_T, S_{x_T})$ . Thus,  $o(f, b)$  is an efficient and stable allocation in the maximal participation set. This finishes the proof of Theorem 1.  $\blacksquare$

Observe that the unique SPNE outcome  $o(f, b)$  of  $\Gamma(R)$  is uniquely identified from  $R$ , but it depends on the given linear order  $\succ$ .<sup>17</sup> Denote this outcome by  $(x^\succ, S_{x^\succ})$ . That is, given  $\succ$ , define the social choice function  $\varphi_\succ : \mathcal{R} \rightarrow A$  as follows: for each  $R \in \mathcal{R}$ ,  $\varphi_\succ(R) = (x^\succ, S_{x^\succ})$ . Hence, the following Corollary holds.

**Corollary 1** *The extensive-game form  $\Gamma^\succ$  implements in SPNE the social choice function  $\varphi_\succ : \mathcal{R} \rightarrow A$ .*

## 5.4 Order Invariance and No $\varepsilon$ -costs

To state that the SPNE outcome is invariant with respect to the order in which agents take decisions along the game, let  $\sigma : \{1, \dots, n\} \rightarrow N$  be any one-to-one mapping such that  $\sigma(i) \neq i$  for some  $i \in N$ ; *i.e.*,  $\sigma$  represents the order  $\sigma(1), \dots, \sigma(n)$ , which is different from the order  $1, \dots, n$  used to define our mechanism. Fix  $R \in \mathcal{R}$  and let  $\Gamma^\sigma(R)$  be the corresponding game in extensive form defined by our mechanism with the order  $\sigma(1), \dots, \sigma(n)$ .

---

<sup>17</sup>The extensive-game form  $\Gamma$  depends on the linear order  $\succ$ . To make this explicit, write  $\Gamma^\succ$ .

**Corollary 2** *Let  $R \in \mathcal{R}$  be given. The unique SPNE outcome of  $\Gamma^\sigma(R)$  coincides with the unique SPNE outcome of  $\Gamma(R)$ .*

Corollary 2 follows because all the arguments used in the proof of Theorem 1 to show that the game  $\Gamma(R)$  has a unique SPNE outcome do not depend on the order in which agents take decisions along the game. Thus, if  $o(f, b)$  is the unique SPNE outcome of  $\Gamma(R)$  and  $o(f^\sigma, b^\sigma)$  is the unique SPNE outcome of  $\Gamma^\sigma(R)$  then,  $o(f, b) = o(f^\sigma, b^\sigma)$ .

Finally, in Lemma 6 below we state that in equilibrium no agent pays any  $\varepsilon$ -cost.

**Lemma 6** *Let  $R \in \mathcal{R}$  be given. Assume  $(f, b)$  is a SPNE of  $\Gamma(R)$ . Then, for all  $i \in N$ ,  $\varepsilon_i(f, b) = 0$ .*

**Proof** Assume  $\varepsilon_i(f, b) \neq 0$  for some  $i \in N$ . Then, by CLAIM 1 in the proof of Theorem 1, there are at least two agents who made a proposal at Stage 1 and both agents consume at the final allocation. Let  $o_1(f) = (\hat{x}, \hat{k}, \hat{i})$  and  $m$  be the step in which  $\hat{i}$  made the accepted proposal; *i.e.*,  $p_m(f) = (\hat{x}, \hat{k})$ , where  $\hat{i} = m \pmod{n}$  and  $p_{m+1}(f) = \dots = p_{m+n-1}(f) = (NP, 0)$ . Hence, there exist agent  $j \in N \setminus \{\hat{i}\}$ , step  $m'$ , and history  $h^{m'-1} = (p_1(f), \dots, p_{m'-1}(f))$  such that  $m' < m$ ,  $j = m' \pmod{n}$  and  $f_j(h^{m'-1}) = (z, k) \neq (\hat{x}, \hat{k})$ . Thus,  $\varepsilon_j(f, b) = \varepsilon'$ . Consider another consumption strategy  $\tilde{f}_j$  that coincides with  $f_j$  in all histories except in the history  $h^{j-1} = (p_1(f), \dots, p_{j-1}(f))$  where  $\tilde{f}_j(h^{j-1}) = (\hat{x}, \hat{k})$ . Note that since  $(\hat{x}, \hat{k})$  was a feasible proposal after  $(z, k)$  was proposed it has to be also feasible at step  $j$ . If  $o_1(\tilde{f}_j, f_{-j}) = (\hat{x}, \hat{k}, j)$ , then  $f$  is not a SPNE since agent  $j$  does not have to pay the  $\varepsilon'$ -cost, and therefore  $\tilde{f}_j$  is a profitable deviation from  $f$ . If  $o_1(\tilde{f}_j, f_{-j}) = (w, q, l) \neq (\hat{x}, \hat{k}, j)$  then, using a similar argument than the one used in the proof of Theorem 1 we should find a player (the first to propose after  $j$  proposes  $(\hat{x}, \hat{k})$  when playing according to  $(\tilde{f}_j, f_{-j})$  should also have played the same proposal after  $\hat{i}$  proposes  $(\hat{x}, \hat{k})$  when playing according to  $f$ ) with a profitable deviation, contradicting the fact that  $(f, b)$  is a SPNE of  $\Gamma(R)$ . ■

## 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,” *International Journal of Game Theory* 34, 33-54 (2006).
- [6] Bogomolnaia, A. and A. Nicolò. “Stable assignment of public facilities under congestion,” *Journal of Public Economic Theory* 7, 63-95 (2005).
- [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,” *Economic Theory* 28, 453-460 (2006).

- [12] Vartiainen, H. “Subgame perfect implementation: a full characterization,” forthcoming in *Journal of Economic Theory* (2005).
- [13] Vartiainen, H. “Subgame perfect implementation of voting rules via randomized mechanisms,” forthcoming in *Social Choice and Welfare* (2006).