DOCUMENTO DE TRABAJO: E2006/11

Strategy-Proof Coalition Formation

CARMELO RODRÍGUEZ ÁLVAREZ

Centro de Estudios Andaluces
CONSEJERÍA DE LA PRESIDENCIA

Centro de Estudios Andaluces
CONSEJERÍA DE LA PRESIDENCIA

El Centro de Estudios Andaluces es
una entidad de carácter científico y
cultural, sin ánimo de lucro, adscrita
a la Consejería de la Presidencia
de la Junta de Andalucía.
El objetivo esencial de esta institución
es fomentar cuantitativa y cualitativamente
una línea de estudios e investigaciones
científicas que contribuyan a un más
preciso y detallado conocimiento de
Andalucía, y difundir sus resultados
a través de varias líneas estratégicas.

El Centro de Estudios Andaluces desea
generar un marco estable de relaciones
con la comunidad científica e intelectual
y con movimientos culturales en
Andalucía desde el que crear verdaderos
canales de comunicación para dar
cobertura a las inquietudes intelectuales y culturales.

Las opiniones publicadas por los autores en
esta colección son de su exclusiva responsabilidad

1   Introduction

We study simple coalition formation problems in which a group of agents is partitioned into coalitions and agents have preferences over the coalitions they are members of. Following the terminology proposed by Drèze and Greenberg (1980), we focus on problems characterized by the "hedonic" aspect of coalition formation. Agents' preferences only depend on the identity of the members of the coalition they belong to. Hence, we exclude the existence of externalities among different coalitions. Relevant examples of such problems are matching problems such as marriage and roommate problems, or the formation of social clubs, teams, and societies.

The formation of coalitions is a relevant phenomenon in a wide variety of social and economic environments. The rationale behind the formation of coalitions is that agents form groups in order to exploit the joint benefits of cooperation. The literature on Coalitional Game Theory has extensively analyzed the existence of stable partitions in hedonic coalition formation problems.[1] Instead, we propose a social choice and implementation approach. We study coalition formation rules that associate to each profile of agents' preferences a partition of the group of agents. A coalition formation rule can be interpreted as an optimal recommendation for the society that represents an optimal compromise between the conflicting preferences of the agents. However, since preferences are not observable, they must be elicited from the agents. Thus, given a coalition formation rule, a fundamental concern is whether or not agents have the incentive to reveal their *true* preferences. In this paper, we analyze the possibility of devising coalition formation rules that always give agents such an incentive. Hence, we are interested in rules that satisfy *strategy-proofness*. Strategy-proofness is the strongest decentrability property. It implies that it is a dominant strategy for the agents to straight-forwardly reveal their preferences. Moreover, each agent needs to know only her own preferences to compute her best choice.

It is well known that the requirements of strategy-proofness are hard to meet. In the abstract model of social choice, Gibbard (1973) and Satterthwaite (1975) show that – provided there are more than two alternatives at stake– every strategy-proof social choice rule is dictatorial. However, reasonable strategy-proof rules exist if appropriate restrictions are imposed on agents' preferences. In coalition formation problems, such domain

---

[1]For further references, see the recent works by Banerjee, Konishi, and Sönmez (2001), Barberà and Gerber (2003), Bogomolnaina and Jackson (2002), and Pápai (2004).

1

restrictions arise naturally. On the one hand, while coalition formation rules select a partition for each preference profile, each agent only cares about the coalition she is a member of. On the other hand, additional restrictions on how an agent may compare different coalitions can be easily justified. For instance, an interesting class of problems consists of situations in which there are no complementarities among the members of a coalition. That is, the preferences of an agent i regarding the convenience of an agent j joining the coalition i belongs to, do not depend on the coalition to which i is assigned.[2] Then, agents' preferences are additively representable or separable. These domains of preferences have been studied in the general context of abstract social choice by Barberà, Sonnenschein, and Zhou (1991) and Le Breton and Sen (1999), among others, and positive results have been obtained. Yet, the possibility of constructing strategy-proof coalition formation rules when agents' preferences are additively representable or separable has not been addressed in the literature.

Besides strategy-proofness, we would like our rules to satisfy four additional properties. Our rules should be *individually rational*, *Pareto efficient*, *non-bossy*, and *flexible*. Individual rationality is a participation constraint. It means that no agent should ever be worse-off than she would be if staying alone. Pareto efficiency requires that the whole society of agents should not prefer a partition formed by feasible coalitions rather than forming the partition that the rules recommends. Non-bossiness is a collusion-proof requirement. It says that if a change in an agent's preferences does not affect the coalition to which this agent is assigned, then the remaining agents are also unaffected by this change of preferences. Flexibility is implied by Pareto efficiency. It says that every partition formed by a collection of feasible coalitions belongs to the range of the rule. Hence, flexibility implies that feasible disjoint coalitions are mutually compatible.

We provide two characterizations of a family of rules, the family of single-lapping rules, that fulfill the previous axioms in minimally rich domains of preferences (as the domain of additively representable preferences). Single-lapping rules are characterized by strong restrictions over the set of feasible coalitions –the *single-lapping property*– that can be justified by the initial existence of a hierarchical structure of the society. The

---

[2]Think, for example, in the preferences of a senior member of an Economics Department about the job–candidates for two tenure–track positions that are available (but that need not to be filled). Suppose that there are two candidates, a macroeconomist and an econometrician. If the senior economist prefers hiring the macroeconomist rather than not hiring anybody, then the senior economist should also prefer hiring the macroeconomist and the econometrician rather than hiring the econometrician alone.

single-lapping property was firstly introduced by Pápai (2004), who shows that it is a necessary and sufficient condition for the existence of a unique core-stable partition of the society for every profile of agents' preferences. In fact, single-lapping rules always select the unique core-stable partition of the society, in the sense that no feasible coalition of agents unanimously prefer joining each other rather than staying at the coalition they are assigned to. Hence, our results provide further evidence on the relation between the non-cooperative game theory concept of strategy-proofness and the cooperative game theory concept of the existence of a unique core-stable partition.

Before proceeding with the formal analysis, we review the most related literature. This paper is is closely related to the already mentioned Pápai (2004). Pápai's main focus is on finding necessary and sufficient conditions on the set of feasible coalitions for t uniqueness of core-stable partitions. Additionally, this author shows that, given an initial set of coalitions that satisfy the single-lapping property, its associated single-lapping rule is the unique rule that satisfies strategy-proofness, individual rationality, and Pareto efficiency when agents' preferences over coalitions are restricted to prefer any coalition in the initial set to any other coalition. Our analysis complements Pápai's results in several directions. We show that the single-lapping structure of the set of feasible coalition is implied directly by strategy-proofness and the remaining axioms. Moreover, we show that the results also hold in more restricted domains of preferences over coalitions.

The manipulability of coalition formation rules has also been studied by Alcalde and Revilla (2004), Cechlárová and Romero-Medina (2001), Sönmez (1999), and Takamiya (2003). However, these works focus on different domains of preferences that are not consistent with additively representable or separable preferences. More specifically, Alcalde and Revilla (2004), and Cechlárová and Romero-Medina (2001) assume that agents' preferences over coalitions are based on the best or the worst group of agents in each coalition. In these environments, they prove the existence of strategy-proof rules that always select core-stable partitions. Finally, Sönmez (1999) proposes a general model of allocation of indivisible goods which includes our coalition formation model as a special case. This author focuses on problems for which there always exist core-stable partitions. Under some assumptions on agents' preferences, Sönmez (1999) shows that there exist strategy-proof, individually rational, and Pareto efficient rules only if the set of core-stable partitions is always essentially single-valued. Takamiya (2003) proves that the converse result also holds under additional assumptions on preferences –such as strict preferences and no

3

consumption externalities– that are fulfilled in coalition formation problems.

The remainder of the paper is organized as follows. In Section 2, we present the model and basic notation. In Section 3, we present different domains of preferences over coalitions and the notion of minimally rich domain. In Section 4, we introduce the main axioms while in Section 5 we present single-lapping rules and provide the characterization results. In Section 6, we prove Theorem 2. We include the proofs of some intermediate results and supplemental material in the Appendices.

2 Basic Notation

Let $N \equiv \{1, \ldots, n\}$ be a society consisting of a finite set of at least 3 agents, $(n \geq 3)$. We call a non-empty subset $C \subseteq N$ a **coalition**. Let $\mathcal{N}$ denote the set of all non-empty subsets of N. For each $C \in \mathcal{N}$, let $[C] \equiv \{\{i\} : i \in C\}$. A **collection of coalitions** is a set of coalitions $\Pi \subseteq \mathcal{N}$ that contains all singleton sets, $[N] \subseteq \Pi$. Let σ be a partition of N and let Σ denote the set of all partitions of N. For each $i \in N$ and each $\sigma \in \Sigma$, we denote by $\sigma_i \in \sigma$ the coalition in σ to which i belongs.

For each $i \in N$, let $\mathcal{C}_i \equiv \{C \subseteq N, \ i \in C\}$. That is, $\mathcal{C}_i$ is the set of all coalitions to which i belongs. A **preference** for i, $\succsim_i$, is a complete order on $\mathcal{C}_i$.[3] For each $i \in N$, we denote by $\mathcal{D}_i$ the set of all preferences for i. Note that preferences are strict. Hence, for each $i \in N$, each $\succsim_i \in \mathcal{D}_i$, and each $C, C' \in \mathcal{C}_i$, we write $C \succ_i C'$ to indicate that i strictly prefers C to C', and $C \succsim_i C'$ to indicate that either $C \succ_i C'$ or $C = C'$. We assume that agents only care about the coalition they belong to, then agents' preferences over partitions are completely defined by their preferences over coalitions. Thus, abusing notation, we say that for each $i \in N$, each $\succsim \in \mathcal{D}_i$, and each $\sigma, \sigma' \in \Sigma$, σ is at least as good as σ', $\sigma \succsim_i \sigma'$, if and only if $\sigma_i \succsim_i \sigma'_i$.

For each $i \in N$, each set of coalitions $\mathcal{X} \subseteq \mathcal{N}$ with $\mathcal{X} \cap \mathcal{C}_i \neq \varnothing$, and each $\succsim_i \in \mathcal{D}_i$, let $\text{top}(\mathcal{X}, \succsim_i)$ be the coalition in $\mathcal{X} \cap \mathcal{C}_i$ that is ranked first according to $\succsim_i$.

Let $\mathcal{D} \equiv \times_{i \in N} \mathcal{D}_i$. We call $\succsim \in \mathcal{D}$ a preference profile. For each $C \subset N$, $\mathcal{D}_C = \times_{i \in C} \mathcal{D}_i$, while for each $\succsim \in \mathcal{D}$, $\succsim_C \in \mathcal{D}_C$ denotes the restriction of profile $\succsim$ to the preferences of the agents in C.

Let $\tilde{\mathcal{D}} \subseteq \mathcal{D}$, we say that $\tilde{\mathcal{D}}$ is a **cartesian domain** if for each $i \in N$ there is $\tilde{\mathcal{D}}_i \subseteq \mathcal{D}_i$ such that $\tilde{\mathcal{D}} = \times_{i \in N} \tilde{\mathcal{D}}_i$.

---

[3]An order is a reflexive, transitive, and antisymmetric binary relation.

We are interested in rules that associate a partition of the society to each profile of agents' preferences.

Let $\tilde{\mathcal{D}} \subset \mathcal{D}$ be a cartesian domain. A *(coalition formation) rule* defined on the domain $\tilde{\mathcal{D}}$ is a mapping $\varphi : \tilde{\mathcal{D}} \to \Sigma$.

Naturally, for each $i \in N$ and each $\succsim \in \tilde{\mathcal{D}}$, $\varphi_i(\succsim)$ denotes the coalition in $\varphi(\succsim)$ to which i belongs.

Finally, R^φ denotes the range of φ, that is, the set of feasible partitions,

$$R^\varphi \equiv \{\sigma \in \Sigma, \text{ such that there is } \succsim \in \tilde{\mathcal{D}}, \ \varphi(\succsim) = \sigma\},$$

while, F^φ denotes the set of feasible coalitions,

$$F^\varphi \equiv \{C \in \mathcal{N}, \text{ such that for some } \sigma \in R^\varphi, \ C \in \sigma\}.$$

3 Restricted Domains of Preferences over Coalitions

We start by presenting two classes of preferences over coalitions – *top* and *bottom preferences*– that play a crucial role in our analysis. Both domains are contained in other domains of preferences that have been extensively analyzed in the social choice literature, namely, the domains of additively representable and separable preferences. Top and bottom preferences are obtained by extending orders over single agents to orders over coalitions. The basic idea behind top and bottom preferences is that each agent i divides the society into two groups according to some order over the set of agents: the agents that she likes and the agents she dislikes. An agent equipped with top preferences prioritizes (lexicographically) joining the agents she likes the most with respect to avoiding the agents she dislikes. On the other hand, an agent equipped with bottom preferences prioritizes (lexicographically) avoiding the agents she dislikes the most with respect to joining the agents she likes.

Let $\mathcal{P}$ be the set of all complete orders over N. For each $P \in \mathcal{P}$, R denotes the weak order associated to P and it is defined in the usual way. For each $C \subseteq N$ and each $P \in \mathcal{P}$, $\max(C, P)$ and $\min(C, P)$ denote, respectively, the first-ranked and the last-ranked agent of C according to P. Next, for each $i \in N$, each $P \in \mathcal{P}$, and each $C \in \mathcal{C}_i$, let $C_i^+(P) \equiv \{j \in C, \text{ s.t. } j\, R\, i\}$, and $C_i^-(P) \equiv \{j \in C \text{ s.t. } i\, R\, j\}$. Now, define $C_i^+(1, P) \equiv \max(C_i^+(P), P)$ and $C_i^-(1, P) \equiv \min(C_i^-(P), P)$. Once $C_i^+(t, P)$ and

$C_i^-(t, P)$ are defined for some $t \geq 1$, iteratively, let

$$C_i^+(t+1, P) \equiv \max\left(\left[C_i^+(P) \setminus \cup_{k=1}^t C_i^+(k, P)\right], P\right), \text{ and}$$
$$C_i^-(t+1, P) \equiv \min\left(\left[C_i^-(P) \setminus \cup_{k=1}^t C_i^-(k, P)\right], P\right).$$

For each $i \in N$ and each $P \in \mathcal{P}$, the preference $\succsim_i \in \mathcal{D}_i$ is the **top preference associated to P by** i, $\succsim_i = \succsim_i^+(P)$ if for each two distinct coalitions $C, C' \in \mathcal{C}_i$, $C \succ_i C'$ if and only if

- $C_i^+(P) \neq C_i'^+(P)$ and $C_i^+(t, P) \ P \ C_i'^+(t, P)$, where t is the first integer such that $C_i^+(t, P) \neq C_i'^+(t, P)$.

- $C_i^+(P) = C_i'^+(P)$ and $C_i^-(t', P) \ P \ C_i'^-(t', P)$, where t' is the first integer such that $C_i^-(t', P) \neq C_i'^-(t', P)$.

Let $i \in N$, $P \in \mathcal{P}$, and let $C, C' \in \mathcal{C}_i$ be such that $C \neq C'$. When comparing the coalitions C and C', if agent i is equipped with preference $\succsim_i^+(P)$, then she focuses on the sets of agents who are ranked above i according to P, $C_i^+(P)$ and $C_i'^+(P)$. First, i compares C and C' on the basis of the agents who are first-ranked according to P in $C_i^+(P)$ and $C_i'^+(P)$. If these agents are the same, then i compares the second-ranked agents and so on. If $C_i^+(P) = C_i'^+(P)$, then i turns her attention to the agents who are ranked below i according to P, $C_i^-(P)$ and $C_i'^-(P)$, and applies the same lexicographic logic, but starting from the bottom. She compares first the last ranked agents in $C_i^-(P)$ and $C_i'^-(P)$, and she proceeds iteratively in the case that they are the same agent.

The logic behind bottom preferences mimics top preferences.

For each $i \in N$ and each $P \in \mathcal{P}$,, the preference $\succsim_i \in \mathcal{D}_i$ is the **bottom preference associated to P by** i, $\succsim_i = \succsim_i^-(P)$ if for each two distinct coalitions $C, C' \in \mathcal{C}_i$, $C \succ_i C'$ if and only if

- $C_i^-(P) \neq C_i'^-(P)$, and $C_i^-(t, P) \ P \ C_i'^-(t, P)$, where t is the first integer such that $C_i^-(t, P) \neq C_i'^-(t, P)$.

- $C_i^-(P) = C_i'^-(P)$ and $C_i^+(t', P) \ P \ C_i'^+(t', P)$, where t' is the first integer such that $C_i^+(t', P) \neq C_i'^+(t', P)$.

Let $i \in N$, $P \in \mathcal{P}$, and let $C, C' \in \mathcal{C}_i$ be such that $C \neq C'$. When comparing the coalitions C and C', if agent i is equipped with the preference $\succsim_i^- (P)$, then i focuses on the sets of agents who are ranked below i according to P, $C_i^-(P)$ and $C_i'^-(P)$. First, i compares C and C' on the basis of the agents who are last-ranked according to P in $C_i^-(P)$ and $C_i'^-(P)$. If these agents are the same, then i compares the next-to-the-last ranked agents and so on. If $C_i^-(P) = C_i'^-(P)$, then i turns her attention to the agents who are ranked above i according to P, $C_i^+(P)$ and $C_i'^+(P)$, and applies the same lexicographic logic, but starting from the top. First, she compares the first-ranked agents in $C_i^+(P)$ and $C_i'^+(P)$, and she proceeds iteratively in the case that they are the same agent.

For each $i \in N$, let

$$\mathcal{D}_i^+ \equiv \{\succsim_i \in \mathcal{D}_i \text{ such that for some } P \in \mathcal{P}, \ \succsim_i = \succsim_i^+ (P)\},$$
$$\mathcal{D}_i^- \equiv \{\succsim_i \in \mathcal{D}_i \text{ such that for some } P \in \mathcal{P}, \ \succsim_i = \succsim_i^- (P)\},$$
$$\mathcal{D}_i^* \equiv \mathcal{D}_i^+ \cup \mathcal{D}_i^- \text{ and,}$$
$$\mathcal{D}^* \equiv \times_{i \in N} \mathcal{D}_i^*.$$

Let $\bar{\mathcal{D}} \subseteq \mathcal{D}$. We say that $\bar{\mathcal{D}}$ is **minimally rich** if $\bar{\mathcal{D}}$ is cartesian and $\mathcal{D}^* \subseteq \bar{\mathcal{D}}$.

We consider that a domain of preferences over coalitions is minimally rich if it contains top and bottom preferences. Minimal richness also requires that the domain is cartesian. That is, an agent's set of admissible preferences does not depend on the preferences of the remaining agents.

The following remark shows that in minimally rich domains, the preferences of an agent regarding the way in which she may compare the coalition in which she stays on her own and any two other different coalitions she may belong to are not restricted.

Remark 1. *For each $i \in N$ and each two distinct $C, C' \in \mathcal{C}_i \setminus \{i\}$, there exist $\succsim, \succsim', \succsim'' \in \mathcal{D}_i^*$ such that:*

$$\{i\} \succ C \succ C',$$
$$C \succ' \{i\} \succ' C', \text{ and}$$
$$C \succ'' C' \succ'' \{i\}.$$

It can be argued that top and bottom preferences reflect rather extreme preferences over coalitions. However, the domains of additively representable and separable preferences are minimally rich. These domains exclude the possibility of (negative or positive) complementarities among the members of a coalition.

Let $i \in N$. A ***utility function*** for agent i is a mapping $u_i : N \to \mathbb{R}$ such that $u_i(i) = 0$. A preference for agent i, $\succsim_i \in \mathcal{D}_i$ is ***additively representable*** if there is a utility function u_i such that for each $C, C' \in \mathcal{C}_i$, $C \succsim_i C'$ if and only if $\sum_{c \in C} u_i(c) \geq \sum_{c' \in C'} u_i(c')$. For each $i \in N$, $\mathcal{A}_i$ denotes the set of all i's additively representable preferences for agent i and let $\mathcal{A} \equiv \times_{i \in N} \mathcal{A}_i$.

A preference for i, $\succsim_i \in \mathcal{D}_i$, is ***separable*** if for each $j \in N$ and each $C \in \mathcal{C}_i$ such that $j \notin C$, $\{i, j\} \succ_i \{i\}$ if and only if $(C \cup \{j\}) \succ_i C$. Let $\mathcal{S}_i$ be the set of all agent i's separable preferences and let $\mathcal{S} \equiv \times_{i \in N} \mathcal{S}_i$.

The following remark shows that the domain of additively representable preferences and the domain of separable preferences are indeed minimally rich domains. Moreover, for small societies both domains coincide with the smallest minimally rich domain.

Remark 2. *Let $i \in N$.*

(a) *If $n \geq 4$, then $\mathcal{D}_i^* \subset \mathcal{A}_i \subset \mathcal{S}_i$.*

(b) *If $n = 3$, then $\mathcal{D}_i^* = \mathcal{A}_i = \mathcal{S}_i$.*

4   Axioms

This section introduces four properties that rules may satisfy. Let $\tilde{\mathcal{D}} \subseteq \mathcal{D}$ be a cartesian domain and let φ be a rule defined on $\tilde{\mathcal{D}}$.

Our main axiom is an incentive constraint. A rule should never provide an incentive for an agent to misreport her preferences. Only if a rule elicits the true preferences from the agents, the social choice will be based upon the correct information. Of course, this property refers to the specific domain in which the rule is defined.

Strategy-Proofness*.* For each $i \in N$, each $\succsim \in \tilde{\mathcal{D}}$, and each $\succsim_i' \in \tilde{\mathcal{D}}_i$, $\varphi_i(\succsim) \succsim_i \varphi_i(\succsim_{N \setminus \{i\}}, \succsim_i')$. Conversely, we say that $i \in N$ ***manipulates*** φ if there exist $\succsim \in \tilde{\mathcal{D}}$ and $\succsim_i' \in \tilde{\mathcal{D}}_i$ such that $\varphi_i(\succsim_{N \setminus \{i\}}, \succsim_i') \succ_i \varphi_i(\succsim)$.

The Gibbard-Satterthwaite Theorem states that every *strategy-proof* rule on an unrestricted domain either is dictatorial or its range contains only two elements.[4] As we

---

[4]A rule $\varphi : \tilde{\mathcal{D}} \to \Sigma$ is ***dictatorial*** if there is $i \in N$ (a dictator) such that for each $\succsim \in \tilde{\mathcal{D}}$, $\varphi_i(\succsim) = \text{top}(F^\varphi, \succsim_i)$.

assume that agents' preferences over social outcomes are restricted to depend only on the coalitions they are members of and we focus on minimally rich domains, the negative consequences of the Gibbard-Satterthwaite Theorem do not apply to our framework.

We also consider a minimal participation constraint. Agents should not prefer to stay on their own rather than to belong to the coalition that the rule assigns them.

Individual Rationality. For each $i \in N$ and each $\succsim \in \tilde{\mathcal{D}}$, $\varphi_i(\succsim) \succsim_i \{i\}$.

Note that, for every *individually rational* rule, its set of feasible allocations is a collection of coalitions.

We introduce a weak version of efficiency. This notion of efficiency for coalition formation problems is introduced in Pápai (2004).

Pareto efficiency. For each $\succsim \in \tilde{\mathcal{D}}$, there is no $\sigma \in \Sigma$ such that for each $C \in \sigma$, $C \in F^\varphi$, and for every $i \in N$, $\sigma_i \succsim_i \varphi_i(\succsim)$, and for some $j \in N$, $\sigma_j \succ_j \varphi_j(\succsim)$.

Note that *Pareto efficiency* is a version of efficiency restricted to the set of feasible range of the rule. *Pareto efficiency* does not implies onto-ness.[5] That is, it may be the case that a *Pareto efficient* rule does not admit every conceivable coalition as feasible.

We consider rules such that whenever a change in an agent's preference does not change the coalition she is assigned to, then the assignment for the remaining agents does not change.

Non-Bossiness. For each $i \in N$, each $\succsim \in \tilde{\mathcal{D}}$, and each $\succsim'_i \in \tilde{\mathcal{D}}_i$, $\varphi_i(\succsim) = \varphi_i(\succsim_{N \setminus \{i\}}, \succsim'_i)$ implies $\varphi(\succsim) = \varphi(\succsim_{N \setminus \{i\}}, \succsim'_i)$.

We can interpret *non-bossiness* as a collusion-proof or bribe-proof condition. Imagine that there exists a transferable private good and that agents preferences over coalitions and private good allocations are lexicographic. Agents focus first on the coalition they are assigned, and then in the private good allocation. A violation of *non-bossiness* implies a possibility of collusion because an agent might have incentives to misrepresent her

---

[5] A rule $\varphi : \tilde{\mathcal{D}} \to \Sigma$ satisfies **onto-ness** if $F^\varphi = \mathcal{N}$.

preferences in exchange for a positive transfer of the private good from those who benefit from the change in her preference report.

Finally, we introduce a minimal flexibility condition on the range of the rule. We assume that the range of a rule is determined by the set of feasible coalitions.

Flexibility. For each $\sigma = \{C_1, \ldots, C_m\} \in \Sigma$, $C_t \in F^{\varphi}$ for each $t = 1, \ldots, m$, implies $\sigma \in R^{\varphi}$.

Flexibility means that any two disjoint feasible coalitions are mutually compatible. Hence, it implies that the range of the rule is completely determined by the set of feasible coalitions. *Flexibility* is implied by *Pareto efficiency*, but *flexibility* does not implies *ontoness*. By focusing on *flexible* rules, we rule out some coalition formation problems. For instance, any rule defined in a four-agent society, in which every couple of agents is a feasible coalitions but partitions containing two couples are not admissible would violate *flexibility*.

5 Characterization Results

In this section we analyze the implications of the axioms listed above over rules defined on rich domains. First, we introduce additional notation due to Pápai (2004). This author proposes a property over sets of coalitions – the single-lapping property– that ensures the existence and uniqueness of a core-stable partition for every preference profile.[6] We make use of this property to define a class of rules.

A collection of coalitions Π satisfies the ***single-lapping property*** if

Condition (a)**:** For each $C, C' \in \Pi$, $C \neq C'$ implies $\#(C \cap C') \leq 1$.

Condition (b)**:** For each $\{C_1, \ldots, C_m\} \subseteq \Pi$ with $m \geq 3$ and for each $t = 1, \ldots, m$, $\#(C_t \cap C_{t+1}) \geq 1$ (where $m + 1 = 1$), there is $i \in N$ such that for each $t = 1, \ldots, m$, $C_t \cap C_{t+1} = \{i\}$.

---

[6]Given a preference profile $\succsim \in \mathcal{D}$ and a collection of coalitions $\Pi \subseteq \mathcal{N}$, the partition $\sigma \in \Sigma$ is ***core-stable*** if there is no $C \in \Pi$ such that for each $j \in C$, $C \succ_j \sigma_j$.

Condition (a) states that if there is an overlap between any two coalitions in the collection, there cannot be more than one agent who is member of these two coalitions. *Condition (b)* is a non-cycle condition. It requires that if a set of coalitions in the collection form a cycle in which every two neighbor coalitions have a common member, then all these coalitions have the same common member.

Pápai (2004) shows that single-lapping collections of coalitions can be associated to a non-directed graph endowed with a tree or network structure. Tree structures are characteristic to many hierarchical societies or networks in which only members of adjacent levels in the hierarchy (network) are connected and can form a coalition.[7]

The following remark presents a prominent property of single-lapping collections of coalitions. For every single-lapping collection of coalitions and for every preference profile, there is a coalition in the collection such that all its members think that this coalition is the best coalition in the collection.

Remark 3. *(Pápai, 2004, Theorem 1). Let Π be a single-lapping collection of coalitions. For each $\succsim \in \mathcal{D}$ there is $C \in \Pi$ such that for each $i \in C$, $C = \text{top}(\Pi, \succsim_i)$.*

Remark 3 implies that for every single-lapping collection of coalitions and every preference profile there is a unique core-stable partition of the society. Pápai (2004) presents the following algorithm to find such partition.

For each $\succsim \in \mathcal{D}$ and each single-lapping collection of coalitions $\Pi \subset \mathcal{N}$, the ***core-stable partition associated to Π at profile $\succsim$, $\bar{\sigma}^{\Pi}(\succsim)$***, can be identified by the following algorithm:

Algorithm: *(Pápai, 2004). Let $\succsim \in \mathcal{D}$ and let Π be a single-lapping collection of coalitions. Find $C \in \Pi$ such that for each $i \in C$, $\text{top}(\Pi, \succsim_i) = C$. As Π is single-lapping, such coalition exists. Note that there may be several such coalitions, and all these coalitions are disjoint. Let*

$$\Pi(1, \succsim) \equiv \Pi,$$
$$M^{\Pi}(1, \succsim) \equiv \{C \in \Pi \text{ such that for each } i \in C, \text{top}(\Pi, \succsim_i) = C\}$$
$$T^{\Pi}(1, \succsim) \equiv \cup_{C \in M^{\Pi}(1, \succsim)} C$$

Hence, $M^{\Pi}(1, \succsim)$ denotes the set of all the coalitions that are formed in this first stage and $T^{\Pi}(1, \succsim)$ denotes the set of agents that are matched in the first stage.

---

[7]See Demange (2004, 2005) for more on the relation of hierarchical structures and coalitional stability.

Once $\Pi(t, \succsim)$, $M^{\Pi}(t, \succsim)$, and $T^{\Pi}(t, \succsim)$ are defined for some $t \geq 1$, let,

$$\Pi(t+1, \succsim) \equiv \{C \in \Pi \text{ such that } C \cap T^{\Pi}(t, \succsim) = \{\varnothing\}\},$$

$$M^{\Pi}(t+1, \succsim) \equiv \{C \in \Pi(t+1, \succsim) \text{ such that for each } i \in C, \text{top}(\Pi(t+1, \succsim), \succsim_i) = C\} \text{ and,}$$

$$T^{\Pi}(t+1, \succsim) \equiv \cup_{C \in M^{\Pi}(1, \succsim) \cup \ldots \cup M^{\Pi}(t+1, \succsim)} C.$$

Note that, for each $t = 1, \ldots, m$, $\Pi(t, \succsim) \subset \Pi$, $\Pi(t, \succsim)$ is a collection of coalitions for the reduced society $N \setminus T^{\Pi}(t, \succsim)$. Moreover, $\Pi(t, \succsim)$ satisfies the single-lapping property. Let $m \leq n$ be the smallest integer such that $T^{\Pi}(m, \succsim) = N$. Then, the algorithm identifies a unique partition,

$$\bar{\sigma}^{\Pi}(\succsim) \equiv \{C \in \Pi \text{ such that for some } t \leq m, \ C \in M^{\Pi}(t, \succsim)\}.$$

For each single-lapping collection of coalitions and each preference profile there is a unique core-stable partition. Thus, each single-lapping collection of coalitions defines a unique rule.

Let $\tilde{\mathcal{D}} \subseteq \mathcal{D}$ be a cartesian domain of preferences and let φ be a rule defined on $\tilde{\mathcal{D}}$. The rule φ is a **_single-lapping rule_** if there is a single-lapping collection of coalitions Π such that for each $\succsim \in \tilde{\mathcal{D}}$, $\varphi(\succsim) = \bar{\sigma}^{\Pi}(\succsim)$.

Pápai (2004, Theorem 1) shows that, given a fixed single-lapping collection of coalitions Π, if agents are restricted to prefer standing on their own to any other coalition $C \notin \Pi$, then the single-lapping rule associated to Π is the unique rule that satisfies _strategy-proofness_, _individual rationality_, and _Pareto efficiency_. Note that for every single-lapping rule, for each preference profile there is always a feasible coalition such that all its members think that it is their best preferred feasible coalition. Thus, single-lapping rules clearly satisfy _strategy-proofness_ in any minimally rich domain. In fact, single-lapping rules also satisfy _individual rationality_, _non-bossiness_, _flexibility_, and _Pareto efficiency_.

Theorem 1. _Let $\bar{\mathcal{D}}$ be a minimally rich domain. If a rule $\varphi : \bar{\mathcal{D}} \to \Sigma$ is single-lapping, then φ satisfies strategy-proofness, individual rationality, non-bossiness, flexibility, and Pareto efficiency._

Proof. Let $F^{\varphi} = \Pi$. Because φ is a single-lapping rule, Π is a single-lapping collection of coalitions. Let us check that φ satisfies _strategy-proofness_.[8] Let $\succsim \in \bar{\mathcal{D}}$. For each $i \in$

---

[8]This fact is proven in Pápai (2004, Theorem 3). We include the proof for the sake of completeness.

$T^\Pi(1, \succsim)$, $\varphi_i(\succsim) = \text{top}(\Pi, \succsim_i)$. Thus, agents in $T^\Pi(1, \succsim)$ cannot manipulate. Moreover, by the definition of single-lapping rule for each $\succsim' \in \bar{\mathcal{D}}$ such that for each $i \in T^\Pi(1, \succsim)$ $\succsim_i = \succsim'_i$, $\varphi_i(\succsim) = \varphi_i(\succsim')$. Now, let $j \in T^\Pi(2, \succsim)$. If there exists $C \in \Pi$ such that $C \succ_j \varphi_j(\succsim)$, then there is $i \in T^\Pi(1, \succsim))$ such that $i \in C$. Note that for each $\succsim'_j \in \bar{\mathcal{D}}_j$ and each $i \in T^\Pi(1, \succsim)$, $\varphi_i(\succsim_{N \setminus \{j\}}, \succsim'_j) = T^\Pi(1, \succsim)$. Thus, $\varphi_j(\succsim) \succsim_j \varphi_j(\succsim_{N \setminus \{j\}}, \succsim'_j)$ and j cannot manipulate. Repeating iteratively the argument with the remaining steps of the algorithm, we obtain that no agent can manipulate.

Let us check that φ satisfies *individual rationality*. By the definition of single-lapping rule, for each $i \in N$ and each $\succsim \in \bar{\mathcal{D}}$, there is $t \leq n$ such that $\varphi_i(\succsim) \in M^\Pi(t, \succsim)$. Note that $\{i\} \in \Pi(t, \succsim)$. By the definition of single-lapping rule, $\varphi_i(\succsim) \equiv \text{top}(\Pi(t, \succsim), \succsim_i)$. Thus, $\varphi_i(\succsim) \succsim_i \{i\}$, which proves *individual rationality*.

Let us check that φ satisfies *non-bossiness*. Let $i \in N$, $\succsim \in \bar{\mathcal{D}}$, and $\succsim'_i \in \bar{\mathcal{D}}_i$ be such that $\varphi_i(\succsim) = \varphi_i(\succsim_{N \setminus \{i\}}, \succsim'_i)$. Let $i \in T^\Pi(t, \succsim)$. Because φ is a single-lapping rule, for each $j \in \cup_{t' \leq t} T^\Pi(t', \succsim)$, $\varphi_j(\succsim) = \varphi_j(\succsim_{N \setminus \{i\}}, \succsim'_i)$. Moreover, because $\varphi_i(\succsim) = \varphi_i(\succsim_{N \setminus \{i\}}, \succsim'_i)$, for each $k \in \cup_{t' \geq t} T^\Pi(t', \succsim)$, we have $\varphi_k(\succsim) = \varphi_k(\succsim_{N \setminus \{i\}}, \succsim'_i)$. Then, $\varphi(\succsim) = \varphi(\succsim')$, which proves *non-bossiness*.

Let us check that φ satisfies *flexibility*. Let $\sigma = \{C_1, \ldots, C_m\} \in \Sigma$ be such that for each $t = 1, \ldots, k$, $C_t \in \Pi$. Let $\succsim \in \bar{\mathcal{D}}$ be such that for each $t = 1, \ldots, m$ and each $i \in C_t$, $\text{top}(\mathcal{N}, \succsim_i) = C_t$. By the definition of single-lapping rule, $\varphi(\succsim) = \sigma$ and $\sigma \in R^\varphi$. Thus, φ satisfies *flexibility*.

Finally, *Pareto efficiency* follows immediately from the definition of single-lapping rule. Note that for each $i \in N$ and each $\succsim \in \bar{\mathcal{D}}$ there is $t \leq n$ such that $\varphi_i(\succsim) = \text{top}(\Pi(t, \succsim), \succsim_i)$. $\qquad \square$

Note that we only need to assume that the domain of the rule is minimally rich in proving *flexibility*. The proof of the remaining axioms is domain independent. Note also that single-lapping rules satisfy *strategy-proofness* even in the unrestricted domain of preferences over coalitions $\mathcal{D}$. By restricting the set of feasible coalitions, single-lapping rules eliminate agents' opportunities for profitable misrepresentation of preferences. Our next results are, in some way, more surprising. In every minimally rich domain, single-lapping rules are the only rules that satisfy our list of axioms. Hence, reducing the set of admissible preferences for the agents does not allow for additional rules.

Theorem 2. *Let $\bar{\mathcal{D}}$ be a minimally rich domain. A rule $\varphi : \bar{\mathcal{D}} \to \Sigma$ satisfies strategy-*

proofness, individual rationality, non-bossiness, and flexibility if and only if φ is a single-lapping rule.

We present the proof of necessity part of Theorem 2 in the next section. The intuition runs as follows. For every rule that satisfies our axioms, when the members of a feasible coalition of individuals agree that this coalition is the best preferred feasible coalition, this coalition is formed. Then, it remains to check that the set of feasible coalitions satisfies the single-lapping property. This step is far from being immediate and constitutes the bulk of the proof. The analysis is relatively simple for three agents societies. We use an induction argument to extend the result to arbitrary societies.

Theorem 2 shows that only rules that select the unique core-stable partition given an initial set of feasible coalition satisfy our list of axioms. Hence, Theorem 2 provides further evidence on the relation between the concepts of *strategy-proofness* and *unique core-stability*. This relation has been already presented in previous works as Sönmez (1999) and Pápai (2004).[9] However, Theorem 2 provides several novelties with respect to previous results. We do not impose any restrictions either on preferences or on feasible coalitions that ensure the existence of core-stable partitions. Instead, we obtain that the rule selects the unique core-stable partition directly from our axioms. This fact allows us to obtain a characterization result that applies to every kind of coalition formation problem instead of impossibility results. In addition, our results apply to very restricted domains of preferences as the smallest minimally rich domain. Finally, we do not use *Pareto efficiency* in the characterization, instead we use two axioms, *non-bossiness* and *flexibility*, that are not included in the definition of *core-stability*.

The domains of additively representable and separable preferences are minimally rich domains. Hence, we obtain the following corollaries to Theorem 2.

Corollary 1. *A rule $\varphi : \mathcal{A} \to \Sigma$ satisfies strategy-proofness, individual rationality, non-bossiness, and flexibility if and only if φ is a single-lapping rule.*

Corollary 2. *A rule $\varphi : \mathcal{S} \to \Sigma$ satisfies strategy-proofness, individual rationality, non-bossiness, and flexibility if and only if φ is a single-lapping rule.*

---

[9]The line of research that investigates the existence of *strategy-proof* rules in core selecting organizations was initiated by Ledyard (1977).

Corollaries 1 and 2 are in sharp contrast with the results of Barberà, Sonnenschein, and Zhou (1991). These authors analyze problems in which the founding members of a society select new members for the society and their preferences over candidates are additively representable (or separable). They show that for those coalition formation problems, *strategy-proof* rules can be decomposed in a set of yes/no rules, one for each possible candidate. There are two differences between their framework and ours. Barberà, Sonnenschein, and Zhou (1991) do not consider the preferences of the candidates as relevant for the social choice. Moreover, they do not consider the restrictions imposed by *individual rationality*, that we consider indispensable for the analysis of coalition formation rules.

Theorem 2 is tight if there are at least four agents. When there are only three agents, *flexibility* is directly implied by *individual rationality*. The following examples show the independence of the axioms for any arbitrary minimally rich domain $\bar{\mathcal{D}}$.[10]

Example 1 (*Strategy-proofness*)**.** Let $N = \{i, j, k\}$. For each $\succsim \in \bar{\mathcal{D}}$, let

$$IR_i(\succsim) \equiv \{C \in \mathcal{C}_i, \text{ such that for each } j \in C, C \succsim_j \{j\}\}.$$

Let φ^{-SP} be such that for each $\succsim \in \bar{\mathcal{D}}$, $\varphi_i^{-SP}(\succsim) \equiv \text{top}(IR_i(\succsim), \succsim_i)$ and for each $j \notin \text{top}(IR_i(\succsim), \succsim_i)$, $\varphi_j^{-SP}(\succsim) \equiv \{j\}$. Note that φ^{-SP} satisfies individual rationality, non-bossiness, and flexibility. However, φ^{-SP} violates strategy-proofness.[11]

Example 2 (*Individual rationality*)**.** Let $N = \{i, j, k\}$. Let φ^{-IR} be such that for each $\succsim \in \bar{\mathcal{D}}$, $\varphi_i^{-IR}(\succsim) = \text{top}(\mathcal{N}, \succsim_i)$, and for each $j \notin \text{top}(\mathcal{N}, \succsim_i)$, $\varphi_j^{-IR}(\succsim) = \{j\}$. The rule φ^{-IR} is dictatorial. Note that φ^{-IR} satisfies strategy-proofness, non-bossiness, and flexibility. However, φ^{-IR} violates individual rationality.

Example 3 (*Non-Bossiness*)**.** Let $N = \{i, j, k\}$. Let φ^{-NB} be such that for each $\succsim \in \bar{\mathcal{D}}$,

$$\varphi^{-NB}(\succsim) = \begin{cases} \{i, j, k\} & \text{if for each } i' \in N, \{i, j, k\} \succsim_{i'} \{i'\}, \\ (\{i, j\}, \{k\}) & \text{if } \{i, j\} \succ_i \{i\}, \{i, j\} \succ_j \{j\} \text{ and } \text{top}(\mathcal{N}, \succsim_k) = \{k\}, \\ [N] & \text{otherwise.} \end{cases}$$

---

[10]The following examples are stated in three and four-agent societies. These examples can be easily generalized to arbitrary societies. This point is discussed in Appendix B.

[11]In order to check that φ^{-SP} is *manipulable*, let $N = \{i, j, k\}$, $\succsim \in \mathcal{D}^*$, and $\succsim'_j \in \mathcal{D}_j^*$ be such that $\{i, j\} \succ_i \{i, j, k\} \succ_i \{i\}$, $\{i, j, k\} \succ_j \{i, j\} \succ_j \{j, k\} \succ_j \{j\}$, and $\{i, k\} \succ_k \{i, j, k\} \succ_k \{k\}$; while $\{j, k\} \succ'_j \{i, j, k\} \succ'_j \{j\}$. Note that $\varphi^{-SP}(\succsim) = (\{i, j\}, \{k\})$, while $\varphi^{-SP}(\succsim_{N \setminus \{j\}}, \succsim'_j) = \{i, j, k\}$. Then, $\varphi_j^{-SP}(\succsim_{N \setminus \{j\}}, \succsim'_j) \succ_j \varphi_j^{-SP}(\succsim)$.

Note that φ^{-NB} satisfies individual rationality, strategy-proofness, and flexibility. However, φ^{-NB} violates non-bossiness.[12]

Example 4 (*Flexibility*). *Let $N = \{i, j, k, l\}$. Let φ^{-F} be such that for each $\succsim \in \bar{\mathcal{D}}$,*

$$\varphi^{-F}(\succsim) = \begin{cases} (\{i,j\}, \{k,l\}) & \text{if for each } m \in N, \; (\{i,j\}, \{k,l\}) \succsim_m [N], \\ [N] & \text{otherwise.} \end{cases}$$

Note that φ^{-F} satisfies individual rationality, strategy-proofness, and non-bossiness. However, φ^{-F} violates flexibility.

At this point, we clarify the relation between *strategy-proofness*, *non-bossiness*, and *Pareto efficiency*. In many frameworks, *strategy-proofness* and *non-bossiness* directly imply *Pareto efficiency*. However, this is not the case in our framework.[13] On the other hand, although we cannot find a general and straight-forward argument that shows that *strategy-proofness*, *individual rationality*, and *Pareto efficiency*, imply *non-bossiness*, it turns out that the arguments in the proof of Theorem 2 are also valid (with minimal modifications) if we use *Pareto efficiency* instead of *non-bossiness* and *flexibility*. Hence, we can state the following theorem that parallels Theorem 2.

Theorem 3. *Let $\bar{\mathcal{D}}$ be a minimally rich domain. A rule $\varphi : \bar{\mathcal{D}} \to \Sigma$ satisfies strategy-proofness, individual rationality, and Pareto efficiency if and only if φ is a single-lapping rule.*

---

[12]In order to check that φ^{-NB} violates *non-bossiness*, let $\succsim \in \mathcal{D}^*$, $\succsim'_k \in \mathcal{D}^*_k$ be such that $\{i,j\} \succ_i \{i\}$, $\{i,j\} \succ_j \{j\}$, $\text{top}(\mathcal{N}, \succsim_k) = \{k\}$, while $\{j,k\} \succ'_k \{k\} \succ'_k \{i,j,k\}$. Note that $\varphi(\succsim) = (\{i,j\}, \{k\})$ and $\varphi(\succsim_{N\setminus\{k\}}, \succsim'_k) = (\{i\}, \{j\}, \{k\})$.

[13]Consider a society formed by four agents $N = \{i, j, k, l\}$. Define the rule $\bar{\varphi}$ in the domain of separable preferences. Let $\bar{\varphi} : \mathcal{S} \to \Sigma$. Agents i and j are the founding members of a club and they are always together. Then, for each $\succsim \in \mathcal{S}$, $\{i\} \in \bar{\varphi}_j(\succsim)$, $\{j\} \in \bar{\varphi}_i(\succsim)$. Preferences of agents k and l are irrelevant for the social choice. Agent k enters the club if i likes agent k. Thus, $\{k\} \in \bar{\varphi}_i(\succsim)$ if $\{i,k\} \succsim_i \{i\}$. Agent l enters the club if j likes l. Thus, $\{l\} \in \bar{\varphi}_j(\succsim)$ if $\{j,l\} \succsim_l \{j\}$. The rule $\bar{\varphi}$ satisfies *strategy-proofness*, *non-bossiness*, and *flexibility*. However, $\bar{\varphi}$ violates *individual rationality* and *Pareto efficiency*. Let $\succsim \in \mathcal{S}$ be such that $\{i,j,k\} \succ_i \{i,k\} \succ_i \{i,j\} \succ_i \{i\} \succ_i C$ for each $C \in \mathcal{C}_i \setminus (\{i,j,k\}, \{i,k\}, \{i,j\}, \{i\})$, $\{i,j,l\} \succ_j \{j,l\} \succ_j \{i,j\} \succ_j \{j\} \succ_j C'$ for each $C' \in \mathcal{C}_j \setminus (\{i,j,l\}, \{j,l\}, \{i,j\}, \{j\})$, $\{k\} = \text{top}(\mathcal{N}, \succsim_k)$, and $\{l\} = \text{top}(\mathcal{N}, \succsim_l)$. Basically, i likes j and k but strongly dislikes l, j likes i and l and strongly dislikes k, whereas k and l would rather stay alone. Note that $\bar{\varphi}(\succsim) = \{i,j,k,l\}$, but for each $i' \in N$, $(\{i,j\}, \{k\}, \{l\}) \succ_{i'} \bar{\varphi}(\succsim)$.

Theorem 3 shows that, when applied to *strategy-proof* and *individually rational* rules, *non-bossiness* and *flexibility* are equivalent to *Pareto efficiency*.[14] While sometimes *Pareto efficiency* may seem a more palatable axiom, we think that in coalition formation problems, *non-bossiness* is also easily justifiable. We have chosen to use *non-bossiness* instead of *Pareto efficiency* because *Pareto efficiency* is part of the definition of *core-stability*. We feel that by introducing *individual rationality*–no single agent prefers stay on her own rather than accepting the coalition proposed by the rule– together with *Pareto efficiency* –all the members of the society do not prefer an alternative partition to the partition proposed by the rule–, we would introduce too many ingredients of the core in our framework.

Before moving to the proof of Theorem 2, several remarks are in order.

Our characterization theorems have direct implications for matching problems as marriage, room-mate, and college admission problems. These are coalition formation problems characterized by initial restrictions of the set of feasible coalitions. In a *marriage problem*, agents are divided in two disjoint groups that are usually interpreted as a set of men and a set of women, and the set of feasible coalitions consists of all single agents and all the couples formed by a man and a woman. A generalization of marriage problems is known as the *roommate problem*. There is a set of agents that have to be organized in couples or in groups of a given cardinality. These problems can be interpreted as situations in which there are a number of rooms available and we can assign either 1 or 2 persons to each room, while some room may remain empty. Another generalization of the marriage problem is known as the *college-admission problem*. There are two disjoint sets of agents, a set of colleges C, and a set of new students S. Each college $c \in C$ has a number of free slots and may admit up to a quota of q_c new students. Colleges have preferences over cohorts of new students. New students have preferences over colleges and classmates. A coalition is feasible if either is a singleton or it contains exactly one college and the number of students assigned to each college is not larger than its respective quota q_c. The literature has provided a series of impossibility results for *strategy-proof* rules in those environments. (See Alcalde and Barberà, 1994; Sönmez, 1999; Pápai, 2004, and references therein.) As we do not impose any initial condition on the set of feasible, we can apply directly our

---

[14]Note that the examples that show the independence of the axioms in Theorem 2 are also valid to show the independence of the axioms in Theorem 3. It is easy to check that φ^{-SP}, φ^{-IR}, and φ^{-F} satisfy *Pareto efficiency*. However, φ^{-NB} violates *Pareto efficiency*.

Theorems 2 and 3 to obtain characterization of the rules that satisfy our axioms in these environments. Thus, for instance, in marriage problems, if there are at least two men and two women and every different-sex couple is feasible, then the set of feasible coalitions violates Condition (*b*) of the single-lapping property. In this case, Theorems 2 and 3 imply that we may construct single-lapping rules that satisfy our axioms by not allowing some couples to form.

Next, we address the issue of whether Theorems 2 and 3 hold for domains of preferences strictly contained in $\mathcal{D}^*$. As $\mathcal{D}^*$ consists of the union of the domains of bottom and top preferences, it is natural to check whether there exist non-single-lapping rules that satisfy our axioms in those domains. It turns out that new possibilities arise in both domains. The rule φ^{-SP} presented in Example 1 satisfies *strategy-proofness* when defined in the domain of bottom preferences $\times_{i \in N} \mathcal{D}_i^{-}$.[15] On the the other hand, the domain of top preferences is included in the domain of top-responsive preferences proposed by Alcalde and Revilla (2004). These authors provide an algorithm – the top-covering algorithm– that always select a core-stable partition of the society if agents' preferences are top-responsive. In addition, their top-covering algorithm defines the unique *Pareto efficient* rule that satisfies our axioms in their domain. We must note that in Alcalde and Revilla's top-responsive domain, there are preference profiles with multiple core stable partitions. This fact highlights the key role of bottom preferences in obtaining the relation between *strategy-proofness* and *unique core stability*.[16] In the light of these examples, we can interpret Theorems 2 and 3 as minimal domain results. The smallest minimally rich domain $\mathcal{D}^*$ is a minimal domain for which the single-lapping rules are the unique rules that satisfy *strategy-proofness*, *individual rationality*, and either *non-bossiness* and *flexibility*, or *Pareto efficiency*.

Finally, we conclude relating our results to those by Sönmez (1999). This author proves that for coalition formation problems in which there is always a core-stable partition, there is a rule that satisfies *strategy-proofness*, *individual rationality*, and *Pareto efficiency* if the set of core-stable partitions is always essentially single-valued. Our results are independent

---

[15] See Appendix B for additional details.

[16] Indeed, *individual rationality* is easily satisfied in the domain of top preferences. Note that any agent prefers to join the whole society rather than staying alone as long unless her best preferred coalition is staying alone. Finally, we have to note that Alcalde and Revilla's results depend crucially on the fact that every conceivable coalition is feasible. Their top-covering algorithm works if there are no restrictions in the set of feasible coalitions.

of Sönmez's results. The main difference between our framework and Sönmez's one relies on the domain of preferences over coalitions. Sönmez (1999) assumes the existence of certain preferences that need not exist in a minimally rich domain. Basically, in Sönmez's framework for each $i \in N$, and each $C \in (F^\varphi \cap \mathcal{C}_i)$, if there is an admissible preference $\succsim_i$ such that $C \succ_i \{i\}$, then there is another admissible preference $\succsim_i'$ such that for each $C' \in (F^\varphi \cap \mathcal{C}_i) \setminus \{i\}$, $C' \succsim_i' C$ if and only if $C' \succsim_i C$, while $C \succsim_i C'$ if and only if $C \succsim_i' C'$ and $C \succsim_i' \{i\} \succsim_i' C'$. There are minimally rich domains, namely the domain of additively representable preferences, for which such preferences are not admissible. Let $i, j, k \in N$, and assume $\{i, j\}, \{i, k\}, \{i, j, k\} \in F^\varphi$. Let $\succsim_i \in \mathcal{A}_i$ be such that $\{i, j, k\} \succ_i \{i, j\} \succ \{i, k\} \succ \{i\}$, but there is no $\succsim_i' \in \mathcal{A}_i$ such that $\{i, j, k\} \succ_i' \{i\}$, $\{i\} \succ_i' \{i, j\}$, and $\{i\} \succ_i' \{i, k\}$.

6   Proof of Theorem 2

We begin this section by introducing some properties that are implied by our axioms. These properties incorporate the idea that a rule cannot be against the preferences of the members of the society. When there is a partition that each agent considers at least as good as every other partition, a rule should choose that best-preferred partition. A stronger requirement would be that whenever the members of a coalition consider this coalition as the best coalition, this coalition should form, independently of the preferences of the remaining agents in society. Of course, the following axioms refer to rules defined on a minimally rich domain $\bar{\mathcal{D}}$.

Unanimity.   Let $\sigma = \{C_1, \ldots, C_m\} \in \Sigma$ be such that for each $t = 1, \ldots, m$, $C_t \in F^\varphi$. For each $\succsim \in \bar{\mathcal{D}}$, each $t = 1, \ldots, m$, and each $i \in C_t$, $\text{top}(F^\varphi, \succsim_i) = C_t$ implies $\varphi(\succsim) = \sigma$.

Top-Coalition.   Let $C \in F^\varphi$ and $\succsim \in \bar{\mathcal{D}}$. If for each $i \in C$, $\text{top}(F^\varphi, \succsim_i) = C$, then for each $i \in C$, $\varphi_i(\succsim) = C$.

It is clear that *top-coalition* and *Pareto efficiency* imply *unanimity*. However, *Pareto efficiency* and *top-coalition* are logically independent. Note that *top-coalition* is a property of rules. Banerjee, Konishi, and Sönmez (2001) use the term top-coalition to name a property of preference profiles. These authors say that a preference profile satisfies the top-coalition property if for every group of agents $V \subseteq N$ there is a coalition $C \subseteq V$ that is

mutually the best coalitions for all the members of C. Basically, our *top-coalition* implies that at a preference profile that satisfies Banerjee, Konishi, and Sönmez's top-coalition property, then the rule selects a partition in which all the coalitions such that all their members consider as the best feasible coalition are formed.

Lemma 1. *Let $\bar{\mathcal{D}}$ be a minimally rich domain. If a rule $\varphi : \bar{\mathcal{D}} \to \Sigma$ satisfies strategy-proofness, non-bossiness, and flexibility, then φ satisfies unanimity.*

Proof. Let $\sigma = \{C_1, \ldots, C_m\} \in \Sigma$ be such that for each $t = 1, \ldots, m$, $C_t \in F^\varphi$. Let $\succsim \in \bar{\mathcal{D}}$ be such that for each $t = 1, \ldots, m$ and each $i \in C_t$, $\mathrm{top}(F^\varphi, \succsim_i) = C_t$. By *flexibility*, $\sigma \in R^\varphi$. Then, there is $\succsim' \in \bar{\mathcal{D}}$, such that $\varphi(\succsim') = \sigma$. Let $i \in N$. Let $\succsim'' \in \bar{\mathcal{D}}$ be such that $\succsim''_i = \succsim_i$ while for each $j \in N \setminus \{i\}$, $\succsim''_j = \succsim'_j$. By *strategy-proofness*, $\varphi_i(\succsim'_{N \setminus \{i\}}, \succsim_i) \succsim_i \varphi_i(\succsim') = \mathrm{top}(F^\varphi, \succsim_i)$. Then, $\varphi_i(\succsim'_{N \setminus \{i\}}, \succsim_i) = \varphi_i(\succsim') = \mathrm{top}(F^\varphi, \succsim_i)$. By *non-bossiness*, $\varphi(\succsim'_{N \setminus \{i\}}, \succsim_i) = \varphi(\succsim')$. Repeating the argument as many times as necessary, we obtain $\varphi(\succsim) = \varphi(\succsim')$. $\square$

Lemma 2. *Let $\bar{\mathcal{D}}$ be a minimally rich domain. If a rule $\varphi : \bar{\mathcal{D}} \to \Sigma$ satisfies strategy-proofness, individual rationality, non-bossiness, and flexibility, then φ satisfies top-coalition.*

Proof. Let $C \in F^\varphi$. Let $\succsim \in \bar{\mathcal{D}}$ be such that for each $i \in C$, $\mathrm{top}(F^\varphi, \succsim_i) = C$. If $\#C = 1$, then the result follows from *individual rationality*. If $C = N$, then the result is immediate by *unanimity*. Let $\succsim' \in \bar{\mathcal{D}}$ be such that for each $i \in C$, $\mathrm{top}(F^\varphi, \succsim'_i) = C$, for each $C' \in \mathcal{C}_i$ such that there is $j \in (C' \setminus C)$, $\{i\} \succ_i C'$, and for each $k \notin C$, $\succsim_k = \succsim'_k$.[17] By *individual rationality*, for each $i \in C$, $\varphi_i(\succsim') \subseteq C$. Let $\succsim'' \in \mathcal{A}$ be such that for each $i \in C$, $\succsim'_i = \succsim''_i$ while for each $k \in (N \setminus C)$, $\varphi_k(\succsim') = \mathrm{top}(F^\varphi, \succsim''_k)$. By *strategy-proofness*, $\varphi_k(\succsim'_{N \setminus \{k\}}, \succsim''_k) = \varphi_k(\succsim')$. By *non-bossiness*, $\varphi(\succsim'_{N \setminus \{k\}}, \succsim''_k) = \varphi(\succsim')$. Repeating the arguments for each $k \in (N \setminus C)$, $\varphi(\succsim') = \varphi(\succsim'')$. By *unanimity*, for each $i \in C$, $\varphi_i(\succsim'') = C$. Then, $\varphi_i(\succsim') = C$. Finally, let $i \in C$. By *strategy-proofness*, $\varphi_i(\succsim'_{N \setminus \{i\}}, \succsim_i) \succsim_i \varphi_i(\succsim')$. Then, $\varphi_i(\succsim'_{N \setminus \{i\}}, \succsim_i) = C$. Repeating the argument as many times as necessary, we obtain that for each $i \in C$, $\varphi_i(\succsim) = C$. $\square$

In the following lemma we prove that agents' preferences over infeasible coalitions are irrelevant for the social choice.

---

[17]Note that $\times_{i \in N} \mathcal{D}_i^- \subset \bar{\mathcal{D}}$. Thus, $\succsim'_C \in \bar{\mathcal{D}}_C$.

Lemma 3. *Let $\bar{\mathcal{D}}$ be a minimally rich domain. If a rule $\varphi : \bar{\mathcal{D}} \to \Sigma$ satisfies strategy-proofness and non-bossiness, then, for each $\succsim, \succsim' \in \bar{\mathcal{D}}$ such that for each $i \in N$, and each $C, C' \in (F^{\varphi} \cap \mathcal{C}_i)$, $C \succ_i C'$ if and only if $C \succ'_i C'$, $\varphi(\succsim) = \varphi(\succsim')$.*

Proof. Let $\succsim, \succsim' \in \bar{\mathcal{D}}$ be such that for each $i \in N$, and each $C, C' \in (F^{\varphi} \cap \mathcal{C}_i)$, $C \succ_i C'$ if and only if $C \succ'_i C'$. Let $i \in N$. By *strategy-proofness*, $\varphi_i(\succsim_{N\setminus\{i\}}, \succsim'_i) \succsim'_i \varphi_i(\succsim)$ and $\varphi_i(\succsim) \succsim_i \varphi_i(\succsim_{N\setminus\{i\}}, \succsim'_i)$. Because for each $C, C' \in (F^{\varphi} \cap \mathcal{C}_i)$, $C \succ_i C'$ if and only if $C \succ'_i C'$, we have $\varphi_i(\succsim) = \varphi_i(\succsim_{N\setminus\{i\}}, \succsim'_i)$. By *non-bossiness*, $\varphi(\succsim) = \varphi(\succsim_{N\setminus\{i\}}, \succsim'_i)$. Repeating the argument as many times as necessary, we get $\varphi(\succsim) = \varphi(\succsim')$. $\qquad\square$

The following lemma presents the crucial step in the proof of Theorem 2.

Lemma 4. *Let $\bar{\mathcal{D}}$ be a minimally rich domain. If a rule $\varphi : \bar{\mathcal{D}} \to \Sigma$ satisfies strategy-proofness, individual rationality, non-bossiness, and flexibility, then F^{φ} satisfies the single-lapping property.*

Proof. The proof is by induction on the number of agents. We first focus on three-agent societies. Then, we extend the result to arbitrary societies. We use extensively throughout the proof the fact that $\mathcal{D}^* \subseteq \bar{\mathcal{D}}$.

Claim 1. Let $n = 3$, then F^{φ} satisfies Condition (a) of the single-lapping property.

Proof. Let $N = \{i, j, k\}$. Assume to the contrary that F^{φ} does not satisfy Condition (a). Then, there are $C, C' \in F^{\varphi}$ such that $\#(C \cap C') \geq 2$. We have two cases.

Case (1.1): $F^{\varphi} = \{\{i\}, \{j\}, \{k\}, \{i, j\}, \{i, j, k\}\}$.

Let $\bar{\succsim}_k \in \mathcal{D}^*_k$ be such that $\{i, j, k\} \succ_k \{i, k\} \succ_k \{j, k\} \succ_k \{k\}$. Let the rule $\bar{\varphi}^{\{i,j\}} : \bar{\mathcal{D}}_{\{i,j\}} \to \Sigma$ be such that for each $\succsim_{\{i,j\}} \in \bar{\mathcal{D}}_{\{i,j\}}$, $\bar{\varphi}^{\{i,j\}}(\succsim_{\{i,j\}}) \equiv \varphi(\succsim_{\{i,j\}}, \bar{\succsim}_k)$. By φ's *strategy-proofness*, $\bar{\varphi}^{\{i,j\}}$ satisfies *strategy-proofness*. By φ's *top-coalition*,

$$R^{\bar{\varphi}^{\{i,j\}}} = \{(\{i\}, \{j\}, \{k\}), (\{i, j\}, \{k\}), \{i, j, k\}\}.$$

By Remark 1, agent i and agent j's preferences over the partitions in $R^{\bar{\varphi}^{\{i,j\}}}$ are unrestricted. Hence, $\bar{\varphi}^{\{i,j\}}$ satisfies *strategy-proofness*, its range contains three elements, and

21

agents' preferences over the elements of the range are unrestricted. Then, by the Gibbard-Satterthwaite Theorem, $\bar{\varphi}^{\{i,j\}}$ is *dictatorial*. Assume that i is the dictator for $\bar{\varphi}^{\{i,j\}}$. Let $\succsim_{\{i,j\}} \in \mathcal{D}^*_{\{i,j\}}$ be such that $\{i,j,k\} \succ_i \{i,j\} \succ_i \{i\}$ and $\{j\} \succ_j \{i,j\} \succ_j \{i,j,k\}$. Then, $\varphi(\succsim_{\{i,j\}}, \bar{\succsim}_k) = \{i,j,k\}$, but $\{j\} \succ_j \varphi_j(\succsim)$, which violates *individual rationality*, a contradiction.

Case (**1.2**) $\{\{i\}, \{j\}, \{k\}, \{i,j\}, \{j,k\}, \{i,j,k\}\} \subseteq F^{\varphi}$.

Let $\succsim^1 \in \mathcal{D}^*$ be such that,

$$
\begin{array}{ccc}
\underline{\succsim^1_i:} & \underline{\succsim^1_j:} & \underline{\succsim^1_k:} \\
\{i,j\} & \{i,j\} & \{j,k\} \\
\{i\} & \{j\} & \{i,j,k\} \\
\{i,j,k\} & \{i,j,k\} & \{k\} \\
\{i,k\} & \{j,k\} & \{i,k\}
\end{array}
$$

By *top-coalition*, $\varphi\left(\succsim^1\right) = (\{i,j\}, \{k\})$.

Let $\succsim^2 \in \mathcal{D}^*$ be such that $\succsim^2_{N\setminus\{i\}} = \succsim^1_{N\setminus\{i\}}$ and $\{i,j,k\} \succsim^2_i \{i,j\} \succsim^2_i \{i,k\} \succsim^2_i \{i\}$. By *strategy-proofness*, $\varphi_i(\succsim^2) \succsim^2_i \varphi_i(\succsim^1)$. Then, $\varphi_i(\succsim^2)$ is either $\{i,j,k\}$ or $\{i,j\}$. Because $\{j\} \succ^2_j \{i,j,k\}$, by *individual rationality*, $\varphi_i(\succsim^2) = \{i,j\}$. Then, by *non-bossiness*, $\varphi(\succsim^2) = \varphi(\succsim^1)$.

Let $\succsim^3 \in \mathcal{D}^*$ be such that $\succsim^3_{N\setminus\{j\}} = \succsim^2_{N\setminus\{j\}}$ and $\{i,j\} \succsim^3_j \{i,j,k\} \succsim^3_j \{j\}$. By *strategy-proofness*, $\varphi_j\left(\succsim^3\right) \succsim^3_j \varphi_j(\succsim^2)$. Then, $\varphi_j(\succsim^3) = \{i,j\}$. By *non-bossiness*, $\varphi\left(\succsim^3\right) = \varphi\left(\succsim^2\right)$.

Now, let $\succsim^4 \in \mathcal{D}^*$ be such that $\succsim^4_{N\setminus\{i\}} = \succsim^3_{N\setminus\{i\}}$ and $\{i,k\} \succsim^4_i \{i,j,k\} \succsim^4_i \{i\}$. Then,

$$
\begin{array}{ccc}
\underline{\succsim^4_i:} & \underline{\succsim^4_j:} & \underline{\succsim^4_k:} \\
\{i,k\} & \{i,j\} & \{j,k\} \\
\{i,j,k\} & \{i,j,k\} & \{i,j,k\} \\
\{i\} & \{j\} & \{k\} \\
\{i,j\} & \{j,k\} & \{i,k\}
\end{array}
$$

By *individual rationality*, $\varphi_i(\succsim^4) \neq \{i,j\}$, $\varphi_k(\succsim^4) \neq \{i,k\}$, and $\varphi_j(\succsim^4) \neq \{j,k\}$. By *strategy-proofness*, $\varphi_i(\succsim^3) \succsim^3_i \varphi(\succsim^4)$. Note that, $\{i,j,k\} \succ^3_i \varphi\left(\succsim^3\right)$. Then, $\varphi\left(\succsim^4\right) = (\{i\}, \{j\}, \{k\})$.

Let $\succsim^5 \in \mathcal{D}^*$ be such that $\succsim_i^5 = \succsim_i^4$, $\{j,k\} \succ_j^5 \{j\} \succ_j^5 \{i,j,k\} \succsim_j^5 \{i,j\}$, and $\{i,j,k\} \succsim_k^5$ $\{j,k\} \succsim_k^5 \{i,k\} \succsim_k^5 \{k\}$. By *top-coalition*, $\varphi_k(\succsim_{N\setminus\{k\}}^5, \succsim_k^4) = \{j,k\}$. By *strategy-proofness*, $\varphi_k(\succsim^5) \succsim_k^5 \{j,k\}$. Because $\{j\} \succ_j^5 \{i,j,k\}$, by *individual rationality*, $\varphi(\succsim^5) = (\{i\}, \{j,k\})$.

Let $\succsim^6 \in \mathcal{D}^*$ be such that $\succsim_{N\setminus\{j\}}^6 = \succsim_{N\setminus\{j\}}^5$ and $\{i,j,k\} \succsim_j^6 \{j,k\} \succsim_j^6 \{i,j\} \succsim_j^6 \{j\}$. Note that, by *unanimity*, $\varphi(\succsim_{N\setminus\{i\}}^6, \succsim_i^3) = \{i,j,k\}$. Hence, by *strategy-proofness*, $\varphi_i(\succsim^6) \succsim_i \{i,j,k\}$. Then, $\varphi(\succsim^6) = \{i,j,k\}$.

Finally, let $\succsim^7 \in \mathcal{D}^*$ be such that $\succsim_{N\setminus\{j\}}^7 = \succsim_{N\setminus\{j\}}^6$ and $\succsim_j^7 = \succsim_j^4$. Then

$$
\begin{array}{ccc}
\succsim_i^7: & \succsim_j^7: & \succsim_k^7: \\
\{i,k\} & \{i,j\} & \{i,j,k\} \\
\{i,j,k\} & \{i,j,k\} & \{j,k\} \\
\{i\} & \{j\} & \{i,k\} \\
\{i,j\} & \{j,k\} & \{k\}
\end{array}
$$

Note that the only difference between $\succsim^4$ and $\succsim^7$ consists of k's preference. By *strategy-proofness*, $\varphi_j(\succsim^7) \succsim_j^7 \varphi_j(\succsim^6) = \{i,j,k\}$. By *individual rationality*, if $j \in \varphi_i(\succsim^7)$, then $\varphi_i(\succsim^7) = \{i,j,k\}$. Hence, $\varphi(\succsim^7) = \{i,j,k\}$. However, $\varphi_k(\succsim^7) \succ_k^4 \varphi_k(\succsim^4)$, which violates *strategy-proofness*, a contradiction.

Cases (1.1) and (1.2) exhaust (up to a relabelling the agents) all the possibilities. Then, F^φ satisfies Condition (a), which concludes the proof of Claim 1. $\square$

Claim 2. Let $n = 3$, then F^φ satisfies Condition (b) of the single-lapping property.

Proof. Assume, to the contrary, that F^φ does not satisfy Condition (b). Then, there is a list of coalitions $\{C_1, \ldots, C_m\} \subset F^\varphi$, with $m \geq 3$ and $m + 1 = 1$ such that for each $t = 1, \ldots, m$, $\#(C_t \cap C_{t+1}) \geq 1$ and for no $i \in N$, $(C_t \cap C_{t+1}) = \{i\}$. By Claim 1, φ satisfies Condition (a). Then, we have $F^\varphi = \{\{i\}, \{j\}, \{k\}, \{i,j\}, \{j,k\}, \{i,k\}\}$. Thus, for each $\succsim \in \bar{\mathcal{D}}$, there is $i' \in \{i,j,k\}$ such that

$$\varphi_{i'}(\succsim) = \{i'\} \tag{*}$$

Let $\succsim \in \mathcal{D}^*$ be such that $\{i,j\} \succ_i \{i,k\} \succ_i \{i\}$,[18] $\{j,k\} \succ_j \{i,j\} \succ_j \{j\}$, and $\{i,k\} \succ_k \{j,k\} \succ_k \{k\}$. Let $P \in \mathcal{P}$ be such that $k\ P\ i\ P\ j$, and let $\succsim_i' \in \mathcal{D}_i^*$ be

---

[18]Note that by Lemma 3, we only need specify agents' preferences over feasible coalitions.

$\succsim'_i = \succsim_i^-$ (P). Because $\text{top}(F^\varphi, \succsim'_i) = \text{top}(F^\varphi, \succsim_k) = \{i, k\}$, by *top-coalition*, we have that $\varphi(\succsim_{N \setminus \{i\}}, \succsim'_i) = (\{i, k\}, \{j\})$. By *strategy-proofness*, $\varphi_i(\succsim) \succ_i \varphi(\succsim_{N \setminus \{i\}}, \succsim'_i)$. Then, we have that $\varphi_i(\succsim) \neq \{i\}$. Using parallel arguments, we get $\varphi_j(\succsim) \neq \{j\}$ and $\varphi_k(\succsim) \neq \{k\}$, which contradicts (*) and concludes the proof of Claim 2. $\square$

Now, we extend the result to arbitrary finite societies.

Induction Step. There is $m \geq 3$ such that for $n = m$, if the n-agent rule φ satisfies *strategy-proofness*, *individual rationality*, *non-bossiness*, and *flexibility*, then F^φ satisfies the single-lapping property. We prove that this is true for $n = m + 1$.

By Claims 1 and 2, the induction hypothesis is true for $m = 3$. Let $n = m+1$. Assume that φ satisfies *strategy-proofness*, *individual rationality*, *non-bossiness*, and *flexibility*. First, we prove two facts.

Fact 1. For each $C, C' \in F^\varphi$ such that $C \cup C' \neq N$, $\#(C \cap C') = 1$.

Proof. Let $C, C' \in F^\varphi$ be such that $(C \cup C') \neq N$. Let $j \in N \setminus (C \cup C')$. Let $\bar{\succsim}_j \in \mathcal{D}_j^*$ be such that for each $C \in \mathcal{C}_j$, $C \neq \{j\}$, $\{j\} \succ_j C$. Let $\Sigma_{N \setminus \{j\}}$ denote all the partitions of the reduced society $N \setminus \{j\}$. Define the rule $\bar{\varphi}^{N \setminus \{j\}} : \bar{\mathcal{D}}_{N \setminus \{j\}} \to \Sigma_{N \setminus \{j\}}$ in such a way that for each $\succsim_{N \setminus \{j\}} \in \bar{\mathcal{D}}_{N \setminus \{j\}}$, $(\bar{\varphi}^{N \setminus \{j\}}(\succsim_{N \setminus \{j\}}), \{j\}) \equiv \varphi(\succsim_{N \setminus \{j\}}, \bar{\succsim}_j)$. By φ's *strategy-proofness*, *individual rationality*, *non-bossiness*, and *flexibility*, $\bar{\varphi}^{N \setminus \{j\}}$ satisfies *strategy-proofness*, *individual rationality*, *non-bossiness*, and *flexibility*. By the induction hypothesis, $F^{\bar{\varphi}^{N \setminus \{j\}}}$ satisfies the *single-lapping* property. By φ's *flexibility*, $C, C' \in F^{\bar{\varphi}^{N \setminus \{j\}}}$, then $\#(C \cap C') = 1$. $\square$

Similar arguments apply to prove the following fact.

Fact 2. For each $\{C_1, \ldots, C_m\} \subseteq \Pi$ with $m \geq 3$, $\cup_{t=1}^m C_t \neq N$, and for each $t = 1, \ldots, m$, $\#(C_t \cap C_{t+1}) \geq 1$ (where $m + 1 = 1$), there is $i \in N$ such that for each $t = 1, \ldots, m$, $C_t \cap C_{t+1} = \{i\}$.

Claim 1′. F^φ satisfies Condition (a).

Proof. Assume, to the contrary, that there are $C, C' \in F^\varphi$ such that $(C \cup C') = N$, and $\#(C \cap C') \geq 2$. We replicate the arguments of three-agent societies. There are three cases:

Case (1.0$'$) Let $C, C' \neq N$.

By Fact 1, either $F^\varphi = \{[N], C, C'\}$, or $F^\varphi = \{[N], C, C', N\}$. Let $\bar{\succsim}_{N\setminus(C\cap C')} \in \mathcal{D}^*_{N\setminus(C\cap C')}$ be such that for each $j \in (C \setminus C')$, $\text{top}(F^\varphi, \succsim_j) = C$, whereas for each $k \in (C' \setminus C)$, $\text{top}(F^\varphi, \succsim_k) = C'$. Define the rule $\bar\varphi^{C\cap C'} : \bar{\mathcal{D}}_{C\cap C'} \to \Sigma$ in such a way that for each $\succsim_{C\cap C'} \in \bar{\mathcal{D}}_{C\cap C'}$, $\bar\varphi^{C\cap C'}(\succsim_{C\cap C'}) \equiv \varphi(\succsim_{C\cap C'}, \bar{\succsim}_{N\setminus(C\cap C')})$. Because φ is *strategy-proof*, $\bar\varphi^{C\cap C'}$ is *strategy-proof*. Moreover, by *top-coalition*, $R^{\bar\varphi^{C\cap C'}} = \{[N], (C, [C' \setminus C]), (C', [C \setminus C'])\}$. By Remark 1, the preferences of the agents in $(C \cap C')$ over the partitions in $R^{\bar\varphi^{C\cap C'}}$ are not restricted. By the Gibbard-Satterthwaite Theorem, $\varphi^{C\cap C'}$ is *dictatorial*. Let $i \in (C \cap C')$ be a dictator for $\varphi^{C\cap C'}$. Let $\succsim_{C\cap C'} \in \mathcal{D}^*_{C\cap C'}$ be such that $\text{top}(F^{\bar\varphi^{C\cap C'}}, \succsim_i) = C'$, and for each $j \in (C\cap C')\setminus\{i\}$, $\text{top}(F^{\bar\varphi^{C\cap C'}}, \succsim_j) = \{j\}$. Then, $\varphi(\succsim_{C\cap C'}, \bar{\succsim}_{N\setminus(C\cup C')}) = (C', [C\setminus C'])$, which violates φ's *individual rationality*, a contradiction.

Case (1.1$'$) Let $C' = N$, and for no $j \in C$ there is $k \in N \setminus C$ and $C'' \subset N$ with $C'' \in F^\varphi$ such that $\{j, k\} \subseteq C''$.

Let $\succsim_{N\setminus C} \in \mathcal{D}^*_{N\setminus C}$ be such that for each $j \in (N\setminus C)$, $\text{top}(F^\varphi, \bar{\succ}_j) = N$. Define now the rule $\bar\varphi^C : \bar{\mathcal{D}}_C \to \Sigma$ in the following way. For each $\succsim_C \in \bar{\mathcal{D}}^*$, $\bar\varphi^C(\succ_C) \equiv \varphi(\succsim_C, \bar{\succsim}_{N\setminus C})$. Clearly, $\bar\varphi^C$ satisfies *strategy-proofness*. Moreover, by *top-coalition*, for each $i \in C$, $F^\varphi \cap \mathcal{C}_i = F^{\bar\varphi^C}$. Hence, by Remark 1, the preferences of the agents in C over partitions in $R^{\bar\varphi^C}$ are not restricted. By the Gibbard-Satterthwaite Theorem, $\bar\varphi^C$ is *dictatorial*, which, by an already familiar argument, violates φ's *individual rationality*, a contradiction.

Case (1.2$'$) Let $C' = N$, and for some $j \in C$ there is $k \in N \setminus C$ and $C'' \subset N$ with $C'' \in F^\varphi$ such that $\{j, k\} \subseteq C''$.

Note first that, by Fact 1, for each $C'' \in (F^\varphi \setminus N)$, $\#(C \cap C'') \leq 1$. Moreover, by Fact 2, there is no cycle of three coalitions in F^φ that does not involve the grand coalition N.

Let $C, N \in F^\varphi$, let $j \in C$ be such that for some $T \subseteq N \setminus C$, $T \cup \{j\} \in F^\varphi$. Let $\bar{C} \equiv C \setminus \{j\}$. Let $T' \in F^\varphi \setminus \{C, N\}$. By *Fact 1*, there is no $i \in C \setminus \{j\}$, such that $\{i, j\} \subseteq T'$. By *Fact 2*, there is no $k \in T$ such that $\{i, k\} \subseteq T'$.

Let $\succsim^1 \in \mathcal{D}^*$ be such that for each $i \in \bar{C}$, there is $P_i^1 \in \mathcal{P}$ with $j = \max(N, P_i^1)$, $N_i^+(P) = C$, and $\succsim_i^1 = \succsim_i^- (P_i^1)$, for j there is $P_j^1 \in \mathcal{P}$ with $N_j^+(P) = C$, and $\succsim_j^1 = \succsim_j^- (P_j^1)$, while for each $k \in N \setminus C$, there is $P_k^1 \in \mathcal{P}$ with $N_k^+(P) = \{j\} \cup \{k\}$, and $\succsim_k^1 = \succsim_k^+ (P_k^1)$. By *top-coalition*, for each $i \in C$, $\varphi_i(\succsim^1) = C$.

Next, let $\succsim^2 \in \mathcal{D}^*$ be such that $\succsim_{N \setminus \bar{C}}^1 = \succsim_{N \setminus \bar{C}}^2$, while for each $i \in \bar{C}$ there is $P_i^2 \in \mathcal{P}$ such that $j = \max(N, P_i^2)$, $N_i^+(P_i^2) = N$, and $\succsim_i^2 = \succsim_i^+ (P_i^2)$. Note that for each $i \in \bar{C}$, $N = \text{top}(F^\varphi, \succsim_i^2)$ and $C = \text{top}(F^\varphi \setminus N, \succsim_i^2)$. Let $i \in \bar{C}$, by *strategy-proofness*, $\varphi_i(\succsim_{N \setminus \{i\}}^1, \succsim_i^2) \succsim_i^2 \varphi_i(\succsim^1) = C$. By *individual rationality*, $\varphi_j(\succsim_{N \setminus \{i\}}^1, \succsim_i^2) \neq N$. Then, $\varphi_i(\succsim_{N \setminus \{i\}}^1, \succsim_i^2) = C$. By *non-bossiness*, $\varphi(\succsim_{N \setminus \{i\}}^1, \succsim_i^2) = \varphi(\succsim^1)$. Repeating the same argument iteratively with each $i \in \bar{C}$, we get $\varphi(\succsim^2) = \varphi(\succsim^1)$.

Let $\succsim^3 \in \mathcal{D}^*$ be such that $\succsim_{N \setminus \{j\}}^2 = \succsim_{N \setminus \{j\}}^3$ and $\succsim_j^3 = \succsim_j^+ (P_j^1)$. Note that $\text{top}(F^\varphi, \succsim_j^3) = C$. By *strategy-proofness*, $\varphi_j(\succsim^3) \succsim_j^3 \varphi(\succsim^2)$. Then, $\varphi_j(\succsim^3) = C$, and by *non-bossiness*, $\varphi(\succsim^3) = \varphi(\succsim^2)$.

Let $\succsim^4 \in \mathcal{D}^*$ be such that $\succsim_{N \setminus \bar{C}}^3 = \succsim_{N \setminus \bar{C}}^4$, while for each $i \in \bar{C}$ there is $P_i^4 \in \mathcal{P}$ such that for some $\bar{k} \in T$, $\max(N, P_i^4) = \bar{k}$, $N_i^+(P_i^4) = T \cup \{i\}$, and $\succsim_i^4 = \succsim_i^+ (P_i^4)$. Note that by *Fact 2* and our assumptions on F^φ, for each $i \in \bar{C}$, $\text{top}(F^\varphi, P_i^4) = N$, and for each $C \in F^\varphi \cap \mathcal{C}_i$, if $C \neq N$, then $\{i\} \succsim_i^4 C$. Let $i \in \bar{C}$, by *strategy-proofness*, $\varphi_i(\succsim^3) \succsim_i^3 \varphi_i(\succsim_{N \setminus \{i\}}^3, \succsim_i^4)$. Hence, $\varphi_i(\succsim_{N \setminus \{i\}}^3, \succsim_i^4) \neq N$. Repeating the argument for each $i \in \bar{C}$, we obtain that $\varphi(\succsim^4) \neq N$. Clearly, for each $i \in \bar{C}$, N is the only coalition in F^φ that is preferred to staying on her own. On the other hand, for agent j, the coalitions that are preferred to staying alone include some member of $\bar{C}$. Finally, each agent $k \in N \setminus C$ requires the presence of agent j in order to consider a coalition better than staying on her own. Then, by *individual rationality*, we have that $\varphi(\succsim^4) = [N]$.

Consider now the profile $\succsim^5 \in \mathcal{D}^*$, such that for each $i \in \bar{C}$, $\succsim_i^5 = \succsim_i^4$, for some $P_j^5 \in \mathcal{P}$ such that there is $\bar{k} \in T$, with $\max(N, P_j^5) = \bar{k}$ and $N_j^+(P_j^5) = N$, and $\succsim_j^5 = \succsim_j^+ (P_j^5)$, while for each $k \in N \setminus C$, there is $P_k^5 \in \mathcal{P}$ such that $j = \max(N, P_k^5)$, $N = N_k^+(P_k^5)$, and $\succsim_k^5 = \succsim_k^+ (P_k^5)$. By *unanimity*, $\varphi(\succsim^5) = N$.

Finally, let $\succsim^6 \in \mathcal{D}^*$, be such that for each $\succsim_C^6 = \succsim_C^4$, while $\succsim_{N \setminus C}^6 = \succsim_{N \setminus C}^5$. That is, we only change agent j's preferences with respect to the previous profile. By *strategy-proofness*, $\varphi_j(\succsim^6) \succsim^6 \varphi_j(\succsim^5) = N$. By *individual rationality*, for each $i \in \bar{C}$, if $j \in \varphi_i(\succsim^6)$,

then $\varphi_i(\succsim^6) = N$. Hence, $\varphi(\succsim^6) = N$. Clearly, $\succsim^6$ only differs from $\succsim^4$ in the preferences of the agents who belong to $N \setminus C$. Let $k \in N \setminus T$. By *strategy-proofness*, we have that $\varphi_k(\succsim^6_{N \setminus \{k\}}, \succsim^4_k) \succsim^4_k \varphi_k(\succsim^6) = N$. Then, $j \in \varphi_k(\succsim^6_{N \setminus \{k\}}, \succsim^4_k)$. By *individual rationality*, there is $i \in \bar{C}$ such that $i \in \varphi_j(\succsim^6_{N \setminus \{k\}}, \succsim^4_k)$. By Fact 1 and our assumptions over F^φ, $\varphi(\succsim^6_{N \setminus \{k\}}, \succsim^4_k) = N$. Repeating the argument as many times as necessary, we get that $\varphi(\succsim^4) = N$, a contradiction.

Cases $(1.0')$, $(1.1')$ and $(1.2')$ exhaust all the possibilities. Then, this suffices to prove that F^φ satisfies Condition (a). $\square$

Claim 2′. F^φ satisfies Condition (b).

Proof. Assume, to the contrary, that φ does not satisfy Condition (b). Then, there is a list of coalitions $\{C_1, \ldots, C_m\}$, with $m \geq 3$ such that for each $t = 1, \ldots, m$, $(m+1 = 1)$, $(C_t \cap C_{t+1}) \neq \{\varnothing\}$, and there is no $i \in N$ such that for each $t = 1, \ldots, m$, $\{i\} = (C_t \cap C_{t+1})$. Because we have just proved that F^φ satisfies Condition (a) of the single-lapping property, we have that for each $t = 1, \ldots, m$, $\#(C_t \cap C_{t+1}) = 1$. By Fact 2, $\cup_{t=1}^m C_t = N$. Moreover, $F^\varphi = \{C_1, \ldots, C_m\} \cup [N]$.

For each $t = 1, \ldots, m$, let $i_t \equiv (C_t \cap C_{t+1})$. Note that for each $t = 1, \ldots, m$ and each $j \in (C_t \setminus \{i_{t-1}, i_t\})$, $F^\varphi \cap \mathcal{C}_j = \{C_t, \{j\}\}$. On the other hand, for each $t = 1, \ldots, m$, $F^\varphi \cap \mathcal{C}_{i_t} = \{C_t, C_{t+1}, \{i_t\}\}$. Then, by Remark 1, minimal richness of the domain of preferences does not introduce any restriction on how the agents may order the different coalitions they may belong to. From now on, we only describe agents' preferences over feasible coalitions.

For each $t = 1, \ldots, m$, let $i_t \equiv (C_t \cap C_{t+1})$. Let $\succsim \in \bar{\mathcal{D}}$ be such that for each $t = 1, \ldots, m$ and each $j \in (C_t \setminus \{i_{t-1}, i_t\})$, $\text{top}(F^\varphi, \succsim_j) = C_t$, and for each $t = 1, \ldots, m$, $\text{top}(F^\varphi, \succsim_{i_t}) = C_{t+1}$, and $C_t \succ_{i_t} \{i_t\}$. Let $t \in \{1, \ldots, m\}$. Let $\succsim'_{i_t} \in \bar{\mathcal{D}}_{i_t}$ be such that $\text{top}(F^\varphi, \succsim'_{i_t}) = C_t$. By *top-coalition*, $\varphi_{i_t}(\succsim_{N \setminus \{i_t\}}, \succsim'_{i_t}) = C_t$. By *strategy-proofness*, $\varphi_{i_t}(\succsim) \succsim_{i_t} \varphi_{i_t}(\succsim_{N \setminus \{i_t\}}, \succsim'_{i_t})$. Thus, for each $t = 1, \ldots, m$; $\varphi_{i_t}(\succsim) \succsim_{i_t} C_t$.

Assume first that m is odd. Then, there is $t' \in \{1, \ldots, m\}$ such that $\varphi_{i_{t'}}(\succsim) = \{i_{t'}\}$, a contradiction with $\varphi_{i_t}(\succsim) \succsim_{i_t} C_t$ for each $t = 1, \ldots, m$.

Assume now that m is even. Without loss of generality, assume that for each t odd, $\varphi_{i_t}(\succsim) = C_{t+1}$ and for each t' even, $\varphi_{i_{t'}}(\succsim) = C_{t'}$. Let $\bar{t}$ be even. Let $P_{\bar{t}} \in \mathcal{P}$ be such that

27

$N_{\bar{t}}^+(P_{\bar{t}}) = C_{\bar{t}+1}$. Let $\succsim'_{i_{\bar{t}}}=\succsim_{\bar{t}}^- (P_{\bar{t}}) \in \mathcal{D}_{i_{\bar{t}}}^*$. Note that $\mathrm{top}(F^\varphi, \succsim'_{i_{\bar{t}}}) = C_{\bar{t}+1}$ and for each $T \nsubseteq C_{\bar{t}+1}$, $\{i_{\bar{t}}\} \succ'_{i_{\bar{t}}} T$. By *individual rationality*, $\varphi_{i_{\bar{t}}}(\succsim_{N\setminus\{i_{\bar{t}}\}}, \succsim'_{i_{\bar{t}}}) \neq C_{\bar{t}}$. Let $\succsim'_{i_{\bar{t}-1}} \in \mathcal{D}_{i_{\bar{t}-1}}^*$ be such that $\mathrm{top}(F^\varphi, \succsim'_{i_{\bar{t}-1}}) = C_{\bar{t}-1}$. By *top-coalition*, $\varphi_{i_{\bar{t}-1}}(\succsim_{N\setminus\{i_{\bar{t}-1}, i_{\bar{t}}\}}, \succsim'_{\{i_{\bar{t}-1}, i_{\bar{t}}\}}) = C_{\bar{t}-1}$. By *strategy-proofness*, we have that $\varphi_{i_{\bar{t}-1}}(\succsim_{N\setminus\{i_{\bar{t}}\}}, \succsim'_{\{i_{\bar{t}}\}}) \succsim_{i_{\bar{t}-1}} \varphi_{i_{\bar{t}-1}}(\succsim_{N\setminus\{i_{\bar{t}-1}, i_{\bar{t}}\}}, \succsim'_{\{i_{\bar{t}-1}, i_{\bar{t}}\}})$. Then, $\varphi_{i_{\bar{t}-1}}(\succsim_{N\setminus\{i_{\bar{t}}\}}, \succsim'_{i_{\bar{t}}}) = C_{\bar{t}-1}$, and $\varphi_{i_{\bar{t}-2}}(\succsim_{N\setminus\{i_{\bar{t}}\}}, \succsim'_{i_{\bar{t}}}) = C_{\bar{t}-1}$. Repeating the argument as many times as necessary, for each t odd, $\varphi_{i_t}(\succsim_{N\setminus\{i_{\bar{t}}\}}, \succsim'_{i_{\bar{t}}}) = C_t$, while for each t' even $\varphi_{i_{t'}}(\succsim_{N\setminus\{i_{\bar{t}}\}}, \succsim'_{i_{\bar{t}}}) = C_{t'+1}$, and $\varphi_{i_{\bar{t}}}(\succsim_{N\setminus\{i_{\bar{t}}\}}, \succsim'_{i_{\bar{t}}}) = C_{\bar{t}+1}$. Then, we get $\varphi_{i_{\bar{t}}}(\succsim_{N\setminus\{i_{\bar{t}}\}}, \succsim'_{i_{\bar{t}}}) \succ_{i_{\bar{t}}} \varphi_{i_{\bar{t}}}(\succsim)$, which violates *strategy-proofness*, and suffices to prove Claim 2$'$ and Lemma 4. $\qquad\square$

Proof of Theorem 2. From Theorem 1, every single-lapping rule satisfies *strategy-proofness*, *individual rationality*, *non-bossiness*, and *flexibility*. Hence, we focus on the converse implication. Let φ satisfy *strategy-proofness*, *individual rationality*, *non-bossiness*, and *flexibility*. By Lemma 2, φ satisfies *top-coalition*. Let $\succsim \in \bar{\mathcal{D}}$. By Lemma 4, F^φ satisfies the single-lapping property. Thus, there is $C \in F^\varphi$ such that for each $i \in C$, $\mathrm{top}(F^\varphi, \succsim_i) = C$. By *top-coalition*, for each $i \in C$, $\varphi_i(\succsim) = C$. Moreover, by *top-coalition*, for each $\succsim' \in \bar{\mathcal{D}}$ such that $\succsim_C = \succsim'_C$, for each $i \in C$, $\varphi_i(\succsim') = C$. Let $\Sigma_{N\setminus C}$ denote the set of all possible partitions of the reduced society $N \setminus C$. Define now the restricted social choice function $\bar{\varphi}^{N\setminus C} : \bar{\mathcal{D}}_{N\setminus C} \to \Sigma_{N\setminus C}$, in such a way that for each $\succsim_{N\setminus C} \in \bar{\mathcal{D}}_{N\setminus C}$, $(\bar{\varphi}^{N\setminus C}(\succsim_{N\setminus C}), C) \equiv \varphi(\succsim_{N\setminus C}, \succsim_C)$. Clearly, $\bar{\varphi}^{N\setminus C}$ satisfies *strategy-proofness*, *individual rationality*, *non-bossiness*, and *flexibility*. Moreover, $F^{\bar{\varphi}^{N\setminus C}} = \{C' \in F^\varphi, C \cap C' = \{\varnothing\}\}$, and $F^{\bar{\varphi}^{N\setminus C}}$ satisfies the single-lapping property. Repeating the same arguments as many times as necessary, we get $\varphi(\succsim) = \bar{\sigma}^{F^\varphi}(\succsim)$. $\qquad\square$

References

ALCALDE, J., AND S. BARBERÀ (1994): "Top Dominance and the Possibility of Stable Rules for Matching Problems," *Econ. Theory*, 4, 417–425.

ALCALDE, J., AND P. REVILLA (2004): "Researching with Whom? Stability and Manipulation," *J. Math. Econ.*, 40, 869–887.

BANERJEE, S., H. KONISHI, AND T. SÖNMEZ (2001): "Core in a Simple Coalition Formation Game," *Soc. Choice Welfare*, 18, 135–153.

BARBERÀ, S., AND A. GERBER (2003): "On Coalition Formation: Durable Coalition Structures," *Math. Soc. Sci.*, 45, 185–203.

BARBERÀ, S., H. SONNENSCHEIN, AND L. ZHOU (1991): "Voting by Committees," *Econometrica*, 59, 595–609.

BOGOMOLNAINA, A., AND M. JACKSON (2002): "The Stability of Hedonic Coalition Structures," *Games Econ. Behav.*, 38, 201–230.

CECHLÁROVÁ, C., AND A. ROMERO-MEDINA (2001): "Stability in Coalition Formation Games," *Int. J. Game Theory*, 4, 487–494.

DEMANGE, G. (2004): "On Group Stability in Hierarchies and Networks," *J. Polit. Economy*, 114, 754–777.

——— (2005): "The Strategy Structure of Some Coalition Formation Games," Unpublished Manuscript, EHESS - Paris-Jourdan Sciences Economiques.

DRÈZE, J., AND J. GREENBERG (1980): "Hedonic Coalitions: Optimality and Stability," *Econometrica*, 48, 987–1003.

GIBBARD, A. (1973): "Manipulation of Voting Schemes: A General Result," *Econometrica*, 41, 587–601.

LE BRETON, M., AND A. SEN (1999): "Separable Preferences, Strategy-Proofness and Decomposability," *Econometrica*, 67, 605–628.

LEDYARD, J. (1977): "Incentive Compatible Behavior in Core-Selecting Organizations," *Econometrica*, 45, 1607–1621.

PÁPAI, S. (2004): "Unique Stability in Simple Coalition Formation Games," *Games Econ. Behav.*, 48, 337–354.

SATTERTHWAITE, M. (1975): "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *J. Econ. Theory*, 45, 187–217.

SÖNMEZ, T. (1999): "Strategy-Proofness and Essentially Single-Valued Cores," *Econometrica*, 67, 677–689.

TAKAMIYA, K. (2003): "On Strategy-Proofness and Essentially Single-Valued Cores: A Converse Result," *Soc. Choice Welfare*, 20, 77–83.