Theoretical Economics 4 (2009), 1–15 1555-7561/20090001

Coalition formation under power relations

Michele Piccione
Department of Economics, London School of Economics

Ronny Razin
Department of Economics, London School of Economics

We analyze the structure of a society driven by power relations. Our model has an exogenous power relation over the set of coalitions of agents. Agents determine the social order by forming coalitions. The power relations determine the ranking of agents in society for any social order. We study a cooperative game in partition function form and introduce a solution concept, the stable social order, which exists and includes the core. We investigate a refinement, the strongly stable social order, which incorporates a notion of robustness to variable power relations. We provide a complete characterization of strongly stable social orders.

Keywords. Power, coalition formation, stability.

JEL classification. D0, D7.

1. Introduction

Power relations are a fundamental component of human interaction. In social environments, two types of power shape a significant number of human relations: individual power and group power. Individual power manifests itself in one-to-one relations and generally originates from material or psychological strength. Group power manifests itself in interactions between sets of individuals or in one-to-one interactions between individuals belonging to different sets. The objective of this paper is to study theoretically the joint influence of individual and group power in the determination of social arrangements. Although the term "individual" usually refers to "one person," in this paper "individuals" can be entities such as families, factions, or other groupings, the unity of which is solid and based on exogenous, non-strategic factors such as blood, loyalty, or friendship. Henceforth, such individuals or families are referred to as "agents."

We are interested in analyzing the structure of a society driven by power relations. Our model has the following basic ingredients. Power is represented by an exogenous binary relation over coalitions. Agents determine the social order by forming coalitions.

Michele Piccione: M.Piccione@lse.ac.uk
Ronny Razin: R.Razin@lse.ac.uk

We wish to thank John Moore, Neeve Park, Ariel Rubinstein, and especially Martin Osborne and three referees who provided several valuable and constructive suggestions. We are grateful to William Lauderdale and Iddo Tuvnel for their support.

The power relation and the structure of the social order determine the ranking of agents in society.

Coalitions, in our model, are held together only by strategic considerations. We assume that the objective of each agent is to maximize his/her position in the societal ranking. We study a cooperative game in partition function form and introduce a solution concept, the *stable social order*. We show that for any power relation, the set of stable social orders is not empty and contains the core. We investigate a refinement, the *strongly stable social order*, which requires that a social order be stable for all power relations. We provide a complete characterization (Theorem 1) of strongly stable social orders.

Our framework is too abstract to fit specific historical examples. However, several implications of our results are broadly consistent with stylized historical and political anecdotes. In particular, in a strongly stable social order:

1. Powerful coalitions are large and each coalition is immune from the threat of a unified challenge coming from all less powerful coalitions.

2. Leaders are critical. The elimination of society's most powerful member causes a regime switch: almost all the members of the coalition in power divide into smaller coalitions and significantly drop in status.

As we shall see, the robustness criterion implicit in strongly stable social orders is rather demanding. Hence, we conclude the paper by focusing on social orders that are stable (not necessarily strongly stable) for special power relations.

1.1 *A simple example*

As an illustration of our model and of stable social orders, consider a special case in which the power of individual agents and coalitions is modeled as follows. Each agent i is represented by a number $q(i)$; agent i is more powerful than agent j if and only if $q(i) > q(j)$. When comparing disjoint coalitions of individuals, the power relation is determined additively, i.e., coalition A is more powerful than coalition B whenever $\sum_{i\in A} q(i) > \sum_{i\in B} q(i)$. Suppose that the numbers $q(i)$ are decreasing in i; that is, agent 1 is the most powerful, agent 2 is the second most powerful, and so on. Also suppose that all numbers $q(i)$ are *approximately the same*; that is, a coalition of m agents is more powerful than any coalition with fewer than m agents, and that no two coalitions have the same power.

The agents care only about their social ranking, which is determined by their own individual power and by the power of the coalitions to which they belong.

Suppose, for example, that the set of agents is $I = \{1,2,3,4,5,6,7\}$, and consider the partition (social order)

$$\hat{\Sigma} = \{\{1,3,5,7\},\{2,6\},\{4\}\}.$$

Given the above power relation, the most powerful coalition is $\{1,3,5,7\}$ and the second most powerful coalition is $\{2,6\}$. The social ranking of agents in $\hat{\Sigma}$ is derived as follows. First, agents in a more powerful coalition are ranked higher than agents in a less powerful one. Second, within a coalition, a more powerful agent is ranked higher than a less powerful one. Thus, agent 1 is ranked first as he is in the most powerful coalition and is the most powerful individual in this coalition. Agent 3 is ranked second, agent 2 is ranked fifth and so on.

We analyze the stability of partitions such as $\hat{\Sigma}$. One possible stability notion is the core. We say that a social order is in the *core* if there does not exist a subset of agents who can strictly improve their social rank by forming a new coalition. The above social order is not in the core. If agents 3, 5, and 7 form a new coalition C' dropping agent 1, they strictly improve their social rank in the resulting social order

$$\hat{\Sigma}' = \{\{3,5,7\},\{2,6\},\{1\},\{4\}\}.$$

Piccione and Rubinstein (2004) show, for the above power relation, that the core is empty when $N > 6$. In this paper, we provide a solution concept for which existence is not problematic and that offers interesting insights into coalition formation in the presence of power relations. We follow the traditional route of reducing the set of profitable deviations by allowing counter-deviations. In particular, our *stable social order* incorporates two features:

(i) A recursive definition of "*durable*" deviations and counter-deviations.

(ii) A sequential notion of counter-deviations: members of a deviating coalition do not participate in any immediately subsequent counter-deviation.

The social order $\hat{\Sigma}$ is stable according to our definition. In particular, (all) members of the coalition $C' = \{3,5,7\}$ in $\hat{\Sigma}'$ are made worse off by the "durable" counter-deviation $C'' = \{1,2,4,6\}$. Although the formal definition of durable counter-deviations is recursive, for the moment it is sufficient to note that C'' is durable in that agents 1, 2, 4, and 6 are better off than they are in $\hat{\Sigma}'$ and cannot subsequently be made worse off by any coalition of agents who are not in C''.

As we shall see, the social order $\hat{\Sigma}$ is also *strongly stable*. That is, it is stable for any selection of the numbers $q(i)$ that are decreasing in i; irrespective of the cardinal properties of q, the agents in C' are made worse off by the counter-deviation C'' and the agents in C'' cannot be made worse off.

1.2 *Related literature*

This paper is obviously part of the vast literature on cooperative games, solution concepts, and coalition formation. We refer the reader to Ray (2007) for a detailed and insightful overview. Games in partition function form are studied in Thrall and Lucas (1963), Myerson (1977), and Ray and Vohra (1999). Our solution concept is related to the notion of the "Bargaining Set" of Aumann and Maschler (1964) and, in particular, to a modification due to Dutta et al. (1989); for other notions of stability see Chwe (1994), Ray and Vohra (1997), Greenberg (1990), and Diamantoudi and Xue (2007). Formal models of power relations are analysed in Jordan (2006a,b), Piccione and Rubinstein (2007), and

Acemoglu et al. (2008). Jordan (2006a) considers a model in which power is endogenous and is affected by the wealth that is appropriated from other agents through the exercise of power. Jordan (2006b) incorporates dynamic factors such as histories into the notion of stability, thus introducing a notion of "legitimacy" into the appropriation process. Piccione and Rubinstein (2007) study a model in which the allocation of resources is driven by exogenous power. In this paper, we report a result from Piccione and Rubinstein (2004) that is omitted from Piccione and Rubinstein (2007). Acemoglu et al. (2008) also assume that power is exogenous and study the formation of coalitions under an allocation rule for which the winning coalition takes all.

2. The model

The set of agents is $I = \{1, \dots, N\}$. Although the term "agent" is commonly associated with "one person", in our model an agent can be a clan, a family, or any group of people held together by non-strategic factors. The agents are ordered by an exogenous power relation. We define a "coalitional" power relation over sets of agents as a binary relation P between subsets (coalitions) of agents $A, B \subset I$ such that $A \cap B = \varnothing$. The relation P is asymmetric, acyclic,[1] and such that either $A\,P\,B$ or $B\,P\,A$. The statement $A\,P\,B$ is interpreted as "coalition A is more powerful than coalition B." We assume that $A\,P\,\varnothing$ whenever $A \subset I$ is non-empty. Note that two disjoint coalitions cannot be equally powerful.

Without loss of generality we assume that P agrees with the naming of agents $\{1, \dots, N\}$. That is, $\{1\}\,P\,\{2\}$, $\{2\}\,P\,\{3\}$, $\dots$, $\{N-1\}\,P\,\{N\}$. In what follows, quantifiers such as "for any power relation P" should be interpreted as "for any power relation P for which $\{1\}\,P\,\{2\}$, $\{2\}\,P\,\{3\}$, $\dots$, $\{N-1\}\,P\,\{N\}$." With some abuse of notation we sometimes replace $\{i\}\,P\,\{j\}$ with $i\,P\,j$.

We define a *social order* as a partition of the set of agents. We often denote a social order by Σ and adopt the convention that in the social order $\{C_1, \dots, C_K\}$, $C_i\,P\,C_j$ if and only if $i < j$.

The power relation P is *separable* if, for any subsets of agents A_1, A_2, A_3, and A_4 such that $A_i \cap A_j = \varnothing$ for $i \neq j$, $A_1\,P\,A_3$ and $A_2\,P\,A_4$ implies that

$$(A_1 \cup A_2)\,P\,(A_3 \cup A_4)$$

Consider two coalitions of agents $A, B \subset I$ such that $A \cap B = \varnothing$. Coalition A *dominates* coalition B if there exists a subset $C \subset A$ and a one-to-one mapping $\sigma : B \longrightarrow C$ such that $i\,P\,\sigma^{-1}(i)$ for any $i \in \sigma(B)$. The next lemmas are useful later.

Lemma 1. *Suppose P is separable. If $A\,P\,B$ and $C \subset B$, then $A\,P\,C$.*

Proof. Suppose not. A contradiction is obtained by defining $A_1 = C$, $A_2 = B \backslash C$, $A_3 = A$, and $A_4 = \varnothing$. □

[1] The relation P is acyclic if, given any collection Θ of subsets of agents, there exists $A \in \Theta$ such that $B\,P\,A$ for no $B \in \Theta$.

LEMMA 2. *Suppose P is separable. Then $A\,P\,B$ whenever A dominates B.*

This result follows from a simple application of separability.

The power relation P is *monotonic* if for any two subsets of agents A_1 and A_2 such that $A_1 \cap A_2 = \varnothing$, $A_1\,P\,A_2$ implies that

$$((A_1 \cup \{i\}) \setminus \{j\})\,P\,((A_2 \cup \{j\}) \setminus \{i\})$$

whenever $i \in A_2$ and $i\,P\,j$.

LEMMA 3. *If P is separable, then P is monotonic.*

PROOF. Consider two subsets of agents A_1 and A_2 such that $A_1 \cap A_2 = \varnothing$ and $A_1\,P\,A_2$. Take $i \in A_2$ and any j such that $i\,P\,j$. First suppose that $(A_2 \setminus \{i\})\,P\,(A_1 \setminus \{j\})$. Then, by separability,

$$((A_2 \setminus \{i\}) \cup \{i\})\,P\,((A_1 \setminus \{j\}) \cup \{j\}).$$

Since $A_1 \subset (A_1 \setminus \{j\}) \cup \{j\}$, a contradiction is obtained by Lemma 1. Hence $(A_1 \setminus \{j\})\,P\,(A_2 \setminus \{i\})$. The claim follows by separability. □

In the remainder of the paper, we assume that power relations are separable. Finally, we define the social ranking that is induced by a social order. Given a power relation P and a social order $\Sigma = \{C_1, \dots, C_K\}$, let $V_i^P(\Sigma)$ denote agent i's position in the ranking induced by Σ. That is, $V_i^P(\Sigma) = 1$ indicates that agent i is ranked the highest, $V_i^P(\Sigma) = 5$ indicates that agent i is ranked fifth, and so on. Formally, $V_i^P(\Sigma)$ assigns to each agent i an integer in $\{1, 2, \dots, N\}$ and satisfies

$$\begin{aligned} V_i^P(\Sigma) < V_j^P(\Sigma) \text{ if and only if} \\ &\text{either } i, j \in C_k \text{ for some } k \text{ and } i\,P\,j \\ &\text{or } i \in C_k,\ j \in C_{k'}, \text{ and } C_k\,P\,C_{k'}. \end{aligned} \tag{*}$$

We say that agent i is ranked "higher" in Σ than in Σ' whenever $V_i^P(\Sigma) < V_i^P(\Sigma')$. We assume that each agent's preferences over social orders are determined by the induced social rankings. In particular, each agent strictly prefers to be ranked higher in the social ranking to being ranked lower. That is, each agent i strictly prefers the social order Σ to the social order Σ' if and only if $V_i^P(\Sigma) < V_i^P(\Sigma')$. We also say, given $\Sigma = \{C_1, \dots, C_K\}$, that C_k is ranked k^{th} and that C_k is ranked higher than $C_{k'}$ whenever $k < k'$ (recall that $C_k\,P\,C_{k'}$ by convention).

3. STABILITY

We introduce a cooperative solution concept for social orders that we call *stable social order*. For any subset C of agents who deviate from a social order $\Sigma = \{C_1, \dots, C_K\}$, with some abuse of the conventional notation let $\Sigma \top C$ be the partition $\{C_1 \setminus C, C_2 \setminus C, \dots, C_K \setminus C, C\}$. We say that a deviation by C from Σ is *profitable* if $V_i^P(\Sigma \top C) < V_i^P(\Sigma)$ for any $i \in C$.

Our stability notion is based on the durability of deviations by coalitions of agents. Two criteria need to be satisfied by a durable deviation. First, all members in the deviating coalition are better off. Second, there does not exist a durable counter-deviation that makes some member in the original deviating coalition worse off than in the social order prior to the deviation. It should be noted that members of the deviating coalition are excluded from the counter-deviating coalition.

We now define durable deviations. Let Ξ be the set of social orders and $\mathscr{I}$ be the set of all possible subsets of I. Define the correspondence $\mathscr{S}^P : \Xi \Longrightarrow \mathscr{I}$ such that $C \in \mathscr{S}^P(\Sigma)$ if and only if

(a) C is a profitable deviation from Σ

(b) there does not exist $C' \in \mathscr{S}^P(\Sigma \top C)$ such that

  (i) $C \cap C' = \varnothing$

  (ii) $V_i^P((\Sigma \top C) \top C')) > V_i^P(\Sigma)$ for some $i \in C$.

A deviation C from a social order Σ is *durable* if $C \in \mathscr{S}^P(\Sigma)$. The following proposition shows that the mapping $\mathscr{S}^P : \Xi \Longrightarrow \mathscr{I}$ exists and is unique, notwithstanding its self-referential nature.

Proposition 1. *There exists a unique correspondence $\mathscr{S}^P : \Xi \Longrightarrow \mathscr{I}$ that satisfies (a) and (b).*

Proof. Given a social order Σ and a coalition C with a profitable deviation from Σ, let

$$\tilde{\mathscr{S}}(\Sigma, C) = \{C' \subset I : \begin{array}{l} \text{(i) } C \cap C' = \varnothing \\ \text{(ii) } C' \text{ is a profitable deviation from } \Sigma \top C. \\ \text{(iii) } V_i^P((\Sigma \top C) \top C')) > V_i^P(\Sigma) \text{ for some } i \in C\}. \end{array}$$

Consider all finite sequences $\{B^t\}_{t=0}^{\tau}$ of subsets of agents such that

- $B^0 = C$
- $\Sigma^0 = \Sigma$, $\Sigma^{t+1} = \Sigma^t \top B^t$
- $B^t \in \tilde{\mathscr{S}}(\Sigma^{t-1}, B^{t-1})$, $B^t \cap B^{t-1} = \varnothing$ for $t > 0$
- either $\tilde{\mathscr{S}}(\Sigma^{t-1}, B^{t-1}) \neq \varnothing$ for any $t \leq \tau$ and $\tilde{\mathscr{S}}(\Sigma^{\tau}, B^{\tau}) = \varnothing$, or $\tilde{\mathscr{S}}(\Sigma^{t-1}, B^{t-1}) \neq \varnothing$ for any t and $\tau = \infty$.

Note that, by (ii) and (iii) in the definition of $\tilde{\mathscr{S}}$, each member of B^t is better off in $\Sigma^{\tau+1}$ than in Σ^{τ} and that at least one member of B^t is worse off in $\Sigma^{\tau+2}$ than in Σ^{τ}. Hence, $B^t \; P \; B^{t-1}$ for every $t > 0$. Thus, by acyclicity, there exists a finite bound for τ that is common to all sequences $\{B^t\}_{t=0}^{\tau}$. Since $\tilde{\mathscr{S}}(\Sigma^{\tau}, B^{\tau}) = \varnothing$, if $B \in \mathscr{S}^P(\Sigma^{\tau} \top B^{\tau})$ then $B \cap B^{\tau} \neq \varnothing$. Hence, $B^{\tau} \in \mathscr{S}^P(\Sigma^{\tau})$ and $B^{\tau-1} \notin \mathscr{S}^P(\Sigma^{\tau-1})$. Consider now a directed graph in which each B^t is a node and a directed edge links B^t to B^{t+1} if and only if B^t

immediately precedes B^{t+1} in the same sequence. If none of the immediate successors of B^t is in $\mathscr{S}^P(\Sigma^t \curlywedge B^t)$, then $B^t \in \mathscr{S}^P(\Sigma^t)$. If at least one immediate successor of B^t is in $\mathscr{S}^P(\Sigma^t \curlywedge B^t)$, then $B^t \notin \mathscr{S}^P(\Sigma^t)$. Proceeding by backward induction in this fashion, we determine uniquely whether $C \in \mathscr{S}^P(\Sigma)$. $\square$

The example in Section 1.1 clarifies the intuition behind this result. The coalition C'' is a durable deviation from $\hat{\Sigma}'$ since the agents in C'' cannot be made worse off by any coalition of agents who are not in C''. Working backwards, one deduces that the coalition C' is not a durable deviation from $\hat{\Sigma}$.

We are now ready to define the stability of a social order.

Definition 1. A social order $\Sigma = \{C_1, \dots, C_K\}$ is *stable* for a power relation P if $\mathscr{S}^P(\Sigma) = \varnothing$.

One can refine the stability notion by requiring that social orders are stable for any power relation.

Definition 2. A social order $\Sigma = \{C_1, \dots, C_K\}$ is *strongly stable* if $\mathscr{S}^P(\Sigma) = \varnothing$ for any power relation P.

The requirements for a strongly stable social order are severe but can be partially justified on robustness grounds. It is natural to think of the power of coalitions as more variable and harder to assess than the power of individuals.[2] The aggregate strength of a group can depend on characteristics of social interaction that are unobservable and difficult to evaluate.

4. The main result

In this section, we introduce and prove our main result. First we introduce some special social orders.

The social order Σ^* is constructed according to a simple procedure. First, select the odd-indexed agents to form the strongest coalition. Re-index the remaining agents so that the most powerful agent is indexed as agent $1'$, the second most powerful is indexed as agent $2'$ and so on. Select the odd indexed agents from this set to form the second strongest coalition. Repeat this procedure until no agents are left. For example, when $N = 8$, $\Sigma^* = \{\{1,3,5,7\},\{2,6\},\{4\},\{8\}\}$.

Formally, given a set of numbers $Q = \{a, b, c, d, \dots\}$ and a number δ, let δQ denote the set $\{\delta a, \delta b, \delta c, \delta d, \dots\}$. Let $\mathbb{O}_+$ be the set of the positive odd integers. Define the social order Σ^* as the social order $\{C_1^*, \dots, C_K^*\}$ such that $C_k^* = I \cap 2^{k-1}\mathbb{O}_+$, where K is the largest k for which $C_k^* = I \cap 2^{k-1}\mathbb{O}_+$ is non-empty.

Consider the class $\mathscr{F}$ of social orders derived by modifying Σ^* recursively in the following fashion. A social order $\Sigma = \{C_1, \dots, C_K\}$ is in $\mathscr{F}$ if and only if

1. $C_1 = C_1^*$ or $C_1 = C_1^* \cup \{N\}$
2. for $k \geq 2$, $C_k = \{C_k^* \setminus \cup_{j=1}^{k-1} C_j\}$ or $C_k = \{C_k^* \setminus \cup_{j=1}^{k-1} C_j\} \cup \{\max(I \setminus \cup_{j=1}^{k-1} C_j)\}$.

[2]Recall that $1\,P\,2\,P \cdots P\,N$ for any P in the definition of strongly stable social orders.

For $N = 8$, the social order $\{\{1,3,5,7,8\},\{2,6\},\{4\}\}$ is in $\mathscr{F}$ as it is obtained by adding agent 8 to C_1^*. It is worth noting two features common to the social orders in $\mathscr{F}$. First, the coalitions are highly differentiated in that agents who are contiguous in power generally belong to different coalitions: with the possible exception of the two least powerful agents, for any two agents x and y in any coalition C for whom $x\,P\,y$, there exists an agent z not in C such that $x\,P\,z$ and $z\,P\,y$. Second, any coalition C dominates the union of all coalitions that are less powerful than C.

Theorem 1. *A social order Σ is a strongly stable social order if and only if $\Sigma \in \mathscr{F}$.*

The proof of this result is constructive. We first prove that Σ^* is stable for any P (the proof that any $\Sigma \in \mathscr{F}$ is stable for any P is analogous and thus omitted). We then prove that any strongly stable social order must be in $\mathscr{F}$.

Proof of Theorem 1. We first establish some preliminary results.

Claim 1. *For any j, $C_j^*\,P\,(\cup_{i=j+1}^{K} C_i^*)$.*

This result is proved by construction.

Claim 2. *Fix some partition Σ such that $C_1^* \in \Sigma$ and $2 \in C_2$. For any C such that $C_1^* \cap C \neq \varnothing$ and $V_i^P(\Sigma \top C) < V_i^P(\Sigma)$ for each $i \in C$, there exists C' with $C' \cap C = \varnothing$ such that*

(i) C' is a profitable deviation from $\Sigma^* \top C$

(ii) $V_i^P((\Sigma^* \top C) \top C')) > V_i^P(\Sigma^*)$ for some $i \in C$

(iii) $C'\,P\,((I \setminus C) \setminus C')$.

Proof. Denote $C_1^* = \{y_1,\ldots,y_L\}$ and $C = \{x_1,\ldots,x_M\}$. Construct $C' = \{z_1,\ldots,z_L\}$ by first letting $z_1 = y_1 = 1$. Suppose we have defined z_i for all $i \leqslant j$ for some $j \geqslant 1$. Define z_{j+1} as the smallest i such that (i) $i \notin C$, (ii) $i \neq z_1,\ldots,z_j$, (iii) $V_i^P(\Sigma \top C) > j+1$, and (iv) $i \leq y_{j+1}$. We now show that this algorithm is well defined. We consider several cases.

Case 1: In $\Sigma \top C$, C is ranked first and $C_1^* \setminus C$ is ranked second.

First note that either $z_2 = 2$, or $2 \in C$ and therefore $3 \notin C$ implying $z_2 = 3$. Therefore, $z_2 \leqslant y_2$. Now consider z_j, $j > 2$, given that $z_1,\ldots,z_{j-1}$ have been selected using the algorithm. Let G_j be the set of agents smaller than or equal to $y_j = 2j-1$. By hypothesis, $j-1$ agents in G_j have already been allocated to C'. We now show that the set $H_j = \{r \in G_j \setminus \{z_1,\ldots,z_{j-1}\} : r \notin C \text{ and } V_r^P(\Sigma \top C) > j\}$ is not empty. Since $\#G_j \setminus \{z_1,\ldots,z_{j-1}\} = j$, it is impossible that all agents in $G_j \setminus \{z_1,\ldots,z_{j-1}\}$ are in C. If so, agent $2j-1$ would also be in C but ranked at or lower than the j^{th} position in $\Sigma \top C$, contradicting the definition of C. Therefore, there must exist at least one agent $i \in G_j \setminus \{z_1,\ldots,z_{j-1}\}$ such that $i \notin C$. Consider then the agent r^* in $G_j \setminus \{z_1,\ldots,z_{j-1}\}$ who is ranked lowest in $\Sigma \top C$. Agent r^* is not in C as otherwise $G_j \setminus \{z_1,\ldots,z_{j-1}\} \subset C$. Since $C_1^* \setminus C$ is ranked second, agent r^* must be ranked lower than agent 1 in $\Sigma \top C$. Since agent 1 is not in $G_j \setminus \{z_1,\ldots,z_{j-1}\}$ and $\#G_j \setminus \{z_1,\ldots,z_{j-1}\} = j$, $V_{r^*}^P(\Sigma \top C) > j$. Hence, H_j is not empty. Define $z_j = \min H_j$.

Case 2: In $\Sigma \top C$, C is ranked first and $C_1^* \setminus C$ is ranked lower than second.

Again, either $z_2 = 2$, or $2 \in C$ and therefore $3 \notin C$ implying $z_2 = 3$. Consider z_j, $j > 2$. As in Case 1, $\#G_j \setminus \{z_1, \dots, z_{j-1}\} = j$. Denote the agents in $G_j \setminus \{z_1, \dots, z_{j-1}\}$ by $\gamma_1 < \cdots < \gamma_j$. We need to show that in $\Sigma \top C$ one γ_k is ranked strictly lower than the j^{th} position. If agent $2j-1$ is not in C, then the claim is obvious as $C_1^* \setminus C$ is ranked lower than second. Hence, we can suppose that agent $2j-1$ is in C. Note that it is impossible that all agents $\gamma_1, \dots, \gamma_j$ are in C. If so, agent $\gamma_j = 2j-1$ is ranked no higher than j^{th} in C, contradicting the definition of C. Also, if agent $2j-1$ is not the lowest ranked agent in C, agents not in $\gamma_1, \dots, \gamma_j$ are also in C. Since not all agents $\gamma_1, \dots, \gamma_j$ are in C, one γ_k must be ranked strictly lower than the j^{th} position. We can then suppose that $2j-1$ is the lowest ranked agent in C.

If C does not contain an even agent, then agent 2 is the highest ranked agent in the second ranked coalition in $\Sigma \top C$, which we denote by $C_2^{\Sigma \top C}$. In this case, $z_2 = 2$ and agent 2 is not in $\gamma_1, \dots, \gamma_j$. Since some agents in $\gamma_1, \dots, \gamma_j$ are not in C and agent 2 is the highest ranked agent in $C_2^{\Sigma \top C}$, one γ_k must be ranked strictly lower than the j^{th} position. Hence, if C does not contain an even agent, the claim is proved. Suppose then that C does contain an even agent. Note further that for this even agent γ, $\gamma < 2j-1$, as otherwise agent $2j-1$ is not the lowest ranked in C.

To summarize, in order to conclude the proof of Case 2, we suppose that

(i) not all $\gamma_1, \dots, \gamma_j$ are in C

(ii) $2j-1$ is the lowest ranked agent in C

(iii) C contains an even agent.

Also, if $C_2^{\Sigma \top C}$ contains an agent z_k, $k < j$, (i) implies that at least one γ_l must be ranked strictly lower than the j^{th} position in $\Sigma \top C$. If not, agent z_k is ranked strictly lower than the j^{th} position in $\Sigma \top C$, the agent γ_l who is ranked j^{th} in $\Sigma \top C$ is in $C_2^{\Sigma \top C}$, and $\gamma_l < z_k$. As $j > k$, the algorithm should not have selected z_k. Hence, we also suppose that

(iv) $C_2^{\Sigma \top C}$ does not contain any agents in $z_1, \dots, z_{j-1}$.

Now let Θ be the set $C_1^* \cap C$. Given any $(2k-1) \in C$, $k < j$, if q odd agents smaller than or equal to $(2k-1)$ are in C, at least q even agents who are smaller than $(2k-1)$ are in $z_1, \dots, z_{j-1}$. By (ii), at least $\#\Theta - 1$ even agents must be in $z_1, \dots, z_{j-1}$. By (iii), call Θ' the set composed of these even agents and one even agent i' from C. We can now construct a one-to-one mapping $g : \Theta \to \Theta'$ such that $g(z) < z$. First, let $g(2j-1) = i' < 2j-1$ by (ii) and (iii). Let $\Theta = \{\theta_1, \dots, \theta_m\}$, $\theta_i < \theta_{i+1}$. It is easy to see that there must be an even agent $i \in \Theta' \setminus \{i'\}$ such that $i < \theta_1$. Define $g(\theta_1)$ as the lowest such number. Suppose that for $k-1$ agents θ_i, $i < k$, $k > 1$, we have constructed $g(z)$. Since k odd agents smaller than or equal to θ_k are in C, there must exist k even agents in $z_1, \dots, z_k$ who are smaller than θ_k. Hence, there must be an even agent $\theta'_k > g(\theta_{k-1})$ with $\theta'_k \in \Theta' \setminus \{i'\}$ and $\theta'_k < \theta_k$. Let $g(\theta_k)$ be the highest ranked such agent.

Hence, Θ' dominates Θ. By (iv), $\Theta' \cap C_2^{\Sigma \intercal C} = \varnothing$. Since $C_1^* \, P \, (N \setminus C_1^*)$, then $(C_1^* \setminus \Theta) \, P \, ((N \setminus C_1^*) \setminus \Theta')$ by separability. Since $C_2^{\Sigma \intercal C}$ is contained in $(N \setminus C_1^*) \setminus \Theta'$, Lemma 1 implies $(C_1^* \setminus \Theta) \, P \, C_2^{\Sigma \intercal C}$, a contradiction.

Case 3: C is not ranked first in $\Sigma \intercal C$.

Let $C_1^{\Sigma \intercal C}$ be the highest ranked coalition in $\Sigma \intercal C$. Since $C_1^* \setminus C$ is ranked lower than $C_1^{\Sigma \intercal C}$, $\# C_1^{\Sigma \intercal C} \geqslant 2$. Hence, agent 3 cannot be in C. Since agent 3 is in $C_1^* \setminus C$, $\# C_1^{\Sigma \intercal C} \geqslant 3$. Continuing in this fashion, we establish that $C \cap C_1^* = \varnothing$, a contradiction.

By construction, C' is a profitable deviation from $\Sigma^* \intercal C$. To see that $V_i^P((\Sigma^* \intercal C) \intercal C') > V_i^P(\Sigma^*)$ for some $i \in C$, take any $\hat{x} \in C \cap C_1^*$. Agent $\hat{x}$'s position is weakly higher than the L^{th} position in Σ^* and strictly lower than the L^{th} position in $(\Sigma^* \intercal C) \intercal C'$.

Finally, we show that $C' \, P \, ((I \setminus C) \setminus C')$. Indeed, our construction ensures that $C' \, P \, (I \setminus C')$. By Claim 1, $C_1^* \, P \, (I \setminus C_1^*)$. Since C' is derived from C_1^* by exchanging less powerful agents in C_1^* for more powerful agents in $I \setminus C_1^*$, monotonicity implies that $C' \, P \, (I \setminus C')$. $\lhd$

To prove stability, fix some partition Σ such that $C_1^*, C_2^* \in \Sigma$ and $\min C_3^* \in C_3$. For any C such that $C_1^* \cap C = \varnothing$, $C_2^* \cap C \neq \varnothing$, and $V_i^P(\Sigma \intercal C) < V_i^P(\Sigma)$ for each $i \in C$, construct a counter-deviation C'' that is constructed analogously to C' in Claim 2 (ignoring the agents in C_1^*). Namely, denoting $C_2^* = \{y_1, \ldots, y_{L'}\}$ and $C = (x_1, \ldots, x_{M'})$, construct $C'' = \{z_1, \ldots, z_{L'}\}$ by first letting $z_1 = y_1 = 2$. Having defined z_i for all $i \leqslant j$ for some $j \geqslant 1$, define z_{j+1} as the smallest i such that (i) $i \notin C$, (ii) $i \neq z_1, \ldots, z_j$, (iii) $V_i^P(\Sigma \intercal C) > j + 1 + \# C_1^*$, and (iv) $i \leq y_{j+1}$. The deviation C'' is durable; any counter-deviation to C'' in $(\Sigma \intercal C) \intercal C''$ cannot be durable by Claim 2.

The completion of the proof of stability is obtained by an inductive repetition of these arguments.

Finally, we need to show that if Σ is strongly stable then $\Sigma \in \mathscr{F}$. First, we show that any ranking of agents induced by a strongly stable social order must rank the agents in C_1^* as in Σ^*.

Agent 1 needs to be ranked first: consider a power relation such that $\{1\} \, P \, (I \setminus \{1\})$. To see that agent 3 must be at least second, consider P such that $\{2,3\} \, P \, (I \setminus \{2,3\})$. Since agent 1 must be first, if agent 3 is not second, he can deviate forming a coalition with agent 2. Now consider agent 5 and assume he is ranked lower than the third position. Choose P such that $\{2,4,5\} P (I \setminus \{2,4,5\})$. Since agents 1 and 3 are first and second, agents 2 and 4 can form a coalition with agent 5 and improve their rank. Suppose we have shown that all agents $2i-1 \in C_1^*$ are in the i^{th} position. Consider the agent $2i+1 \in C_1^*$ and suppose he is below the $(i+1)^{\text{th}}$ position. Choose P such that $\{2,4,6,\ldots,2i,2i+1\} \, P \, (I \setminus \{2,4,6,\ldots,2i,2i+1\})$. Since $\{1,3,5,\ldots,2i-1\}$ are ranked in the $1,\ldots,i^{\text{th}}$ positions, agents $\{2,4,6,\ldots,2i\}$ can form a coalition with $2i+1$ and improve their rank.

Now it is easy to verify that no agent in C_1^* can belong to a coalition that contains agents who are not in C_1^* with the exception of agent N. To show that all agents in C_1^* must belong to the same coalition, choose P such that $\{2\}\,P\,(I\setminus\{1,2\})$.

To characterise C_2, repeat the arguments above for social orders where the set of agents is $I\setminus C_1$. To ensure that deviations analogous to the ones in the previous paragraphs are durable when the set of agents is I, it is sufficient to consider P's such that $\{1\}\,P\,(I\setminus\{1\})$, as no agent in C_1 would then join a counter-deviation. Repeating these arguments for all C_i's concludes the proof. □

5. Remarks on the stability concept

5.1 *Axioms for strong stability*

The proof of Theorem 1 is quite complex. To gain some insights into the main arguments, we provide a simple and intuitive axiomatic characterization of the social orders in $\mathscr{F}$ that underscores their stability properties.

(K1) A social order $\Sigma=\{C_1,\ldots,C_K\}$ is such that C_i dominates $\cup_{j=i+1}^{K}C_j$ for any $i=1,\ldots,K$.

(K2) A social order $\Sigma=\{C_1,\ldots,C_K\}$ is such that for all $i=1,\ldots,K$, the rank of any agent $j\in C_i$ within C_i is at least as high as his rank within the set $(\cup_{l=i+1}^{K}C_l)\cup\{j\}$.

Axiom (K1) is a criterion for external stability: all coalitions are immune from the threat of a unified challenge coming from all weaker coalitions. Axiom (K2) is a criterion for internal stability in that agents in a coalition never wish to join a united challenge by all weaker coalitions.

Proposition 2. *Σ satisfies (K1) and (K2) if and only if $\Sigma\in\mathscr{F}$.*

Proof. First note that any $\Sigma\in\mathscr{F}$ satisfies (K1) and (K2). Now consider a social order $\Sigma=\{C_1,\ldots,C_K\}$ that satisfies (K1) and (K2). By (K1), agent 1 is in C_1. By (K2), agent 2 cannot be in C_1. By (K1) again, agent 3 is in C_1. By (K2) again, agent 4 cannot be in C_1. Repeating these arguments up to $\max C_1$ implies that $C_1\supset C_1^*$. Recall that $\#C_1^*\geq\frac{1}{2}N$. If $C_1\setminus C_1^*\neq\varnothing$, (K2) implies that any agent $j\in C_1\setminus C_1^*$ must be ranked worse than all agents in $\cup_{l=i+1}^{K}C_l$ and $\#C_1\setminus C_1^*\leq 1$. The same arguments for the other coalitions establish that $\Sigma\in\mathscr{F}$. □

Note that strongly stable social orders depend critically on coalition leaders. In particular, eliminating agent 1 from the set of agents causes a major upset in the social structure. The new most powerful coalition is composed of agents who were not in C_1^*, while those agents who were in C_1^* are now divided into smaller and less powerful coalitions. In contrast, eliminating the lowest individual in society does not affect the social order except for the absence of that agent.

5.2 *The core*

In the previous section, we show that the set of (strongly) stable social orders is not empty. However, a standard solution concept such as the *core* can be empty in our framework. A social order $\Sigma = \{C_1, \dots, C_K\}$ is in the *core* if no coalition of agents is a profitable deviation.

We say that the relation P is *homogeneous* if a coalition of m agents is more powerful than any coalition of strictly fewer than m agents. The next proposition is from Piccione and Rubinstein (2004).

Proposition 3. *If P is homogeneous, the core is empty when $N \geq 7$.*

Proof. We first establish the following claims.

Claim 1. *The least powerful coalition C_K in Σ has either 1 or 2 agents.*

Proof. If not, all agents in C_K except for the most powerful can form a coalition that strictly improves their ranking. ◁

Claim 2. $\#C_{j+1} \leq \#C_j \leq \#C_{j+1} + 1$ *for* $j = 0, \dots, K-1$.

Proof. The left-hand side follows by definition and by homogeneity. For the right-hand side, if $\#C_j > \#C_{j+1} + 1$, all agents in C_j except for the most powerful can form a coalition that strictly improves their ranking. ◁

Claim 3. $\#C_2 \geq 2$ *and* $\#C_1 \geq 3$.

Proof. If $\#C_2 = 1$ then by Claims 1 and 2, $K > 5$ and $\#C_j = 1$ for $j = 2, 3, 4, \dots, K$. Thus merging C_K and C_{K-1} improves the ranking of all the members of the new coalition. Hence, $\#C_2 \geq 2$. If $\#C_1 = 2$, then merging C_2 with one element of C_K improves the ranking of all the members of the new coalition. ◁

Since $N \geq 7$, there are at least two agents who do not belong to either C_1 or C_2. If two such agents form a coalition with the agents in C_2, the ranking of each member of the new coalition improves. □

Note that, for $N = 6$, the social order $\{\{1,5,6\},\{3,4\},\{2\}\}$ is in the core when P is homogeneous and $\{1,5,6\}\, P \,\{2,3,4\}$.

5.3 *Existence and farsightedness*

Our solution concept is rooted in the theory of cooperative games. Stable social orders are defined as collections of coalitions that agents do not find in their interest or are unable to destabilize by forming new coalitions. Specifically, agents consider forming a new coalition if (i) they are all better off and (ii) they do not expect the formation of a retaliatory coalition that *excludes* them and makes some of them worse off.

Existence is one of the main advantages of our solution concept, as is highlighted by the emptiness of the core. The assumption that the agents in a deviating coalition do not participate in any immediately subsequent counter-deviation is especially helpful to this end.[3] Moreover, we find this restriction natural if counter-deviations are retaliatory.

A standard objection to this approach is that agents do not consistently foresee the final consequences of their deviations. According to this criticism, the only consideration that should guide the decision by a set of agents forming a new coalition is their equilibrium expectation of the stable social order that ensues. This objection ultimately applies to all cooperative approaches to coalition formation known to us, and a satisfactory resolution is well beyond the scope of this paper. We wish to point out, however, that one need not treat coalition formation as effortless in a cooperative approach. As forming coalitions requires a high degree of coordination and common intent, one can argue that some coalitions are formed more easily than others. It is implicit in our approach that conditions (i) and (ii) above facilitate the coordination of agents forming a new coalition whereas the absence of either condition makes it demanding. Indeed, (i) and (ii) are very natural considerations. Stable social order can be interpreted as "robust" structures from which such considerations are absent and in which the formation of new, destabilizing coalitions is hindered.

6. Special power relations

We are unable to provide a complete characterization of stable social orders under an arbitrary power relation. In this section, we explore social orders that are stable for particular power relations.

6.1 *Congruence*

Generally, stable social orders induce a ranking of agents that differs from the ranking of agents under P. The next proposition shows that a ranking of agents that agrees with P can be induced by a stable social order if and only if agent 1 is more powerful than the coalition of all remaining agents.

Proposition 4. *There exists a stable social order Σ such that $V_i^P(\Sigma) = i$ for $i = 1, 2 \ldots, N$ if and only if $\{1\}\, P\, \{2, 3, \ldots, N\}$.*

Proof. If $\{1\}\, P\, \{2, 3, \ldots, N\}$, consider the social order with only one coalition. For any deviating coalition C, let j be the agent ranked highest in C. It is easy to verify that j cannot be ranked higher than the j^{th} position. Consider a social order Σ such that $V_i^P(\Sigma) = i$, $i = 1, 2 \ldots, N$, and assume that $\{2, 3, \ldots, N\}\, P\, \{1\}$. Obviously $\{2, 3, \ldots, N\}$ is a durable deviation and therefore Σ is not stable. □

6.2 *Homogeneous power*

Suppose that the power of agents is approximately the same. The following result shows that, in a stable social order, the most powerful coalition must exclude some of the most powerful agents.

[3] When deviations are non-nested, existence is problematic in equilibrium concepts that involve bootstrapping (see Ray 2007, p. 240).

Proposition 5. *Consider a stable social order Σ. If P is homogeneous, it is impossible that* $\{1,2,3\} \subset C_1$.

Proof. Consider first the case of N even and suppose that there exists a stable Σ such that $\{1,2,3\} \subset C_1$. To obtain a contradiction, take a coalition with agents 2, 3, and all agents ranked strictly lower than $\frac{1}{2}N+1$ in Σ. This deviation is durable. Now consider the case of N odd and suppose that there exist a stable Σ such that $\{1,2,3\} \subset C_1$. Take a coalition with agent 2 and all agents ranked strictly lower than $\frac{1}{2}(N+1)$ in Σ. This deviation is again durable. □

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2008), "Coalition formation in non-democracies." *Review of Economic Studies*, 75, 987–1009. [4]

Aumann, Robert J. and Michael Maschler (1964), "The bargaining set for cooperative games." In *Advances in Games Theory* (Melvin Dresher, Lloyd S. Shapley, and Alan W. Tucker, eds.), 443–476, Princeton University Press, Princeton. [3]

Chwe, Michael S. (1994), "Farsighted coalitional stability." *Journal of Economic Theory*, 63, 299–325. [3]

Diamantoudi, Effrosyni and Licun Xue (2007), "Coalitions, agreements and efficiency." *Journal of Economic Theory*, 136, 105–125. [3]

Dutta, Bhaskar, Debraj Ray, Kunal Sengupta, and Rajiv Vohra (1989), "A consistent bargaining set." *Journal of Economic Theory*, 49, 93–112. [3]

Greenberg, Joseph (1990), *The Theory of Social Situations*. Cambridge University Press, Cambridge. [3]

Jordan, James S. (2006a), "Pillage and property." *Journal of Economic Theory*, 131, 26–44. [3, 4]

Jordan, James S. (2006b), "Power and legitimacy in pillage games." Unpublished paper, Department of Economics, Pennsylvania State University. [3, 4]

Myerson, Roger B. (1977), "Values of games in partition function form." *International Journal of Game Theory*, 6, 23–31. [3]

Piccione, Michele and Ariel Rubinstein (2004), "Equilibrium in the jungle." Unpublished paper. [3, 4, 12]

Piccione, Michele and Ariel Rubinstein (2007), "Equilibrium in the jungle." *Economic Journal*, 117, 883–896. [3, 4]

Ray, Debraj (2007), *A Game-Theoretic Perspective on Coalition Formation*. Oxford University Press, Oxford. [3, 13]

Ray, Debraj and Rajiv Vohra (1997), "Equilibrium binding agreements." *Journal of Economic Theory*, 73, 30–78. [3]

Ray, Debraj and Rajiv Vohra (1999), "A theory of endogenous coalition structures." *Games and Economic Behavior*, 26, 286–336. [3]

Thrall, Robert M. and William F. Lucas (1963), "n-person games in partition function form." *Naval Research Logistics Quarterly*, 10, 281–298. [3]

Submitted 2008-4-1. Final version accepted 2008-10-17. Available online 2008-10-17.