Coalition Formation in Nondemocracies[*]

Daron Acemoglu
MIT

Georgy Egorov
Harvard

Konstantin Sonin
New Economic School

December 2007

Abstract

We study the formation of a ruling coalition in nondemocratic societies where institutions do not enable political commitments. Each individual is endowed with a level of political power. The ruling coalition consists of a subset of the individuals in the society and decides the distribution of resources. A ruling coalition needs to contain enough powerful members to *win against* any alternative coalition that may challenge it and it needs to be *self-enforcing*, in the sense that none of its subcoalitions should be able to secede and become the new ruling coalition. We present both an axiomatic approach that captures these notions and determines a (generically) unique ruling coalition and the analysis of a dynamic game of coalition formation that encompasses these ideas. We establish that the subgame perfect equilibria of the coalition formation game coincide with the set of ruling coalitions resulting from the axiomatic approach. A key insight of our analysis is that a coalition is made self-enforcing by the failure of its winning subcoalitions to be self-enforcing. This is most simply illustrated by the following example: with "majority rule," two-person coalitions are generically not self-enforcing and consequently, three-person coalitions are self-enforcing (unless one player is disproportionately powerful). We also characterize the structure of ruling coalitions. For example, we determine the conditions under which ruling coalitions are robust to small changes in the distribution of power and when they are fragile. We also show that when the distribution of power across individuals is relatively equal and there is majoritarian voting, only certain sizes of coalitions (e.g., with "majority rule," coalitions of size 3, 7, 15, 31, etc.) can be the ruling coalition.

Keywords: coalition formation, political economy, self-enforcing coalitions, stability.
JEL Classification: D71, D74, C71.

1 Introduction

We study the formation of a *ruling coalition* in a nondemocratic ("weakly institutionalized") environment. A ruling coalition must be powerful enough to impose its wishes on the rest of the society. A key ingredient of our analysis is that because of the absence of strong, well-functioning institutions, binding agreements are not possible.[1] This has two important implications: first, members of the ruling coalition cannot make binding offers on how resources will be distributed; second, and more importantly, members of a candidate ruling coalition cannot commit to not eliminating (sidelining) fellow members in the future. Consequently, there is always the danger that, once a particular coalition has formed and has centralized power in its hands, a subcoalition will try to remove some of the original members of the coalition in order to increase the share of resources allocated to itself. Ruling coalitions must therefore not only be powerful enough to be able to impose their wishes on the rest of the society, but also *self-enforcing* so that none of their subcoalitions are powerful enough and wish to split from or eliminate the rest of this coalition. These considerations imply that the nature of ruling coalitions is determined by a tradeoff between "power" and "self-enforcement".

More formally, we consider a society consisting of an arbitrary number of individuals with different amount of political or military powers ("guns"). Any subset of these individuals can form a coalition and the power of the coalition is equal to the sum of the powers of its members. We formalize the interplay between power and self-enforcement as follows: a coalition with sufficient power is *winning against* the rest of the society and can centralize decision-making powers in its own hands (for example, eliminating the rest of the society from the decision-making process). How powerful a coalition needs to be in order to be winning is determined by a parameter α. When $\alpha = 1/2$, this coalition simply needs to be more powerful than the rest of the society, so this case can be thought of as "majority rule." When $\alpha > 1/2$, the coalition needs "supermajority" or more than a certain multiple of the power of the remainder of the society. Once this first stage is completed, a subgroup can secede from or sideline the rest of the initial winning coalition if it has enough power and wishes to do so. This process continues until a *self-enforcing coalition*, which does not contain any subcoalitions that wish to engage in further rounds of eliminations, emerges. Once this coalition, which we refer to as the *ultimate ruling coalition* (URC), is formed, the society's resources are distributed according to some pre-determined rule (for example, resources may be distributed among the members of this

---

coalition according to their powers). This simple game formalizes the two key consequences of weak institutions mentioned above: (1) binding agreements on how resources will be distributed are not possible; (2) subcoalitions cannot commit to not sidelining their fellow members in a particular coalition.[2]

Our main results are as follows. First, we characterize the equilibria of this class of games under general conditions. We show that a ruling coalition always exists and is "generically" unique. Moreover, the equilibrium always satisfies some natural axioms that are motivated by the power and self-enforcement considerations mentioned above. Therefore, our analysis establishes the equivalence between an axiomatic approach to the formation of ruling coalitions (which involves the characterization of a mapping that determines the ruling coalition for any society and satisfies a number of natural axioms) and a noncooperative approach (which involves characterizing the subgame perfect equilibria of a game of coalition formation). We also show that the URC can be characterized recursively. Using this characterization, we establish the following results on the structure of URCs.

1. Despite the simplicity of the environment, the URC can consist of any number of players, and may include or exclude the most powerful individuals in the society. Consequently, the equilibrium payoff of an individual is not monotonic in his power. The most powerful will belong to the ruling coalition only if he is powerful enough to win by himself or weak enough to be a part of a smaller self-enforcing coalition.

2. An increase in α, that is, an increase in the degree of supermajority needed to eliminate opponents, does not necessarily lead to larger URCs, because it stabilizes otherwise non-self-enforcing subcoalitions, and as a result, destroys larger coalitions that would have been self-enforcing for lower values of α.

3. Self-enforcing coalitions are generally "fragile." For example, under majority rule (i.e., $\alpha = 1/2$), adding or subtracting one player from a self-enforcing coalition necessarily makes it non-self-enforcing.

4. Nevertheless, URCs are (generically) continuous in the distribution of power across individuals in the sense that a URC remains so when the powers of the players are perturbed.

5. Coalitions of certain sizes are more likely to emerge as the URC. For example, with majority rule ($\alpha = 1/2$) and a sufficiently equal distribution of powers among individuals, the URC must have size $2^k - 1$ where k is an integer (i.e., 1, 3, 7, 15,...). A similar formula for the size of the ruling coalition applies when $\alpha > 1/2$.

---

[2]The game also introduces the feature that once a particular group of individuals has been sidelined, they cannot be brought back into the ruling coalition. This feature is adopted for tractability.

We next illustrate some of the main interactions using a simple example.

Example 1 Consider two agents A and B. Denote their powers $\gamma_A > 0$ and $\gamma_B > 0$ and assume that the decision-making rule requires power-weighted majority, that is, $\alpha = 1/2$. This implies that if $\gamma_A > \gamma_B$, then starting with the coalition $\{A, B\}$, the agent A will form a majority by himself. Conversely, if $\gamma_A < \gamma_B$, then agent B will form a majority. Thus, "generically" (i.e., as long as $\gamma_A \neq \gamma_B$), one of the members of the two-person coalition can secede and form a subcoalition that is powerful enough within the original coalition. Since each agent will receive a higher share of the scarce resources in a coalition that consists of only himself than in a two-person coalition, two-person coalitions are generically not self-enforcing.

Now, consider a coalition consisting of three agents, A, B and C with powers γ_A, γ_B and γ_C, and suppose that $\gamma_A < \gamma_B < \gamma_C < \gamma_A + \gamma_B$. Clearly, no two-person coalition is self-enforcing. The lack of self-enforcing subcoalitions of $\{A, B, C\}$ implies that $\{A, B, C\}$ is itself self-enforcing. To see this, suppose, for example, that $\{A, B\}$ considers seceding from $\{A, B, C\}$. They can do so since $\gamma_A + \gamma_B > \gamma_C$. However, we know from the previous paragraph that the subcoalition $\{A, B\}$ is itself not self-enforcing, since after this coalition is established, agent B would secede or eliminate A. Anticipating this, agent A would not support the subcoalition $\{A, B\}$. A similar argument applies for all other subcoalitions. Moreover, since agent C is not powerful enough to secede from the original coalition by himself, the three-person coalition $\{A, B, C\}$ is self-enforcing and will be the ruling coalition.

Next, consider a society consisting of four individuals, A, B, C and D. Suppose that we have $\gamma_A = 3, \gamma_B = 4, \gamma_C = 5$, and $\gamma_D = 10$. D's power is insufficient to eliminate the coalition $\{A, B, C\}$ starting from the initial coalition $\{A, B, C\}$. Nevertheless, D is stronger than any two of A, B, C. This implies that any three-person coalition that includes D would not be self-enforcing. Anticipating this, any two of $\{A, B, C\}$ would decline D's offer to secede. However, $\{A, B, C\}$ is self-enforcing, thus the three agents would be happy to eliminate D. Therefore, in this example, the ruling coalition again consists of three individuals, but interestingly excludes the most powerful individual D.

The most powerful individual is not always eliminated. Consider the society with $\gamma_A = 2, \gamma_B = 4, \gamma_C = 7$ and $\gamma_D = 10$. In this case, among the three-person coalitions only $\{B, C, D\}$ is self-enforcing, and it will eliminate the weakest individual, A, and become the ruling coalition. This example also illustrates why three-person coalitions ($2^2 - 1 = 3$) may be more likely than two-person (and also four-person) coalitions.[3]

---

[3]It also shows that in contrast to approaches with unrestricted side-payments (e.g., Riker, 1962), the ruling coalition will not generally be a minimal winning coalition (the unique minimum winning coalition is $\{A, D\}$,

Although our model is abstract, it captures a range of economic forces that appear salient in nondemocratic, weakly-institutionalized polities. The historical example of Stalin's Soviet Russia illustrates this in a particularly clear manner. The Communist Party Politburo was the highest ruling body of the Soviet Union. All top government positions were held by its members. Though formally its members were elected at Party meetings, for all practical purposes the Politburo determined the fates of its members, as well as those of ordinary citizens. Soviet archives contain execution lists signed by Politburo members; sometimes a list would contain one name, but some lists from the period of 1937-39 contained hundreds or even thousands names (Conquest, 1968).

Of 40 Politburo members (28 full, 12 non-voting) appointed between 1919 and 1952, only 12 survived through 1952. Of these 12, 11 continued to hold top positions after Stalin's death in March 1953. There was a single Politburo member (Petrovsky) in 33 years who left the body and survived. Of the 28 deaths, there were 17 executions decided by the Politburo, 2 suicides, 1 death in prison immediately after arrest, and 1 assassination.

To interpret the interactions among Politburo members through the lenses of our model, imagine that the Politburo consists of five members, and to illustrate our main points, suppose that their powers are given by $\{3, 4, 5, 10, 20\}$. It can be verified that with $\alpha = 1/2$, this five-member coalition is self-enforcing. However, if either of the lower power individuals, $3, 4, 5$, or 10, dies or is eliminated, then the ruling coalition consists of the singleton, 20. If, instead, 20 dies, the ultimate ruling coalition becomes $\{3, 4, 5\}$ and eliminates the remaining most powerful individual 10. This is because 10 is unable to form an alliance with less powerful players. While the reality of Soviet politics in the first half of the century is naturally much more complicated, this simple example sheds light on three critical episodes.

The first episode is the suicides of two members of the Politburo, Tomsky and Ordzhonikidze, during 1937-38. An immediate implication of these suicides was a change in the balance of power, something akin to the elimination of 5 in the $\{3, 4, 5, 10, 20\}$ example above. In less than a year, 11 current or former members of Politburo were executed. Consistent with the ideas emphasized in our model, some of those executed in 1939 (e.g., Chubar, Kosior, Postyshev, and Ezhov) had earlier voted for the execution of Bukharin and Rykov in 1937. The second episode followed the death of Alexei Zhdanov in 1948 from a heart attack. Until Zhdanov's death, there was a period of relative "peace": no member of this body had been executed in nine years. Montefiore (2003) describes how the Zhdanov's death immediately changed the balance in the Politburo. The death gave Beria and Malenkov the possibility to have Zhdanov's supporters and associates

---

which has the minimum power among all winning coalitions).

in the government executed.[4] The third episode followed the death of Stalin himself in March 1953. Since the bloody purge of 1948, powerful Politburo members conspired in resisting any attempts by Stalin to have any of them condemned and executed. When in the Fall of 1952, Stalin charged two old Politburo members, Molotov and Mikoian, with being the "enemies of the people," the other members stood firm and blocked a possible trial (see Montefiore, 2003, or Gorlizki and Khlevniuk, 2004). After Stalin's death, Beria became the most powerful politician in Russia. He was immediately appointed the first deputy prime-minister as well as the head of the ministry of internal affairs and of the ministry of state security, the two most powerful ministries in the USSR. His ally Malenkov was appointed prime-minister, and no one succeeded Stalin as the Secretary General of the Communist Party. Yet in only 4 months, the all-powerful Beria fell victim of a military coup by his fellow Politburo members, was tried and executed. In terms of our simple example with powers $\{3, 4, 5, 10, 20\}$, Beria would correspond to 10. After 20 (Stalin) is out of the picture, $\{3, 4, 5\}$ becomes the ultimate ruling coalition, so 10 must be eliminated.

Similar issues arise in other dictatorships when top figures were concerned with others becoming too powerful. These considerations also appear to be particularly important in international relations, especially when agreements have to be reached under the shadow of the threat of war (e.g., Powell, 1999). For example, following both World Wars, many important features of the peace agreements were influenced by the desire that the emerging balance of power among states should be self-enforcing. In this context, small states were viewed as attractive because they could combine to contain threats from larger states but they would be unable to become dominant players. Similar considerations were paramount after Napoleon's ultimate defeat in 1815. In this case, the victorious nations designed the new political map of Europe at the Vienna Congress, and special attention was paid to balancing the powers of Britain, Germany and Russia, to ensure that "... their equilibrium behaviour... maintain the Vienna settlement" (Slantchev, 2005).[5]

Our paper is related to models of bargaining over resources, particularly in the context of political decision-making (e.g., models of legislative bargaining such as Baron and Ferejohn, 1989, Calvert and Dietz, 1996, Jackson and Moselle, 2002). Our approach differs from these papers, since we do not impose any specific bargaining structure and focus on self-enforcing

---

[4]In contrast to the two other episodes from the Soviet Politburo we discuss here, the elimination of the associates of Zhadanov could also be explained by competition between two groups within the Politburo rather than by competition among all members and lack of commitment, which are the ideas emphasized by our model.

[5]Other examples of potential applications of our model in political games are provided in Pepinsky (2007), who uses our model to discuss issues of coalition formation in nondemocratic societies.

ruling coalitions.[6]

More closely related to our work are the models of on equilibrium coalition formation, which combine elements from both cooperative and noncooperative game theory (e.g., Peleg, 1980, Hart and Kurz, 1983, Greenberg and Weber, 1993, Chwe, 1994, Bloch, 1996, Mariotti, 1997, Ray, 2007, Ray and Vohra, 1997, 1999, 2001, Seidmann and Winter, 1998, Konishi and Ray, 2001, Maskin, 2003, Eguia, 2006, Pycia, 2006). The most important difference between our approach and the previous literature on coalition formation is that, motivated by political settings, we assume that the majority (or supermajority) of the members of the society can impose their will on those players who are not a part of the majority. This feature both changes the nature of the game and also introduces "negative externalities" as opposed to the positive externalities and free-rider problems upon which the previous literature focuses (Ray and Vohra, 1999, and Maskin, 2003). A second important difference is that most of these works assume the possibility of binding commitments (Ray and Vohra, 1997, 1999), while we suppose that players have no commitment power. Despite these differences, there are important parallels between our results and the insights of this literature. For example, Ray (1979) and Ray and Vohra (1997, 1999) emphasize that the internal stability of a coalition influences whether it can block the formation of other coalitions, including the grand coalition. In the related context of risk-sharing arrangements, Bloch, Genicot, and Ray (2006) show that stability of subgroups threatens the stability of a larger group.[7] Another related approach to coalition formation is developed by Moldovanu and Winter (1995), who study a game in which decisions require appoval by all members of a coalition and show the relationship of the resulting allocations to the core of a related cooperative game.[8] Finally, Skaperdas (1998) and Tan and Wang (1999) investigate coalition formation in dynamic contests. Nevertheless, none of these papers study self-enforcing coalitions in political games without commitment, or derive existence, generic uniqueness and characterization results similar to those in our paper.

The rest of the paper is organized as follows. Section 2 introduces the formal setup. Section 3 provides our axiomatic treatment. Section 4 characterizes subgame perfect equilibria of the

---

[6]See also Perry and Reny (1994), Moldovanu and Jehiel (1999), and Gomes and Jehiel (2005) for models of bargaining with a coalition structure.

[7]In this respect, our paper is also related to work on "coalition-proof" Nash equilibrium or rationalizability, e.g., Bernheim, Peleg, and Whinston (1987), Moldovanu (1992), Ambrus (2006). These papers allow deviations by coalitions in noncooperative games, but impose that only stable coalitions can form. In contrast, these considerations are captured in our model by the game of coalition formation and by the axiomatic analysis.

[8]Our game can also be viewed as a "hedonic game" since the utility of each player is determined by the composition of the ultimate coalition he belongs to. However, it is not a special case of hedonic games defined and studied in Bogomolnaia and Jackson (2002), Banerjee, Konishi, and Sonmez (2001), and Barbera and Gerber (2007), because of the dynamic interactions introduced by the self-enforcement considerations. See Le Breton, Ortuno-Ortin, and Weber (2008) for an application of hedonic games to coalition formation.

extensive-form game of coalition formation. It then establishes the equivalence between the ruling coalition of Section 3 and the equilibria of this extensive-form game. Section 5 contains our main results on the nature and structure of ruling coalitions in political games. Section 6 concludes. The Appendix contains the proofs of all the results presented in the text.

2   The Political Game

Let $\mathcal{I}$ denote the collection of all individuals, which is assumed to be finite. The non-empty subsets of $\mathcal{I}$ are *coalitions* and the set of coalitions is denoted by $\mathcal{C}$. In addition, for any $X \subset \mathcal{I}$, $\mathcal{C}_X$ denotes the set of coalitions that are subsets of X and $|X|$ is the number of members in X. In each period there is a designated *ruling coalition*, which can change over time. The game starts with ruling coalition N, and eventually the *ultimate ruling coalition* (URC) forms. We assume that if the URC is X, then player i obtains *baseline* utility $w_i(X) \in \mathbb{R}$. We denote $w(\cdot) \equiv \{w_i(\cdot)\}_{i \in \mathcal{I}}$.

Our focus is on how differences in the powers of individuals map into political decisions. We define a *power* mapping to summarize the powers of different individuals in $\mathcal{I}$:

$$\gamma : \mathcal{I} \to \mathbb{R}_{++},$$

where $\mathbb{R}_{++} = \mathbb{R}_+ \setminus \{0\}$. We refer to $\gamma_i \equiv \gamma(i)$ as the political *power* of individual $i \in \mathcal{I}$. In addition, we denote the set of all possible power mappings by $\mathcal{R}$ and a power mapping γ restricted to some coalition $N \subset \mathcal{I}$ by $\gamma|_N$ (or by γ when the reference to N is clear). The power of a coalition X is $\gamma_X \equiv \sum_{i \in X} \gamma_i$.

Coalition $Y \subset X$ is *winning* within coalition X if and only if $\gamma_Y > \alpha \gamma_X$, where $\alpha \in [1/2, 1)$ is a fixed parameter referring to the degree of (weighted) supermajority. Naturally, $\alpha = 1/2$ corresponds to majority rule. Moreover, since $\mathcal{I}$ is finite, there exists a large enough α (still less than 1) that corresponds to unanimity rule. We denote the set of coalitions that are winning within X by $\mathcal{W}_X$. Since $\alpha \geq 1/2$, if $Y, Z \in \mathcal{W}_X$, then $Y \cap Z \neq \varnothing$.

The assumption that payoffs are given by the mapping $w(\cdot)$ implies that a coalition cannot commit to a redistribution of resources or payoffs among its members (for example, a coalition consisting of two individuals with powers 1 and 10 cannot commit to share the resource equally if it becomes the URC). We assume that the *baseline* payoff functions, $w_i(X) : \mathcal{I} \times \mathcal{C} \to \mathbb{R}$ for any $i \in N$, satisfy the following properties.

Assumption 1 *Let $i \in \mathcal{I}$ and $X, Y \in \mathcal{C}$. Then:*

(1) If $i \in X$ and $i \notin Y$, then $w_i(X) > w_i(Y)$ [i.e., each player prefers to be part of the URC].

7

(2) *For $i \in X$ and $i \in Y$, $w_i(X) > w_i(Y) \iff \gamma_i/\gamma_X > \gamma_i/\gamma_Y$ ($\iff \gamma_X < \gamma_Y$) [i.e., for any two URCs that he is part of, each player prefers the one where his relative power is greater].*

(3) *If $i \notin X$ and $i \notin Y$, then $w_i(X) = w_i(Y) \equiv w_i^-$ [i.e., a player is indifferent between URCs he is not part of].*

This assumption is natural and captures the idea that each player's payoff depends positively on his relative strength in the URC. A specific example of function $w(\cdot)$ that satisfies these requirements is sharing of a pie between members of the ultimate ruling coalition proportional to their power:

$$w_i(X) = \frac{\gamma_{X \cap \{i\}}}{\gamma_X} = \begin{cases} \gamma_i/\gamma_X & \text{if } i \in X \\ 0 & \text{if } i \notin X \end{cases}. \tag{1}$$

The reader may want to assume (1) throughout the text for interpretation purposes, though this specific functional form is not used in any of our results or proofs.

We next define the extensive-form complete information game $\Gamma = (N, \gamma|_N, w(\cdot), \alpha)$, where $N \in \mathcal{C}$ is the initial coalition, γ is the power mapping, $w(\cdot)$ is a payoff mapping that satisfies Assumption 1, and $\alpha \in [1/2, 1)$ is the degree of supermajority; denote the collection of such games by $\mathcal{G}$. Also, let $\varepsilon > 0$ be sufficiently small such that for any $i \in N$ and any $X, Y \in \mathcal{C}$, we have

$$w_i(X) > w_i(Y) \implies w_i(X) > w_i(Y) + 2\varepsilon \tag{2}$$

(this holds for sufficiently small $\varepsilon > 0$ since $\mathcal{I}$ is a finite set). This immediately implies that for any $X \in \mathcal{C}$ with $i \in X$, we have

$$w_i(X) - w_i^- > \varepsilon. \tag{3}$$

The extensive form of the game $\Gamma = (N, \gamma|_N, w(\cdot), \alpha)$ is as follows. Each *stage* j of the game starts with some ruling coalition N_j (at the beginning of the game $N_0 = N$). Then the *stage game* proceeds with the following steps:

1. Nature randomly picks agenda setter $a_{j,q} \in N_j$ for $q = 1$.

2. [Agenda-setting step] Agenda setter $a_{j,q}$ makes proposal $P_{j,q} \in \mathcal{C}_{N_j}$, which is a subcoalition of N_j such that $a_{j,q} \in P_{j,q}$ (for simplicity, we assume that a player cannot propose to eliminate himself).

3. [Voting step] Players in $P_{j,q}$ vote sequentially over the proposal (we assume that players in $N_j \setminus P_{j,q}$ automatically vote against this proposal). More specifically, Nature randomly chooses the first voter, $v_{j,q,1}$, who then casts his vote vote $\tilde{v}(v_{j,q,1}) \in \{\tilde{y}, \tilde{n}\}$ (Yes or No), then Nature chooses the second voter $v_{j,q,2} \neq v_{j,q,1}$, etc. After all $|P_{j,q}|$ players have voted, the game proceeds to step 4 if players who supported the proposal form a winning coalition within N_j (i.e., if $\{i \in P_{j,q} : \tilde{v}(i) = \tilde{y}\} \in \mathcal{W}_{N_j}$), and otherwise it proceeds to step 5.

4. If $P_{j,q} = N_j$, then the game proceeds to step 6. Otherwise, players from $N_j \setminus P_{j,q}$ are eliminated and the game proceeds to step 1 with $N_{j+1} = P_{j,q}$ (and j increases by 1 as a new transition has taken place).

5. If $q < |N_j|$, then next agenda setter $a_{j,q+1} \in N_j$ is randomly picked by Nature among members of N_j who have not yet proposed at this stage (so $a_{j,q+1} \neq a_{j,r}$ for $1 \leq r \leq q$), and the game proceeds to step 2 (with q increased by 1). If $q = |N_j|$, the game proceeds to step 6.

6. N_j becomes the ultimate ruling coalition. Each player $i \in N$ receives total payoff

$$U_i = w_i(N_j) - \varepsilon \sum_{1 \leq k \leq j} \mathbf{I}_{\{i \in N_k\}}, \tag{4}$$

where $\mathbf{I}_{\{\cdot\}}$ is the indicator function taking the value of 0 or 1.

The payoff function (4) captures the idea that an individual's overall utility is the difference between the baseline $w_i(\cdot)$ and disutility from the number of transitions (rounds of elimination) this individual is involved in. The arbitrarily small cost ε can be interpreted as a cost of eliminating some of the players from the coalition or as an organizational cost that individuals have to pay each time a new coalition is formed. Alternatively, ε may be viewed as a means to refine out equilibria where order of moves matters for the outcome. Note that Γ is a finite game: the total number of moves, including those of Nature, does not exceed $4 |N|^3$. Notice also that this game form introduces sequential voting in order to avoid issues of individuals playing weakly-dominated strategies. Our analysis below will establish that the main results hold regardless of the specific order of votes chosen by Nature.[9]

3 Axiomatic Analysis

Before characterizing the equilibria of the dynamic game Γ, we take a brief detour and introduce four *axioms* motivated by the structure of the game Γ. Although these axioms are motivated by game Γ, they can also be viewed as natural axioms to capture the salient economic forces discussed in the introduction. The analysis in this section identifies an outcome mapping $\Phi : \mathcal{G} \rightrightarrows \mathcal{C}$ that satisfies these axioms and determines the set of (admissible) URCs corresponding to each game Γ. This analysis will be useful for two reasons. First, it will reveal certain attractive features of the game presented in the previous section. Second, we will show in the next section that equilibrium URCs of this game coincide with the outcomes picked by the mapping Φ.

More formally, consider the set of games $\Gamma = (N, \gamma|_N, w(\cdot), \alpha) \in \mathcal{G}$. Holding γ, w and α fixed, consider the correspondence $\phi : \mathcal{C} \rightrightarrows \mathcal{C}$ defined by $\phi(N) = \Phi(N, \gamma|_N, w, \alpha)$ for any $N \in \mathcal{C}$.

---

[9]See Acemoglu, Egorov, and Sonin (2006) both for the analysis of a game with simultaneous voting and a stronger equilibrium notion, and for an example showing how, in the absence of the cost $\varepsilon > 0$, the order of moves may matter.

We adopt the following axioms on ϕ (or alternatively on Φ).

Axiom 1 (*Inclusion*) For any $X \in \mathcal{C}$, $\phi(X) \neq \varnothing$ and if $Y \in \phi(X)$, then $Y \subset X$.

Axiom 2 (*Power*) For any $X \in \mathcal{C}$, $Y \in \phi(X)$ only if $Y \in \mathcal{W}_X$.

Axiom 3 (*Self-Enforcement*) For any $X \in \mathcal{C}$, $Y \in \phi(X)$ only if $Y \in \phi(Y)$.

Axiom 4 (*Rationality*) For any $X \in \mathcal{C}$, for any $Y \in \phi(X)$ and for any $Z \subset X$ such that $Z \in \mathcal{W}_X$ and $Z \in \phi(Z)$, we have that $Z \notin \phi(X) \iff \gamma_Y < \gamma_Z$.

Motivated by Axiom 3, we define the notion of a self-enforcing coalition as a coalition that "selects itself". This notion will be used repeatedly in the rest of the paper.

Definition 1 Coalition $X \in P(\mathcal{I})$ is *self-enforcing* if $X \in \phi(X)$.

Axiom 1, inclusion, implies that ϕ maps into subcoalitions of the coalition in question (and that it is defined, i.e., $\phi(X) \neq \varnothing$). It therefore captures the feature introduced in Γ that players that have been eliminated (sidelined) cannot rejoin the ruling coalition. Axiom 2, the power axiom, requires a ruling coalition be a winning coalition. Axiom 3, the self-enforcement axiom, captures the key interactions in our model. It requires that any coalition $Y \in \phi(X)$ should be self-enforcing according to Definition 1. This property corresponds to the notion that in terms of game Γ, if coalition Y is reached along the equilibrium path, then there should not be any deviations from it. Finally, Axiom 4 requires that if two coalitions $Y, Z \subset X$ are both winning and self-enforcing and all players in $Y \cap Z$ strictly prefer Y to Z, then $Z \notin \phi(X)$ (i.e., Z cannot be the selected coalition). Intuitively, all members of winning coalition Y (both those in $Y \cap Z$ by assumption and those in $Y \setminus Z$ because they prefer to be in the URC) strictly prefer Y to Z; hence, Z should not be chosen in favor of Y. This interpretation allows us to call Axiom 4 the Rationality Axiom. In terms of game Γ, this axiom captures the notion that, when he has the choice, a player will propose a coalition in which his payoff is greater.

At the first glance, Axioms 1–4 may appear relatively mild. Nevertheless, they are strong enough to pin down a unique mapping ϕ. Moreover, under the following assumption, these axioms also imply that this unique mapping ϕ is single valued.

Assumption 2 *The power mapping γ is generic in the sense that if for any $X, Y \in \mathcal{C}$, $\gamma_X = \gamma_Y$ implies $X = Y$. We also say that coalition N is generic or that numbers $\{\gamma_i\}_{i \in N}$ are generic if mapping $\gamma|_N$ is generic.*

Intuitively, this assumption rules out distributions of powers among individuals such that two different coalitions have exactly the same total power. Notice that mathematically, genericity assumption is without much loss of generality since the set of vectors $\{\gamma_i\}_{i\in\mathcal{I}} \in \mathbb{R}_{++}^{|\mathcal{I}|}$ that are not generic has Lebesgue measure 0 (in fact, it is a union of a finite number of hyperplanes in $\mathbb{R}_{++}^{|\mathcal{I}|}$).

Theorem 1 *Fix a collection of players $\mathcal{I}$, a power mapping γ, a payoff function $w(\cdot)$ such that Assumption 1 holds, and $\alpha \in [1/2, 1)$. Then:*

1. There exists a unique mapping ϕ that satisfies Axioms 1–4. Moreover, when γ is generic (i.e. under Assumption 2), ϕ is single-valued.

2. This mapping ϕ may be obtained by the following inductive procedure. For any $k \in \mathbb{N}$, let $\mathcal{C}^k = \{X \in \mathcal{C} : |X| = k\}$. Clearly, $\mathcal{C} = \cup_{k\in\mathbb{N}}\mathcal{C}^k$. If $X \in \mathcal{C}^1$, then let $\phi(X) = \{X\}$. If $\phi(Z)$ has been defined for all $Z \in \mathcal{C}^n$ for all $n < k$, then define $\phi(X)$ for $X \in \mathcal{C}^k$ as

$$\phi(X) = \underset{A\in\mathcal{M}(X)\cup\{X\}}{\operatorname{argmin}} \gamma_A, \tag{5}$$

where

$$\mathcal{M}(X) = \{Z \in \mathcal{C}_X \setminus \{X\} : Z \in \mathcal{W}_X \text{ and } Z \in \phi(Z)\}. \tag{6}$$

Proceeding inductively $\phi(X)$ is defined for all $X \in \mathcal{C}$.

The intuition for the inductive procedure is as follows. For each X, (6) defines $\mathcal{M}(X)$ as the set of proper subcoalitions which are both winning and self-enforcing. Equation (5) then picks the coalitions in $\mathcal{M}(X)$ that have the least power. When there are no proper winning and self-enforcing subcoalitions, $\mathcal{M}(X)$ is empty and X becomes the URC), which is captured by (5). The proof of this theorem, like all other proofs, is in the Appendix.

Theorem 1 establishes not only that ϕ is uniquely defined, but also that when Assumption 2 holds, it is single-valued. In this case, with a slight abuse of notation, we write $\phi(X) = Y$ instead of $\phi(X) = \{Y\}$.

Corollary 1 *Take any collection of players $\mathcal{I}$, power mapping γ, payoff function $w(\cdot)$, and $\alpha \in [1/2, 1)$. Let ϕ be the unique mapping satisfying Axioms 1–4. Then for any $X, Y, Z \in \mathcal{C}$, $Y, Z \in \phi(X)$ implies $\gamma_Y = \gamma_Z$. Coalition N is self-enforcing, that is, $N \in \phi(N)$, if and only if there exists no coalition $X \subset N$, $X \neq N$, that is winning within N and self-enforcing. Moreover, if N is self-enforcing, then $\phi(N) = \{N\}$.*

Corollary 1, which immediately follows from (5) and (6), summarizes the basic results on self-enforcing coalitions. In particular, Corollary 1 says that a coalition that includes a winning and

self-enforcing subcoalition cannot be self-enforcing. This captures the notion that the stability of smaller coalitions undermines the stability of larger ones.

As an illustration to Theorem 1, consider again three players A, B and C and suppose that $\alpha = 1/2$. For any $\gamma_A < \gamma_B < \gamma_C < \gamma_A + \gamma_B$, Assumption 2 is satisfied and it is easy to see that $\{A\}$, $\{B\}$, $\{C\}$, and $\{A, B, C\}$ are self-enforcing coalitions, whereas $\phi(\{A, B\}) = \{B\}$, $\phi(\{A, C\}) = \phi(\{B, C\}) = \{C\}$. In this case, $\phi(X)$ is a singleton for any X. On the other hand, if $\gamma_A = \gamma_B = \gamma_C$, all coalitions except $\{A, B, C\}$ would be self-enforcing, while $\phi(\{A, B, C\}) = \{\{A, B\}, \{B, C\}, \{A, C\}\}$ in this case.

4 Equilibrium Characterization

We now characterize the Subgame Perfect Equilibria (SPE) of game Γ defined in Section 2 and show that they correspond to the ruling coalitions identified by the axiomatic analysis in the previous section. The next subsection provides the main results. We then provide a sketch of the proofs. The formal proofs are contained in the Appendix.

4.1 Main Results

The following two theorems characterize the *Subgame Perfect Equilibrium* (SPE) of game $\Gamma = (N, \gamma|_N, w, \alpha)$ with initial coalition N. As usual, a strategy profile σ in Γ is a SPE if σ induces continuation strategies that are best responses to each other starting in any subgame of Γ, denoted Γ_h, where h denotes the history of the game, consisting of actions in past periods (stages and steps).

Theorem 2 *Suppose that $\phi(N)$ satisfies Axioms 1-4 (cfr. (5) in Theorem 1). Then, for any $K \in \phi(N)$, there exists a pure strategy profile σ_K that is an SPE and leads to URC K in at most one transition. In this equilibrium player $i \in N$ receives payoff*

$$U_i = w_i(K) - \varepsilon \mathbf{I}_{\{i \in K\}} \mathbf{I}_{\{N \neq K\}}. \tag{7}$$

This equilibrium payoff does not depend on the random moves by Nature.

Theorem 2 establishes that there exists a pure strategy equilibrium leading to any coalition that is in the set $\phi(N)$ defined in the axiomatic analysis of Theorem 1.[10] This is intuitive in view of the analysis in the previous section: when each player anticipates members of a self-enforcing ruling coalition to play a strategy profile such that they will turn down any offers other than K

---

[10]It can also be verified that Theorem 2 holds even when $\varepsilon = 0$. The assumption that $\varepsilon > 0$ is used in Theorem 3.

and they will accept K, it is in the interest of all the players in K to play such a strategy for any history. This follows because the definition of the set $\phi(N)$ implies that only deviations that lead to ruling coalitions that are not self-enforcing or not winning could be profitable. But the first option is ruled out by induction while a deviation to a non-winning URC will be blocked by sufficiently many players. The payoff in (7) is also intuitive. Each player receives his baseline payoff $w_i(K)$ resulting from URC K and then incurs the cost ε if he is part of K and if the initial coalition N is not equal to K (because in this latter case, there will be one transition). Notice that Theorem 2 is stated without Assumption 2 and does not establish uniqueness. The next theorem strengthens these results under Assumption 2.

Theorem 3 *Suppose Assumption 2 holds and suppose $\phi(N) = K$. Then any (pure or mixed strategy) SPE results in K as the URC. The payoff player $i \in N$ receives in this equilibrium is given by* (7).

Since Assumption 2 holds, the mapping ϕ is single-valued (with $\phi(N) = K$). Theorem 3 then shows that even though the SPE may not be unique in this case, any SPE will lead to K as the URC. This is intuitive in view of our discussion above. Because any SPE is obtained by backward induction, multiplicity of equilibria results only when some player is indifferent between multiple actions at a certain nod. However, as we show, this may only happen when a player has no effect on equilibrium play and his choice between different actions has no effect on URC (in particular, since ϕ is single-valued in this case, a player cannot be indifferent between actions that will lead to different URCs).

It is also worth noting that the SPE in Theorems 2 and 3 is "coalition-proof". Since the game Γ incorporates both dynamic and coalitional effects and is finite, the relevant concept of coalition-proofness is Bernheim, Peleg, and Whinston's (1987) Perfectly Coalition-Proof Nash Equilibrium (PCPNE). This equilibrium refinement requires that the candidate equilibrium should be robust to deviations by coalitions in all subgames when the players take into account the possibility of further deviations. Since Γ introduces more general coalitional deviations explicitly, it is natural to expect the SPE in Γ to be PCPNE. Indeed, if Assumption 2 holds, it is straightforward to prove that the set of PCPNE coincides with the set of SPE.[11]

4.2 Sketch of the Proofs

We now provide an outline of the argument leading to the proofs of the main results presented in the previous subsection and we present two key lemmas that are central for these theorems.

---

[11] A formal proof of this result follows from Lemma 2 below and is available from the authors upon request.

Consider the game Γ and let ϕ be as defined in (5). Take any coalition $K \in \phi(N)$. We will outline the construction of the pure strategy profile σ_K which will be a SPE and lead to K as the URC.

Let us first rank all coalitions so as to "break ties" (which are possible, since we have not yet imposed Assumption 2). In particular, $n : \mathcal{C} \longleftrightarrow \{1, \ldots, 2^{|\mathcal{I}|} - 1\}$ be a one-to-one mapping such that for any $X, Y \in \mathcal{C}$, $\gamma_X > \gamma_Y \Rightarrow n(X) > n(Y)$, and if for some $X \neq K$ we have $\gamma_X = \gamma_K$, then $n(X) > n(K)$ (how the ties among other coalitions are broken is not important). With this mapping, we have thus ranked (enumerated) all coalitions such that stronger coalitions are given higher numbers, and coalition K receives the smallest number among all coalitions with the same power. Now define the mapping $\chi : \mathcal{C} \to \mathcal{C}$ as

$$\chi(X) = \underset{Y \in \phi(X)}{\operatorname{argmin}} n(Y). \tag{8}$$

Intuitively, this mapping picks an element of $\phi(X)$ for any X and satisfies $\chi(N) = K$. Also, note that χ is a *projection* in the sense that $\chi(\chi(X)) = \chi(X)$. This follows immediately since Axiom 3 implies $\chi(X) \in \phi(\chi(X))$ and Corollary 1 implies that $\phi(\chi(X))$ is a singleton.

The key to constructing a SPE is to consider off-equilibrium path behavior. To do this, consider a subgame in which we have reached a coalition X (i.e., j transitions have occurred and $N_j = X$) and let us try to determine what the URC would be if proposal Y is accepted starting in this subgame. If $Y = X$, then the game will end, and thus X will be the URC. If, on the other hand, $Y \neq X$, then the URC must be some subset of Y. Let us define the strategy profile σ_K such that the URC will be $\chi(Y)$. We denote this (potentially off-equilibrium path) URC following the acceptance of proposal Y by $\psi_X(Y)$, so that

$$\psi_X(Y) = \begin{cases} \chi(Y) & \text{if } Y \neq X; \\ X & \text{otherwise.} \end{cases} \tag{9}$$

By Axiom 1 and equations (8) and (9), we have that

$$X = Y \iff \psi_X(Y) = X. \tag{10}$$

We will introduce one final concept before defining profile σ_K. Let $F_X(i)$ denote the "favorite" coalition of player i if the current ruling coalition is X. Naturally, this will be the weakest coalition among coalitions that are winning within X, that are self-enforcing and that include player i. If there are several such coalitions, the definition of $F_X(i)$ picks the one with the smallest n, and if there are none, it picks X itself. Therefore,

$$F_X(i) = \underset{Y \in \{Z : Z \subset X, Z \in \mathcal{W}_X, \chi(Z) = Z, Z \ni i\} \cup \{X\}}{\operatorname{argmin}} n(Y). \tag{11}$$

14

Similarly, we define the "favorite" coalition of players $Y \subset X$ starting with X at the current stage. This is again the weakest coalition among those favored by members of Y, thus

$$F_X\left(Y\right) = \begin{cases} \underset{\{Z : \exists i \in Y : F_X(i) = Z\}}{\operatorname{argmin}} \; n\left(Z\right) & \text{if } Y \neq \varnothing; \\ X & \text{otherwise.} \end{cases} \tag{12}$$

Equation (12) immediately implies that

$$\text{For all } X \in \mathcal{C} : F_X\left(\varnothing\right) = X \text{ and } F_X\left(X\right) = \chi\left(X\right). \tag{13}$$

The first part is true by definition. The second part follows, since for all $i \in \chi\left(X\right)$, $\chi\left(X\right)$ is feasible in the minimization (11), and it has the lowest number n among all winning self-enforcing coalitions by (8) and (5) (otherwise there would exist a self-enforcing coalition Z that is winning within X and satisfies $\gamma_Z < \gamma_{\chi(X)}$, which would imply that ϕ violates Axiom 4). Therefore, it is the favorite coalition of all $i \in \chi\left(X\right)$ and thus $F_X\left(X\right) = \chi\left(X\right)$.

Now we are ready to define profile σ_K. Take any history h and denote the player who is supposed to move after this history $a = a\left(h\right)$ if after h, we are at an agenda-setting step, and $v = v\left(h\right)$ if we are at a voting step deciding on some proposal P (and in this case, let a be the agenda-setter who made proposal P). Also denote the set of potential agenda setters at this stage of the game by A. Finally, recall that $\tilde{n}$ denotes a vote of "No" and $\tilde{y}$ is a vote of "Yes". Then σ_K is the following simple strategy profile where each agenda setter proposes his favorite coalition in the continuation game (given current coalition X) and each voter votes "No" against proposal P if the URC following P excludes him or he expects another proposal that he will prefer to come shortly.

$$\sigma_K = \begin{cases} \text{agenda-setter } a \text{ proposes } P = F_X\left(a\right); \\ \text{voter } v \text{ votes } \begin{cases} \tilde{n} & \begin{array}{l} \text{if either } v \notin \psi_X\left(P\right) \text{ or} \\ v \in F_X\left(A\right), \, P \neq F_X\left(A \cup \{a\}\right), \text{ and } \gamma_{F_X(A)} \leq \gamma_{\psi_X(P)}; \end{array} \\ \tilde{y} & \text{otherwise.} \end{cases} \end{cases} \tag{14}$$

In particular, notice that $v \in F_X\left(A\right)$ and $P \neq F_X\left(A \cup \{a\}\right)$ imply that voter v is part of a *different* coalition proposal that will be made by some future agenda setter at this stage of the game if the current voting fails, and $\gamma_{F_X(A)} \leq \gamma_{\psi_X(P)}$ implies that this voter will receive weakly higher payoff under this alternative proposal. This expression makes it clear that σ_K is similar to a "truth-telling" strategy; each individual makes proposals according to his preferences (constrained by what is feasible) and votes truthfully.

With the strategy profile σ_K defined, we can state the main lemma, which will essentially constitute the bulk of the proof of Theorem 2. For this lemma, also denote the set of voters

who already voted "Yes" at history h by V^+, the set of voters who already voted "No" by V^-. Then, $V = P \setminus (V^+ \cup V^- \cup \{v\})$ denotes the set of voters who will vote after player v.

Lemma 1 *Consider the subgame Γ_h of game Γ after history h in which there were exactly j_h transitions and let the current coalition be X. Suppose that strategy profile σ_K defined in (14) is played in Γ_h. Then:*

(a) If h is at agenda-setting step, the URC is $R = F_X(A \cup \{a\})$; if h is at voting step and $V^+ \cup \{i \in \{v\} \cup V : i \text{ votes } \tilde{y} \text{ in } \sigma_K\} \in \mathcal{W}_X$, then the URC is $R = \psi_X(P)$; and otherwise $R = F_X(A)$.

(b) If h is at the voting step and proposal P will be accepted, player $i \in X$ receives payoff

$$U_i = w_i(R) - \varepsilon \left(j_h + \mathbf{I}_{\{P \neq X\}} \left(\mathbf{I}_{\{i \in P\}} + \mathbf{I}_{\{R \neq P\}} \mathbf{I}_{\{i \in R\}} \right) \right). \tag{15}$$

Otherwise (if proposal P will be rejected or if h is at agenda-setting step), then player $i \in X$ receives payoff

$$U_i = w_i(R) - \varepsilon \left(j_h + \mathbf{I}_{\{R \neq X\}} \mathbf{I}_{\{i \in R\}} \right). \tag{16}$$

The intuition for the results in this lemma is straightforward in view of the construction of the strategy profile σ_K. In particular, part (a) defines what the URC will be. This follows immediately from σ_K. For example, if we are at an agenda-setting step, then the URC will be the favorite coalition of the set of remaining agenda setters, given by $A \cup \{a\}$. This is an immediate implication of the fact that according to the strategy profile σ_K, each player will propose his favorite coalition and voters will vote $\tilde{n}$ ("No") against current proposals if the strategy profile σ_K will induce a more preferred outcome for them in the remainder of this stage. Part (b) simply defines the payoff to each player as the difference between the baseline payoff, $w_i(R)$, as a function of the URC R defined in part (a), and the costs associated with transitions.

Given Lemma 1, Theorem 2 then follows if strategy profile σ_K is a SPE (because in this case URC will be K and it will be reached with at most one transition). With σ_K defined in (14), it is clear that no player can profitably deviate in any subgame.

The next lemma strengthens Lemma 1 for the case in which Assumption 2 holds by establishing that any SPE will lead to the same URC and payoffs as those in Lemma 1.

Lemma 2 *Suppose Assumption 2 holds and $\phi(N) = \{K\}$. Let σ_K be defined in (14). Then for any SPE σ (in pure or mixed strategies) and for any history h, the equilibrium plays induced by σ and by σ_K in the subgame Γ_h will lead to the same URC and to identical payoffs for each player.*

Since $\phi(N) = \{K\}$, Theorem 3 follows as an immediate corollary of this lemma (with $h = \varnothing$).

5 The Structure of Ruling Coalitions

In this section, we present several results on the structure of URCs. Given the equivalence result (Theorems 2 and 3), we will make use of the axiomatic characterization in Theorem 1. Throughout, unless stated otherwise, we fix a game $\Gamma = (N, \gamma, w(\cdot), \alpha)$ with w satisfying Assumption 1 and $\alpha \in [1/2, 1)$. In addition, to simplify the analysis in this section, we assume throughout that Assumption 2 holds and we also impose:

Assumption 3 For no $X, Y \in \mathcal{C}$ such that $X \subset Y$ the equality $\gamma_Y = \alpha\gamma_X$ is satisfied.

Assumption 3 guarantees that a small perturbation of a non-winning coalition Y does not make it winning. Similar to Assumption 2, this assumption fails only in a set of Lebesgue measure 0 (in fact, it coincides with Assumption 2 when $\alpha = 1/2$). All proofs are again provided in the Appendix.

5.1 Robustness

We start with the result that the set of self-enforcing coalitions is open (in the standard topology); this is not only interesting per se but facilitates further proofs. Note that for game $\Gamma = (N, \gamma, w(\cdot), \alpha)$, a power mapping γ (or more explicitly $\gamma|_N$) is given by a $|N|$-dimensional vector $\{\gamma_i\}_{i \in N} \subset \mathbb{R}_{++}^{|N|}$. Denote the subset of vectors $\{\gamma_i\}_{i \in N}$ that satisfy Assumptions 2 and 3 by $\mathcal{A}(N)$, and the subset of $\mathcal{A}(N)$ for which $\Phi\left(N, \{\gamma_i\}_{i \in N}, w, \alpha\right) = N$ (i.e., the subset of power distributions for which coalition N is stable) by $\mathcal{S}(N)$ and let $\mathcal{N}(N) = \mathcal{A}(N) \setminus \mathcal{S}(N)$.

Lemma 3 *1. The set of power allocations that satisfy Assumptions 2 and 3, $\mathcal{A}(N)$, its subsets for which coalition N is self-enforcing, $\mathcal{S}(N)$, and its subsets for which coalition N is not self-enforcing, $\mathcal{N}(N)$, are open sets in $\mathbb{R}_{++}^{|N|}$. The set $\mathcal{A}(N)$ is also dense in $\mathbb{R}_{++}^{|N|}$.*

2. Each connected component of $\mathcal{A}(N)$ lies entirely within either $\mathcal{S}(N)$ or $\mathcal{N}(N)$.

An immediate corollary of Lemma 3 is that if the distributions of powers in two different games are "close," then these two games will have the same URC and also that the inclusion of sufficiently weak players will not change the URC. To state and prove this proposition, endow the set of mappings γ, $\mathcal{R}$, with the sup-metric, so that $(\mathcal{R}, \rho)$ is a metric space with $\rho(\gamma, \gamma') = \sup_{i \in \mathcal{I}} |\gamma_i - \gamma_i'|$. A δ-neighborhood of γ is $\{\gamma' \in \mathcal{R} : \rho(\gamma, \gamma') < \delta\}$.

Proposition 1 *Consider $\Gamma = (N, \gamma, w(\cdot), \alpha)$ with $\alpha \in [1/2, 1)$. Then:*

1. There exists $\delta > 0$ such that if $\gamma' : N \to \mathbb{R}_{++}$ lies within δ-neighborhood of γ, then $\Phi(N, \gamma, w, \alpha) = \Phi(N, \gamma', w, \alpha)$.

2. There exists $\delta' > 0$ such that if $\alpha' \in [1/2, 1)$ satisfies $|\alpha' - \alpha| < \delta'$, then $\Phi(N, \gamma, w, \alpha) = \Phi(N, \gamma, w, \alpha')$.

3. Let $N = N_1 \cup N_2$ with N_1 and N_2 disjoint. Then, there exists $\delta > 0$ such that for all N_2 with $\gamma_{N_2} < \delta$, $\phi(N_1) = \phi(N_1 \cup N_2)$.

This proposition is intuitive in view of the results in Lemma 3. It implies that URCs have some degree of continuity and will not change as a result of small changes in power or in the rules of the game.

5.2 Fragility of Self-Enforcing Coalitions

Although the structure of ruling coalitions is robust to small changes in the distribution of power within the society, it may be fragile to more sizeable shocks. The next proposition shows that in fact the addition or the elimination of a single member of the self-enforcing coalition turns out to be such a sizable shock when $\alpha = 1/2$.

Proposition 2 *Suppose $\alpha = 1/2$ and fix a power mapping $\gamma : \mathcal{I} \to \mathbb{R}_{++}$. Then:*

1. If coalitions X and Y such that $X \cap Y = \varnothing$ are both self-enforcing, then coalition $X \cup Y$ is not self-enforcing.

2. If X is a self-enforcing coalition, then $X \cup \{i\}$ for $i \notin X$ and $X \setminus \{i\}$ for $i \in X$ are not self-enforcing.

The most important implication is that, under majority rule $\alpha = 1/2$, the addition or the elimination of a single agent from a self-enforcing coalition makes this coalition no longer self-enforcing. This result motivates our interpretation in the Introduction of the power struggles in Soviet Russia following random deaths of Politburo members.

5.3 Size of Ruling Coalitions

Proposition 3 *Consider $\Gamma = (N, \gamma, w(\cdot), \alpha)$.*

1. Suppose $\alpha = 1/2$, then for any n and m such that $1 \leq m \leq n$, $m \neq 2$, there exists a set of players N, $|N| = n$, and a generic mapping of powers γ such that $|\phi(N)| = m$. In particular, for any $m \neq 2$ there exists a self-enforcing ruling coalition of size m. However, there is no self-enforcing coalition of size 2.

2. Suppose that $\alpha > 1/2$, then for any n and m such that $1 \leq m \leq n$, there exists a set of players N, $|N| = n$, and a generic mapping of powers γ such that $|\phi(N)| = m$.

These results show that one can say relatively little about the size and composition of URCs without specifying the power distribution within the society further (except that when $\alpha = 1/2$, coalitions of size 2 are not self-enforcing). However, this is largely due to the fact that there can be very unequal distributions of power. For the potentially more interesting case in which the distribution of power within the society is relatively equal, much more can be said about the size of ruling coalitions. In particular, the following proposition shows that, as long as larger coalitions have more power and there is majority rule ($\alpha = 1/2$), only coalitions of size $2^k - 1$ for some integer k (i.e., coalitions of size 3, 7, 15, etc.) can be the URC (Part 1). Part 2 of the proposition provides a sufficient condition for this premise (larger coalitions are more powerful) to hold. The rest of the proposition generalizes these results to societies with values of $\alpha > 1/2$.

Proposition 4 *Consider* $\Gamma = (N, \gamma, w(\cdot), \alpha)$ *with* $\alpha \in [1/2, 1)$.

1. Let $\alpha = 1/2$ and suppose that for any two coalitions $X, Y \in \mathcal{C}$ such that $|X| > |Y|$ we have $\gamma_X > \gamma_Y$ (i.e., larger coalitions have greater power). Then $\phi(N) = N$ if and only if $|N| = k_m$ where $k_m = 2^m - 1$, $m \in \mathbb{Z}$. Moreover, under these conditions, any ruling coalition must have size $k_m = 2^m - 1$ for some $m \in \mathbb{Z}$.

2. For the condition $\forall X, Y \in \mathcal{C} : |X| > |Y| \Rightarrow \gamma_X > \gamma_Y$ to hold, it is sufficient that there exists some $\lambda > 0$ such that

$$\sum_{j=1}^{|N|} \left| \frac{\gamma_j}{\lambda} - 1 \right| < 1. \tag{17}$$

3. Suppose $\alpha \in [1/2, 1)$ and suppose that γ is such that for any two coalitions $X \subset Y \subset N$ such that $|X| > \alpha |Y|$ ($|X| < \alpha |Y|$, resp.) we have $\gamma_X > \alpha \gamma_Y$ ($\gamma_X < \alpha \gamma_Y$, resp.). Then $\phi(N) = N$ if and only if $|N| = k_{m,\alpha}$ where $k_{1,\alpha} = 1$ and $k_{m,\alpha} = \lfloor k_{m-1,\alpha}/\alpha \rfloor + 1$ for $m > 1$, where $\lfloor z \rfloor$ denotes the integer part of z.

4. There exists $\delta > 0$ such that $\max_{i,j \in N} \{ \gamma_i / \gamma_j \} < 1 + \delta$ implies that $|X| > \alpha |Y|$ ($|X| < \alpha |Y|$, resp.) whenever $\gamma_X > \alpha \gamma_Y$ ($\gamma_X < \alpha \gamma_Y$, resp.). In particular, coalition $X \in \mathcal{C}$ is self-enforcing if and only if $|X| = k_{m,\alpha}$ for some m (where $k_{m,\alpha}$ is defined in Part 3).

This proposition shows that although it is impossible to make any general claims about the size of coalitions without restricting the distribution of power within the society, a tight characterization of the structure of the URC is possible when individuals are relatively similar in terms of their power.

5.4 Power and the Structure of Ruling Coalitions

One might expect that an increase in α—the supermajority requirement—cannot decrease the size of the URC. One might also expect that if an individual increases his power (either exoge-

nously or endogenously), this should also increase his payoff. However, both of these are generally not true. Consider the following simple example: let $w(\cdot)$ be given by (1). Then coalition $(3,4,5)$ is self-enforcing when $\alpha = 1/2$, but is not self-enforcing when $4/7 < \alpha < 7/12$, because $(3,4)$ is now a self-enforcing and winning subcoalition. Next, consider game Γ with $\alpha = 1/2$ and five players A, B, C, D, E with powers $\gamma_A = \gamma'_A = 2$, $\gamma_B = \gamma'_B = 10$, $\gamma_C = \gamma'_C = 15$, $\gamma_D = \gamma'_D = 20$, $\gamma_E = 21$, and $\gamma'_E = 40$. Then $\Phi(N, \gamma, w, \alpha) = \{A, D, E\}$, while $\Phi(N, \gamma', w, \alpha) = \{B, C, D\}$, so player E, who is the most powerful player in both cases, belongs to $\Phi(N, \gamma, w, \alpha)$ but not to $\Phi(N, \gamma', w, \alpha)$.

We summarize these results in the following proposition (proof omitted).

Proposition 5 *1. An increase in α may reduce the size of the ruling coalition. That is, there exists a society N, a power mapping γ and $\alpha, \alpha' \in [1/2, 1)$, such that $\alpha' > \alpha$ but for all $X \in \Phi(N, \gamma, w, \alpha)$ and $X' \in \Phi(N, \gamma, w, \alpha')$, $|X| > |X'|$ and $\gamma_X > \gamma_{X'}$.*

2. There exists a society N, $\alpha \in [1/2, 1)$, two mappings $\gamma, \gamma' : N \to \mathbb{R}_{++}$ satisfying $\gamma_i = \gamma'_i$ for all $i \neq j$, $\gamma_j < \gamma'_j$ such that $j \in \Phi(N, \gamma, w, \alpha)$, but $j \notin \Phi(N, \gamma', w, \alpha)$. Moreover, this result applies even when j is the most powerful player in both cases, i.e. $\gamma'_i = \gamma_i < \gamma_j < \gamma'_j$ for all $i \neq j$.

Intuitively, higher α turns certain coalitions that were otherwise non-self-enforcing into self-enforcing coalitions. But this implies that larger coalitions are now less likely to be self-enforcing and less likely to emerge as the ruling coalition. This, in turn, makes larger coalitions more stable. The first part of the proposition therefore establishes that greater power or "agreement" requirements in the form of supermajority rules do not necessarily lead to larger ruling coalitions. The second part implies that being more powerful may be a disadvantage, even for the most powerful player. This is for the intuitive reason that other players may prefer to be in a coalition with less powerful players so as to receive higher payoffs.

This latter result raises the question of when the most powerful player will be part of the ruling coalition. This question is addressed in the next proposition.

Proposition 6 *Consider the game $\Gamma(N, \gamma, w(\cdot), \alpha)$ with $\alpha \in [1/2, 1)$, and suppose that $\gamma_1, \ldots, \gamma_{|N|}$ is an increasing sequence. If $\gamma_{|N|} \in \left(\alpha \sum_{j=2}^{|N|-1} \gamma_j / (1-\alpha), \alpha \sum_{j=1}^{|N|-1} \gamma_j / (1-\alpha) \right)$, then either coalition N is self-enforcing or the most powerful individual, $|N|$, is not a part of the URC.*

6 Concluding Remarks

We presented an analysis of political coalition formation in nondemocratic societies. The absence of strong institutions regulating political decision-making in such societies implies that individuals competing for power cannot make binding promises (for example, they will be unable to commit to a certain distribution of resources in the future) and they will also be unable to commit to abide by the coalitions they have formed. This latter feature implies that once a particular ruling coalition has formed, a subcoalition can try to sideline some of the original members. These considerations imply that ruling coalitions in nondemocratic societies should be *self-enforcing*, in the sense that there should not exist a self-enforcing subcoalition that can sideline some of the members of this ruling coalition. This implies that coalition formation in such political games must be forward-looking; at each point, individuals have to anticipate how future coalitions will behave. Despite this forward-looking element, we showed that self-enforcing ruling coalitions can be determined in a relatively straightforward manner. In particular, we presented both an axiomatic analysis and a noncooperative game of coalition formation, and established that both approaches lead to the same set of self-enforcing ruling coalitions. Moreover, because such coalitions can be characterized recursively (by induction), it is possible to characterize the key properties of self-enforcing ruling coalitions in general societies.

Our main results show that such ruling coalitions always exist and that they are generically unique. Moreover, a coalition will be a self-enforcing ruling coalition if and only if it does not possess any subcoalition that is sufficiently powerful and self-enforcing. We also demonstrated that, although equilibrium ruling coalitions are robust to small perturbations, the elimination of a member of a self-enforcing coalition corresponds to a "large" shock and may change the nature of the ruling coalition dramatically. This result provides us with a possible interpretation for the large purges that took place in Stalin's Politburo following deaths of significant figures. Finally, we showed that although ruling coalitions can, in general, be of any size, once we restrict attention to societies where power is relatively equally distributed, we can make strong predictions on the size of ruling coalitions (for example, with majority rule, $\alpha = 1/2$, the ruling coalition must be of the size $2^k - 1$, where k is an integer).

Naturally, the result that the ultimate ruling coalition always exists and is genetically unique depends on some of our assumptions. In particular, the assumption that there is no commitment to future divisions of resources is crucial both for the uniqueness and the characterization results. Other assumptions can be generalized, however, without changing the major results in the paper. For instance, the payoff functions can be generalized so that individuals may sometimes wish to

be part of larger coalitions, without affecting our main results.

Other interesting areas of study in this context relate to some of the results presented in Section 5. For example, we saw that individuals with greater power may end up worse off. This suggests that individuals may voluntarily want to relinquish their power (for example, their guns) or they may wish to engage in some type of power exchange before the game is played. Some of these issues were discussed in the working paper version of our paper, Acemoglu, Egorov and Sonin (2006), and developing these themes in the context of more concrete problems appears to be a fruitful area for future research. The two most important challenges in future research are to extend these ideas to games that are played repeatedly and are subject to shocks, and to relax the assumption that individuals that are sidelined have no say in future decision-making. Relaxing the latter assumption is particularly important to be able to apply similar ideas to political decision-making in democratic societies.

Appendix

Proof of Theorem 1: Consider first the properties of the set $\mathcal{M}(X)$ in (6) and the mapping $\phi(X)$ in (5) (Step 1). We then prove that $\phi(X)$ satisfies Axioms 1–4 (Step 2) and that this is the unique mapping satisfying Axioms 1–4 (Step 3). Finally, we establish that when Assumption 2 holds, ϕ is single valued (Step 4). These four steps together prove both parts of the theorem.

Step 1: Note that at each step of the induction procedure, $\mathcal{M}(X)$ is well-defined because Z in (6) satisfies $|Z| < |X|$ and thus ϕ has already been defined for Z. The argmin set in (5) is also well defined, because it selects the minimum of a finite number of elements (this number is smaller than $2^{|X|}$; X is a subset of $\mathcal{I}$, which is finite). Non-emptiness follows, since the choice set includes X. This implies that this procedure uniquely defines some mapping ϕ (which is uniquely defined, but not necessarily single-valued).

Step 2: Take any $X \in \mathcal{C}$. Axiom 1 is satisfied, because either $\phi(X) = X$ (if $|X| = 1$) or is given by (5), so $\phi(X)$ contains only subsets of X such that $\phi(X) \neq \varnothing$. Furthermore, in both cases $\phi(X)$ contains only winning (within X) coalitions, and thus Axiom 2 is satisfied.

To verify that Axiom 3 is satisfied, take any $Y \in \phi(X)$. Either $Y = X$ or $Y \in \mathcal{M}(X)$. In the first case, $Y \in \phi(X) = \phi(Y)$, while in the latter, $Y \in \phi(Y)$ by (6).

Finally, Axiom 4 holds trivially when $|X| = 1$, since there is only one winning coalition. If $|X| > 1$, take $Y \in \phi(X)$ and $Z \subset X$, such that $Z \in \mathcal{W}_X$ and $Z \in \phi(Z)$. By construction of $\phi(X)$, we have that

$$Y \in \operatorname*{argmin}_{A \in \mathcal{M}(X) \cup \{X\}} \gamma_A.$$

Note also that $Z \in \mathcal{M}(X) \cup \{X\}$ from (6). Then, if

$$Z \notin \operatorname*{argmin}_{A \in \mathcal{M}(X) \cup \{X\}} \gamma_A,$$

we must have $\gamma_Z > \gamma_Y$, and vice versa, which completes the proof that Axiom 4 holds.

Step 3: We next prove that Axioms 1–4 define a unique mapping ϕ. Suppose that there exist ϕ and $\phi' \neq \phi$ that satisfy these axioms. Then, Axioms 1 and 2 imply that if $|X| = 1$, then $\phi(X) = \phi'(X) = X$; this is because $\phi(X) \neq \varnothing$ and $\phi(X) \subset \mathcal{C}_X$ and the same applies to $\phi'(X)$. Therefore, there must exist $k > 1$ such that for any A with $|A| < k$, we have $\phi(A) = \phi'(A)$, and there exists $X \in \mathcal{C}$, $|X| = k$, such that $\phi(X) \neq \phi'(X)$. Without loss of generality, suppose $Y \in \phi(X)$ and $Y \notin \phi'(X)$. Take any $Z \in \phi'(X)$ (such Z exists by Axiom 1 and $Z \neq Y$ by hypothesis). We will now derive a contradiction by showing that $Y \notin \phi(X)$.

We first prove that $\gamma_Z < \gamma_Y$. If $Y = X$, then $\gamma_Z < \gamma_Y$ follows immediately from the fact that $Z \neq Y$ and $Z \subset X$ (by Axiom 1). Now, consider the case $Y \neq X$, which implies

23

$|Y| < k$ (since $Y \subset X$). By Axioms 2 and 3, $Y \in \phi(X)$ implies that $Y \in \mathcal{W}_X$ and $Y \in \phi(Y)$; however, since $|Y| < k$, we have $\phi(Y) = \phi'(Y)$ (by the hypothesis that for any A with $|A| < k$, $\phi(A) = \phi'(A)$) and thus $Y \in \phi'(Y)$. Next, since $Z \in \phi'(X)$, $Y \in \mathcal{W}_X$, $Y \in \phi'(Y)$ and $Y \notin \phi'(X)$, Axiom 4 implies that $\gamma_Z < \gamma_Y$.

Note also that $Z \in \phi'(X)$ implies (from Axioms 2 and 3) that $Z \in \mathcal{W}_X$ and $Z \in \phi'(Z)$. Moreover, since $\gamma_Z < \gamma_Y$, $Z \neq X$ and therefore $|Z| < k$ (since $Z \subset X$). This again yields $Z \in \phi(Z)$ by hypothesis. Since $Y \in \phi(X)$, $Z \in \phi(Z)$, $Z \in \mathcal{W}_X$, $\gamma_Z < \gamma_Y$, Axiom 4 implies that $Z \in \phi(X)$. Since $Z \in \phi(X)$, $Y \in \phi(Y)$, $Y \in \mathcal{W}_X$, $\gamma_Z < \gamma_Y$, Axiom 4 implies that $Y \notin \phi(X)$, yielding a contradiction. This completes the proof that Axioms 1–4 define at most one mapping.

Step 4: Suppose Assumption 2 holds. If $|X| = 1$, then $\phi(X) = \{X\}$ and the conclusion follows. If $|X| > 1$, then $\phi(X)$ is given by (5); since under Assumption 2 there does not exist $Y, Z \in \mathcal{C}$ such that $\gamma_Y = \gamma_Z$,

$$\underset{A \in \mathcal{M}(X) \cup \{X\}}{\operatorname{argmin}} \gamma_A$$

must be a singleton. Consequently, for any $|X|$, $\phi(X)$ is a singleton and ϕ is single-valued. This completes the proof of Step 4. ∎

Proof of Lemma 1: This lemma is proved by induction on the maximum length of histories of Γ (the number of steps in subgame Γ_h).

Base. If Γ_h includes one last step only, this means that the current coalition is some X and the current step is voting by the last voter v is voting over the last agenda setter's proposal $P = X$. In this case, Γ implies that the URC must be $R = X = \psi_X(X) = F_X(\varnothing)$ and it does not depend on v's vote. Moreover, there are no more eliminations, hence each player i who was not eliminated before receives $w_i(R) - \varepsilon j_h$, which coincides with both (15) and (16).

Step. Suppose that the result is proven for all proper subgames of Γ_h. Consider two cases.

Case 1: The current step is voting. Then, proposal P will be accepted if and only if $V^+ \cup \{i \in \{v\} \cup V : i \text{ votes } \tilde{y} \text{ in } \sigma_K\} \in \mathcal{W}_X$ (recall the definitions of V, V^- and V^+ from the text as the set of future voters, the set of those that have voted $\tilde{n}$ and the set of those that have voted $\tilde{y}$ respectively). If P is accepted, the URC will be X if $P = X$, while if $P \neq X$, the game will have a transition to P, after which the URC will be $F_P(P) = \chi(P)$ by induction (recall that after transition the game proceeds to an agenda-setting step). In both cases, the URC $R = \psi_X(P)$. If $P = X$, player $i \in X$ gets $w_i(X) - \varepsilon j_h$, which equals (15). If $P \neq X$, player $i \in P$ receives $w_i(R) - \varepsilon \left(j_h + 1 + \mathbf{I}_{\{R \neq P\}} \mathbf{I}_{\{i \in R\}}\right)$, which in this case equals (15), while player $i \in X \setminus P$ obtains $w_i^- - \varepsilon j_h$, which again equals (15) because $i \notin \chi(P) \subset P$. On the other hand, if proposal P is rejected, then the game ends when the voting ends if $A = \varnothing$ (then $R = X = F_X(\varnothing) = F_X(A)$

24

and each player $i \in X$ gets $w_i(R) - \varepsilon j_h$, which equals (16)) or, if $A \neq \varnothing$, the game continues with some $b \in X$ as agenda-setter and the remaining set of agenda-setters being $B = A \setminus \{b\}$. In the latter case, we know by induction that $R = F_X(B \cup \{b\}) = F_X(A)$ will be the URC, and the payoff player $i \in X$ receives is given by (16).

Case 2: The current step is agenda-setting; suppose player a is to propose $P = F_X(a)$. Note first that such P satisfies $\psi_X(P) = P$. Indeed, this automatically holds if $P = X$, while if $P \neq X$, then $\psi_X(P) = \chi(P)$, but for $P = F_X(a) \neq X$ we must have $\chi(P) = P$ by (11), so $\psi_X(P) = P$. Consider two subpossibilities. *First,* suppose $P = F_X(A \cup \{a\})$. Then, as (14) suggests, each player $i \in \psi_X(P)$ will vote $\tilde{y}$. Note that $\psi_X(P)$ is necessarily winning within X: if $P = X$ it follows from $X = \psi_X(P) \in \mathcal{W}_X$, while if $P \neq X$, then, as we just showed, $\psi_X(P) = P = F_X(a)$ which is winning by (11). This means that if proposal $P = F_X(A \cup \{a\})$, it is accepted, and $R = P = F_X(A \cup \{a\})$ both in the case $P = X$ and $P \neq X$ (in the latter case, $R = \psi_X(P)$ by induction, and $\psi_X(P) = P$). Payoffs in this case are given by (16) because there are no more transitions if $P = X$ and exactly one more transition if $P \neq X$, and only players in P get the additional $-\varepsilon$. *Second,* consider the case $P \neq F_X(A \cup \{a\})$. Then $\gamma_{F_X(A)} \leq \gamma_{\psi_X(P)}$, for $\gamma_{F_X(A)} > \gamma_{\psi_X(P)}$ would imply that minimum in (12) for $Y = A \cup \{a\} \neq \varnothing$ is reached at $F_X(a) = \psi_X(P) = P$ and thus $P = F_X(A \cup \{a\})$ which leads to a contradiction. But then, as (14) suggests, all players in $F_X(A) \in \mathcal{W}_X$ will vote against proposal P, and thus P will be rejected. We know by induction that then $R = F_X(A)$ and the payoffs are given by (16). This completes the proof of Lemma 1. ∎

Proof of Theorem 2: Profile σ_K involves only pure strategies. Applying Lemma 1 to the first stage where $h = \varnothing$, we deduce that the URC under σ_K is $F_N(N) = \chi(N) = K$, and payoffs are given by (16) which equals (7) because $j_h = 0$ (there were no eliminations before and $N = X$, $K = R$). The theorem is therefore proved if we establish that profile σ_K is a SPE. To do this we show that there is no profitable one-shot deviation, which is sufficient since Γ is finite.

Suppose, to obtain a contradiction, that there is a one-shot profitable deviation after history h; since only one player moves at each history, this is either voter v or agenda-setter a. Let us start with the former case, which is then subdivided into two subcases.

Subcase 1: suppose voter v votes $\tilde{y}$ in σ_K (this means $v \in \psi_X(P) \subset P$), but would be better off if he voted $\tilde{n}$. In profile σ_K, the votes of players who vote after voter v (those in V) do not depend on the vote of player v. Hence, if proposal P is accepted in equilibrium, deviating to $\tilde{n}$ can result in rejection, but not vice versa. This deviation may only be profitable if voter v is pivotal, so we restrict attention to this case. From Lemma 1, the URC will be

$\psi_X(P)$ if proposal P is accepted and $F_X(A)$ if P is rejected; the number of transitions will be between 0 and 2 (including transition from X to P if $P \neq X$) in the first case and either 0 or 1 in the second case, so the number of transitions matters only if $w_v(\psi_X(P)) = w_v(F_X(A))$ (see (2)). Let us prove that $v \in F_X(A)$, $\gamma_{F_X(A)} \leq \gamma_{\psi_X(P)}$, and $P \neq F_X(A \cup \{a\})$. *First*, since deviation is profitable, $v \in F_X(A)$ (recall that $v \in \psi_X(P)$ simply because v votes $\tilde{y}$ in σ_K). *Second*, if instead $\gamma_{F_X(A)} > \gamma_{\psi_X(P)}$, this would imply $w_v(F_X(A)) < w_v(\psi_X(P))$ due to Assumption 1. *Third*, if instead $P = F_X(A \cup \{a\})$, then $\psi_X(P) = P$ (for in this case either $P = X$ or $P = F_X(i)$ for some $i \in X$; the first case is trivial while the second is considered in the proof of Lemma 1). But we just showed that either $\gamma_{F_X(A)} < \gamma_{\psi_X(P)}$ or $\gamma_{F_X(A)} = \gamma_{\psi_X(P)}$. In the first case, the minimum in (12) for $Y = A \cup \{a\}$ cannot be achieved at $\psi_X(P) = P$ because $n(F_X(A)) < n(\psi_X(P))$ and $F_X(A)$ is feasible in this minimization problem, so $P \neq F_X(A \cup \{a\})$ which is a contradiction. In the second case, $w_v(F_X(A)) = w_v(\psi_X(P))$, and since $F_X(A) \neq X$ if and only if $\psi_X(P) \neq X$ ($F_X(A) = X \neq \psi_X(P)$ would contradict $\gamma_{F_X(A)} = \gamma_{\psi_X(P)}$, and so would $F_X(A) \neq X = \psi_X(P)$), the number of additional transitions is the same. Hence, deviation to $\tilde{n}$ is not profitable because with or without this deviation player v would get $w_v(F_X(A)) - \varepsilon \left(j_h + \mathbf{I}_{\{F_X(A) \neq X\}} \right)$. This contradiction proves that $P \neq F_X(A \cup \{a\})$. This, together with $v \in F_X(A)$ and $\gamma_{F_X(A)} \leq \gamma_{\psi_X(P)}$ implies that voter v must vote $\tilde{n}$ in profile σ_K, which contradicts the assumption that he votes $\tilde{y}$.

Subcase 2: suppose that voter v votes $\tilde{n}$ in σ_K, but would be better off voting $\tilde{y}$. Again, this deviation does not change other voters' votes, it can only change the URC from $F_X(A)$ to $\psi_X(P)$ and is only profitable if v is pivotal. Consider two possible cases. If $v \notin \psi_X(P)$, voting $\tilde{y}$ gives v exactly $w_v^- - \varepsilon(j_h + 1)$ ($v \in P$, so v is part of transition from X to P, and Lemma 1 implies that transition from P to $\psi_X(P) \neq P$ will proceed in one step, so v will participate in exactly one more transition). Voting $\tilde{n}$ will then result in at most one additional transition, so v obtains a payoff no less than $w_v(F_X(A)) - \varepsilon(j_h + 1)$. This implies that the payoff of player v from voting $\tilde{n}$ is at least as large as his payoff from deviating to $\tilde{y}$, thus deviation is not profitable. On the other hand, if $v \in \psi_X(P)$, then, as implied by equation (14), $v \in F_X(A)$, $\gamma_{F_X(A)} \leq \gamma_{\psi_X(P)}$, and $P \neq F_X(A \cup \{a\})$. By Lemma 1, if v votes $\tilde{n}$, the URC is $F_X(A)$ and he receives payoff $w_v(F_X(A)) - \varepsilon \left(j_h + \mathbf{I}_{\{F_X(A) \neq X\}} \right)$; if he votes $\tilde{y}$, the URC is $\psi_X(P)$ and he receives $w_v(\psi_X(P)) - \varepsilon \left(j_h + \mathbf{I}_{\{P \neq X\}} \left(1 + \mathbf{I}_{\{\psi_X(P) \neq P\}} \right) \right)$. But $\gamma_{F_X(A)} \leq \gamma_{\psi_X(P)}$ implies $w_v(F_X(A)) \geq w_v(\psi_X(P))$, so the deviation could only be profitable for v if $\mathbf{I}_{\{F_X(A) \neq X\}} > \mathbf{I}_{\{P \neq X\}} \left(1 + \mathbf{I}_{\{\psi_X(P) \neq P\}} \right)$. This can only be true if $F_X(A) \neq X$ and $\psi_X(P) = P = X$. In this case, however, strict inequality $\gamma_{F_X(A)} < \gamma_{\psi_X(P)}$ holds, and therefore $w_v(F_X(A)) > w_v(\psi_X(P))$. Then (2) implies that deviation for v is not profitable. This completes the proof

that no one-shot deviation by a voter may be profitable.

The remaining case is where agenda-setter a has a best response $Q \neq P$, and $P = F_X(a)$ does not belong to the best response set. Consider two subcases. *Subcase 1:* suppose coalition P is accepted if proposed; in that case, Q cannot be rejected under profile σ_K. The reason is as follows: if Q were rejected, then the URC would be $F_X(A)$. If $i \in F_X(A)$, then coalition $F_X(A)$ is feasible in minimization problem (11), which means that $w_a(F_X(A)) \leq w_a(P)$. If this inequality is strict, so $w_a(F_X(A)) < w_a(P)$, then deviation to Q is not profitable; if $w_a(F_X(A)) = w_a(P)$, then the either $F_X(A) = P = X$ or neither $F_X(A)$ nor P equals X, but in both cases a participates in the same number (0 or 1, respectively) of extra transitions, so utility from proposing P and Q is the same and the deviation is not profitable either. If, however, $i \notin F_X(A)$, then proposing Q will give a payoff $w_a^- - \varepsilon j_h$ while proposing P will give at least $w_a(P) - \varepsilon(j_h + 1)$ (again, $\psi_X(P) = P$ for $P = F_X(a)$), so deviation is again not profitable. This proves that Q must be accepted, which immediately implies that $\psi_X(Q) \in \mathcal{W}_X$ (only members of $\psi_X(Q)$ vote for Q in σ_K, see (14)), and then $Q \in \mathcal{W}_X$ because $\psi_X(Q) \subset Q$. We next prove that $\psi_X(Q) = Q$. Suppose, to obtain a contradiction, that $\psi_X(Q) \neq Q$; this immediately implies $Q \neq X$ and thus $\psi_X(Q) = \chi(Q)$. If a proposed $\psi_X(Q)$ instead of Q, it would be accepted, too. Moreover, the fact that χ is a projection implies that $\psi_X(\psi_X(Q)) = \psi_X(\chi(Q)) = \chi(\chi(Q)) = \chi(Q) = \psi_X(Q)$. In addition, any player i who votes $\tilde{y}$ if Q is proposed is part of $\psi_X(Q)$, and therefore would participate in voting for $\psi_X(Q)$; moreover, he would vote $\tilde{y}$ under σ_K in that case, too, because $\psi_X(Q) \neq F_X(A \cup \{a\})$ implies $Q \neq F_X(A \cup \{a\})$ (otherwise Q would satisfy $\psi_X(Q) = Q$), and from (14) one can see that anyone who votes $\tilde{n}$ if $\psi_X(Q)$ is proposed would also vote $\tilde{n}$ if Q were proposed. Therefore, $\psi_X(Q)$ would be accepted if proposed, but proposing $\psi_X(Q)$ would result in only one transition while proposing Q would result in two. Agenda-setter a must be in $\psi_X(Q)$, so proposing $\psi_X(Q)$ is better than proposing Q, which contradicts the assumption that Q is a best response for a and establishes that $\psi_X(Q) = Q$. Finally, for a to propose Q, $a \in Q$ must hold. We have proved that coalition Q is feasible in minimization (11) for $i = a$, and $Q \neq P$ implies $n(P) < n(Q)$. But in that case either $\gamma_P < \gamma_Q$ (then a prefers having P instead of Q as the URC, even if it means an extra transition) or $\gamma_P = \gamma_Q$ (then a is indifferent, because the number of transitions is the same because both $P = \psi_X(P)$ and $Q = \psi_X(Q)$). These arguments together imply that deviation to Q is not profitable for a when P will be accepted under σ_K.

Subcase 2: suppose coalition P is rejected if proposed. Clearly, Q must be accepted, for otherwise the payoffs under the two proposals are identical and Q is not a profitable deviation. Since $P = \psi_X(P)$ is winning within X, but is not accepted, then, from (14), $\gamma_{F_X(A)} \leq \gamma_{\psi_X(P)}$

27

and $P \neq F_X (A \cup \{a\})$. As in the previous case we can show that $Q \in \mathcal{W}_X$, $\psi_X (Q) = Q$, and $a \in Q$. Since Q is accepted, (14) implies that either $\gamma_{F_X(A)} > \gamma_{\psi_X(Q)}$ or $Q = F_X (A \cup \{a\})$, for otherwise members of winning coalition $F_X (A)$ would vote against Q in σ_K and Q would be rejected. In both cases, $n(Q) < n(P)$ (in the first case because $\gamma_Q = \gamma_{\psi_X(Q)} < \gamma_{F_X(A)} \leq \gamma_{\psi_X(P)} = \gamma_P$, and in the second case because $P = F_X (a)$ is feasible in minimization (12) for $Y = A \cup \{a\}$, and $P \neq Q$). This means, however, that P cannot be the outcome in minimization (11) for $i = a$ because Q is also feasible ($Q \in \mathcal{W}_X$, $a \in Q$, and $\chi(Q) = Q$ because $\psi_X (Q) = Q$ and $Q \neq X$ where the latter follows from $n(Q) < n(P)$). This, however, contradicts that $P = F_X (a)$ by construction of profile σ_K in (14). Therefore, there is no profitable deviation at the agenda setting step either. This completes the proof of Theorem 2. ∎

Proof of Lemma 2: This proof also uses induction on the number of steps in Γ_h.

Base: If only one step remains, then the current ruling coalition is some X, and this step must be voting by the last voter v over proposal $P = X$ made by the last agenda-setter. Regardless of the vote (and therefore in either profile), coalition X will be the URC, and each player $i \in X$ will receive payoff $w_i (X) - \varepsilon j_h$; each players in $N \setminus X$ will receive the same payoff under both profiles, because the intermediate coalitions and the number of transitions each player faced is the same because histories up to the last step are identical.

Step. Take any history h and denote the first player to act in subgame Γ_h by b and the payoff to player i when b plays action ξ by $U_i (\xi)$. By induction this value is the same both if profile σ and σ_K is played thereafter. Consequently, if some action is optimal for player b if profile σ is played in subgames of Γ_h, the same is true if profile σ_K is played, and vice versa. Let ξ_K be the action played by b in profile σ_K and ξ_0 be an action played in profile σ with a non-zero probability. Then both ξ_K and ξ_0 must yield the same payoff for b because both are optimal when σ_K is played thereafter. Thus $U_b (\xi_K) = U_b (\xi_0)$.

It therefore suffices to show that both action ξ_0 followed by equilibrium play of profile σ_K and action ξ_K followed by equilibrium play of the same profile σ_K result in the same URC and the same payoff for all players $i \in N$ (then by induction, action ξ_0 followed by equilibrium play of profile σ will result in the same URC and payoffs). This is clearly true when $\xi_0 = \xi_K$. Now consider the case where $\xi_0 \neq \xi_K$. Both action ξ_K and action ξ_0, accompanied by equilibrium play of profile σ_K, will result in 0, 1, or 2 additional eliminations, as follows from Lemma 1: after the first elimination, if any, equilibrium play will have at most one more elimination. This, together with (2), implies that $|w_b (R_0) - w_b (R_K)| \leq 2\varepsilon$ and $w_b (R_0) = w_b (R_K)$, where R_0 and R_K are URCs if ξ_0 and ξ_K are played by b, respectively. There are two possibilities.

First, consider the case $w_b (R_0) = w_b (R_K) \neq w_b^-$, then $b \in R_0$, $b \in R_K$, hence $\gamma_{R_0} = \gamma_{R_K}$,

and by Assumption 2, $R_0 = R_K$, so that the URC is the same in both cases. The number of transitions is also the same (because b participates in all transitions and is indifferent between ξ_0 and ξ_K). If there are no more transitions, then each player $i \in X$ obtains utility $w_i(R_0) - \varepsilon j_h$ for both actions. If there is exactly one transition from X to R_0, then player $i \in R_0$ gets $w_i(R_0) - \varepsilon(j_h + 1)$ and player $i \in X \setminus R_0$ gets $w_i^- - \varepsilon j_h$. Consider the case where there are exactly two transitions both after ξ_0 and ξ_K. This cannot be the case if the first step of Γ_h is agenda-setting, for in that case Lemma 1 implies that if action ξ_K played under profile σ_K, the equilibrium play involves only one more transition. Then the first step of Γ_h is voting over some proposal P; moreover, both action ξ_0 and ξ_K will result in acceptance of this proposal, for a rejection, again by Lemma 1, would lead to only one extra transition. But in that case the two additional transitions are from X to P and from P to $\psi_X(P) \neq P$. This establishes that each player $i \in X$ receives the same payoff regardless of whether b plays action ξ_0 or ξ_K.

Second, consider the case $w_b(R_0) = w_b(R_K) = w_b^-$. Suppose first that b is agenda-setter; then action ξ_K corresponds to making proposal $P = F_X(b)$. Then, as implied by Lemma 1, the URC is $F_X(b)$ that b is part of (this happens if P is accepted) or $F_X(A)$ which b may or may not be part of (this happens if P is rejected); here A is the set of would-be agenda-setters. In the case under consideration $b \notin R_K$, hence $R_K = F_X(A)$ and $U_b(\xi_K) = w_b^- - \varepsilon j_h$. Action ξ_0 is the proposal of some coalition $Q \neq P$ such that $b \in Q$. If Q is accepted, but $b \notin R_0$, then b has an extra transition to Q but is eventually eliminated, so he receives $U_b(\xi_0) = w_b^- - \varepsilon(j_h + 1)$ and this contradicts $U_b(\xi_K) = U_b(\xi_0)$. Therefore, Q must be rejected, the URC must be $R_0 = F_X(A) = R_K$, and each player $i \in R_0$ will receive $w_i(R_0) - \varepsilon(j_h + 1)$ while $i \in X \setminus R_0$ gets $w_i(R_0) - \varepsilon j_h$ in the case of either action. Now suppose that b is voting over some proposal P, then $b \in P$. Then one of actions ξ_0 and ξ_K is $\tilde{y}$ and the other $\tilde{n}$. For the action to matter, proposal must be accepted if $\tilde{y}$ is played and rejected if $\tilde{n}$ is played (or vice versa, but this is impossible under σ_K). But recall that voter b is not a member of R_0 and R_K. Therefore, b votes $\tilde{y}$, he receives $w_i^- - \varepsilon(j_h + 1)$ (he does not participate in the second transition, which will happen under σ_K because $b \in P$ and $b \notin R_0$, $b \notin R_k$). On the other hand, if b votes $\tilde{n}$, then by Lemma 1 he receives $w_i^- - \varepsilon j_h$, because there is only one transition in which b is eliminated. But this means that $U_b(\tilde{y}) \neq U_b(\tilde{n})$, so $U_b(\xi_0) \neq U_b(\xi_K)$, which implies that b cannot be indifferent between the two actions ξ_0 and ξ_K, thus yielding a contradiction.

We have therefore proved that after either of the two actions ξ_0 and ξ_K is played, the URC is the same and each player $i \in X$ is indifferent. But any player $i \in N \setminus X$ is indifferent, too, because in this case the payoffs are entirely determined by history h. This completes the proof of the step of induction, and therefore of Lemma 2. ∎

Proof of Theorem 3: The proof follows immediately from the application of Lemma 2 to the entire game Γ, which is starting with history $h = \varnothing$. The lemma then implies that the URC in any SPE must coincide with that under the strategy profile σ_K, i.e. K, and payoffs must be given by (7), as implied by Lemma 2. ∎

Proof of Lemma 3: (**Part 1**) The set $\mathcal{A}(N)$ may be obtained from $\mathbb{R}_{++}^{|N|}$ by subtracting a finite number of hyperplanes given by equations $\gamma_X = \gamma_Y$ for all $X, Y \in P(N)$ such that $X \neq Y$ and by equations $\gamma_Y = \alpha\gamma_X$ for all $X, Y \in P(N)$ such that $X \subset Y$. These hyperplanes are closed sets (in the standard topology of $\mathbb{R}_{++}^{|N|}$), hence, a small perturbation of powers of a generic point preserves this property (genericity). This ensures that $\mathcal{A}(N)$ is an open set; it is dense because hyperplanes have dimension lower than $|N|$. The proofs for $\mathcal{S}(N)$ and $\mathcal{N}(N)$ are by induction. The base follows immediately since $\mathcal{S}(N) = \mathbb{R}_{++}$ and $\mathcal{N}(N) = \varnothing$ are open sets. Now suppose that we have proved this result for all $k < |N|$. For any distribution of powers $\{\gamma_i\}_{i \in N}$, N is self-enforcing if and only if there are no proper winning self-enforcing coalitions within N. Now take some small (in the sup-metric) perturbation of powers $\{\gamma_i'\}_{i \in N}$. If this perturbation is small, then the set of winning coalitions is the same, and, by induction, the set of proper self-enforcing coalitions is the same as well. Therefore, the perturbed coalition $\{\gamma_i'\}$ is self-enforcing if and only if the initial coalition with powers $\{\gamma_i\}$ is self-enforcing; which completes the induction step.

(**Part 2**) Take any connected component $A \subset \mathcal{A}(N)$. Both $\mathcal{S}(N) \cap A$ and $\mathcal{N}(N) \cap A$ are open in A in the topology induced by $\mathcal{A}(N)$ (and, in turn, by $\mathbb{R}_{++}^{|N|}$) by definition of induced topology. Also, $(\mathcal{S}(N) \cap A) \cap (\mathcal{N}(N) \cap A) = \varnothing$ and $(\mathcal{S}(N) \cap A) \cup (\mathcal{N}(N) \cap A) = A$, which, given that A is connected, implies that either $\mathcal{S}(N) \cap A$ or $\mathcal{N}(N) \cap A$ is empty. Hence, A lies either entirely within $\mathcal{S}(N)$ or $\mathcal{N}(N)$. This completes the proof. ∎

Proof of Proposition 1: The first two parts follow by induction. If $N = 1$, for any γ and α, $\Phi(N, \gamma, w, \alpha) = \{N\}$. Now suppose that this is true for all N with $|N| < n$; take any society N with $|N| = n$. We then use the inductive procedure for determining $\Phi(N, \gamma, w, \alpha)$, which is described in Theorem 1. In particular, Assumptions 2 and 3 imply that the set $\mathcal{M}(N)$ in (6) is identical for $\Gamma(N, \gamma, w, \alpha)$, $\Gamma(N, \gamma', w, \alpha)$, and $\Gamma(N, \gamma, w, \alpha')$, provided that δ is sufficiently small (the result self-enforcing coalitions remain self-enforcing after perturbation follows from Lemma 3). Moreover, if δ is small, then $\gamma_X > \gamma_Y$ is equivalent to $\gamma_X' > \gamma_Y'$. Therefore, (5) implies that $\Phi(N, \gamma, w, \alpha) = \Phi(N, \gamma', w, \alpha) = \Phi(N, \gamma, w, \alpha')$. This completes the proof of parts 1 and 2.

The proof of part 3 is also by induction. Let $|N_1| = n$. For $n = 1$ the result follows straightforwardly. Suppose next that the result is true for n. If δ is small enough, then $\phi(N_1)$ is

winning within $N = N_1 \cup N_2$; we also know that it is self-enforcing. Thus we only need to verify that there exists no $X \subset N_1 \cup N_2$ such that $\phi(X) = X$ (i.e., X that is self-enforcing, winning in $N_1 \cup N_2$ and has $\gamma_X < \gamma_{\phi(N_1)}$). Suppose, to obtain a contradiction, that this is not the case (i.e., that the minimal winning self-enforcing coalition $X \in P(N_1 \cup N_2)$ does not coincide with $\phi(N_1)$). Consider its part that lies within N_1, $X \cap N_1$. By definition, $\gamma_{N_1} \geq \gamma_{\phi(N_1)} > \gamma_X \geq \gamma_{X \cap N_1}$, where the strict inequality follows by hypothesis. This string of inequalities implies that $X \cap N_1$ is a proper subset of N_1, thus must have fewer elements than n. Then, by induction, for small enough δ, $\phi(X \cap N_1) = \phi(X) = X$ (since X is self-enforcing). However, $\phi(X \cap N_1) \subset N_1$, and thus $X \subset N_1$. Therefore, X is self-enforcing and winning within N_1 (since it is winning within $N_1 \cup N_2$). This implies that $\gamma_{\phi(N_1)} \leq \gamma_X$ (since $\phi(N_1)$ is the minimal self-enforcing coalition that is winning within N_1). But this contradicts the inequality $\gamma_{\phi(N_1)} > \gamma_X$ and implies that the hypothesis is true for $n + 1$. This completes the proof of part 3. ∎

Proof of Proposition 2: (Part 1) Either X is stronger than Y or vice versa. The stronger of the two is a winning self-enforcing coalition that is not equal to $X \cup Y$. Therefore, $X \cup Y$ is not the minimal winning self-enforcing coalition, and so it is not the URC in $X \cup Y$.

(Part 2) For the case of adding, it follows directly from Part 1, since coalition of one player is always self-enforcing. For the case of elimination: suppose that it is wrong, and the coalition is self-enforcing. Then, by Part 1, adding this person back will result in a non-self-enforcing coalition. This is a contradiction which completes the proof of Part 2. ∎

Proof of Proposition 3: (Part 1) Given Part 3 of Proposition 1, it is sufficient to show that there is a self-enforcing coalition M of size m (then adding $n - m$ players with negligible powers to form coalition N would yield $\phi(N) = \phi(M) = M$). Let $i \in M = \{1, \ldots, m\}$ be the set of players. If $m = 1$, the statement is trivial. Fix $m > 2$ and construct the following sequence recursively: $\gamma_1 = 2$, $\gamma_k > \sum_{j=1}^{k-1} \gamma_j$ for all $k = 2, 3, \ldots, m-1$, $\gamma_m = \sum_{j=1}^{m-1} \gamma_j - 1$. It is straightforward to check that numbers $\{\gamma_i\}_{i \in M}$ are generic. Let us check that no proper winning coalition within M is self-enforcing. Take any proper winning coalition X; it is straightforward to check that $|X| \geq 2$, for no single player forms a winning coalition. If coalition X includes player m (with power γ_m), then it excludes some player k with $k < m$; his power $\gamma_k \geq 2$ by construction. Hence,

$$\gamma_m = \sum_{j=1}^{m-1} \gamma_j - 1 > \sum_{j=1}^{m-1} \gamma_j - \gamma_k \geq \gamma_{X \setminus \{m\}},$$

which means that player m is stronger than the rest, and thus coalition M is non-self-enforcing. If X does not include γ_m, then take the strongest player in X; suppose it is k, $k \leq m - 1$. However, by construction he is stronger than all other players in X, and thus X is not self-

31

enforcing. This proves that M is self-enforcing. However, if $|X| = 2$ and Assumption 2 holds, then one of the players, say player i, is stronger than the other one, and thus $\{i\}$ is a winning self-enforcing coalition. But then, by Corollary 1, X cannot be self-enforcing.

(**Part 2**) The proof is identical to Part 1. The recursive sequence should be constructed as follows: $\gamma_1 = 2$, $\gamma_k > \alpha \sum_{j=1}^{k-1} \gamma_j$ for all $k = 2, 3, \ldots, m-1$, $\gamma_m = \alpha \sum_{j=1}^{m-1} \gamma_j - 1$. ∎

Proof of Proposition 4: (Part 1) This part follows as a special case of Part 3. To see this, note that the condition in Part 4 is satisfied , since for any $X \subset Y \subset N$, $|X| \gtrless \alpha |Y| \iff |X| \gtrless |Y \setminus X| \implies \gamma_X \gtrless \gamma_{Y \setminus X} \iff \gamma_X \gtrless \alpha \gamma_Y$ for $\alpha = 1/2$. Moreover, the sequences of k_m's in Part 1 and in Part 3 are equal since $k_1 = 2^1 - 1 = 1$, and if $k_{m-1} = 2^{m-1} - 1$ then $k_m = 2^m - 1 = \lfloor 2k_{m-1} \rfloor + 1$ and thus the desired result follows by induction.

(**Part 2**) Suppose, to obtain a contradiction, that the claim is false, i.e., that for some $X, Y \subset N$ such that $|X| > |Y|$ we have $\gamma_X \leq \gamma_Y$. Then the same inequalities hold for $X' = X \setminus (X \cap Y)$ and $Y' = Y \setminus (X \cap Y)$, which do not intersect, so that $\sum_{j \in X'} \gamma_j \leq \sum_{j \in Y'} \gamma_j$. This implies $\sum_{j \in X'} \gamma_j / \lambda \leq \sum_{j \in Y'} \gamma_j / \lambda$, and thus $\sum_{j \in X'} (\gamma_j / \lambda - 1) + |X'| \leq \sum_{j \in Y'} (\gamma_j / \lambda - 1) + |Y'|$. Rearranging, we have

$$1 \leq |X'| - |Y'| \leq \sum_{j \in Y'} \left(\frac{\gamma_j}{\lambda} - 1 \right) - \sum_{j \in X'} \left(\frac{\gamma_j}{\lambda} - 1 \right) \leq \sum_{j \in X' \cup Y'} \left| \frac{\gamma_j}{\lambda} - 1 \right|.$$

However, X' and Y' do not intersect, and therefore this violates (17). This contradiction completes the proof of Part 2.

(**Part 3**) The proof is by induction. The base is trivial: a one-player coalition is self-enforcing, and $|N| = k_1 = 1$. Now assume the claim has been proved for all $q < |N|$, let us prove it for $q = |N|$. If $|N| = k_m$ for some m, then any winning (within N) coalition X must have size at least $\alpha (\lfloor k_{m-1}/\alpha \rfloor + 1) > k_{m-1}$ (if it has smaller size then $\gamma_X < \alpha \gamma_N$). By induction, all such coalitions are not self-enforcing, and this means that the grand coalition is self-enforcing. If $|N| \neq k_m$ for any m, then take m such that $k_{m-1} < |N| < k_m$. Now take the coalition of the strongest k_{m-1} individuals. This coalition is self-enforcing by induction. It is also winning (this follows since $k_{m-1} \geq \alpha \lfloor k_{m-1}/\alpha \rfloor = \alpha (k_m - 1) \geq \alpha |N|$, which means that this coalition would have at least α share of power if all individuals had equal power, but since this is the strongest k_{m-1} individuals, the inequality will be strict). Therefore, there exists a self-enforcing winning coalition, different from the grand coalition. This implies that the grand coalition is non-self-enforcing, completing the proof.

(**Part 4**) This follows from Part 3 and Proposition 3. ∎

Proof of Proposition 6: Inequality $\gamma_{|N|} > \alpha \sum_{j=2}^{n-1} \gamma_j / (1 - \alpha)$ implies that any coalition that includes $|N|$, but excludes even the weakest player will not be self-enforcing. The inequality

$\gamma_{|N|} < \alpha \sum_{j=2}^{n-1} \gamma_j / (1 - \alpha)$ implies that player $|N|$ does not form a winning coalition by himself. Therefore, either N is self-enforcing or $\phi(N)$ does not include the strongest player. $\blacksquare$

References

Acemoglu, Daron, Georgy Egorov and Konstantin Sonin (2006) "Coalition Formation in Political Games" NBER Working Paper 12749.

Acemoglu, Daron and James Robinson (2006) *Economic Origins of Dictatorship and Democracy*, Cambridge University press, Cambridge.

Acemoglu, Daron, James A. Robinson and Thierry Verdier (2004) "Kleptocracy and Divide and Rule: A Theory of Personal Rule," *Journal of the European Economic Association*, 2, 162-192.

Ambrus, Attila (2006) "Coalitional Rationalizability" *Quarterly Journal of Economics*, 121, 903-29.

Banerjee, S., Hideo Konishi and Tayfun Sonmez (2001) "Core in a Simple Coalition Formation Game," *Social Choice and Welfare*, 18, 135-153.

Baron, David and John Ferejohn (1989) "Bargaining in Legislatures," *American Political Science Review* 83: 1181-1206.

Bernheim, Douglas, Bezalel Peleg, and Michael Whinston (1987) "Coalition-proof Nash equilibria: I Concepts," *Journal of Economic Theory,* 42(1):1–12.

Bloch, Francis (1996) "Sequential Formation of Coalitions with Fixed Payoff Division," *Games and Economic Behavior* 14, 90–123.

Bloch, Francis, Garance Genicot and Debraj Ray (2006) "Informal Insurance in Social Networks," New York University, mimeo.

Calvert, Randall and Nathan Dietz. 1996. "Legislative Coalitions in a Bargaining Model with Externalities." mimeo, University of Rochester.

Chwe, Michael S. Y. (1994) "Farsighted Coalitional Stability," *Journal of Economic Theory*, 63: 299-325.

Conquest, Robert (1968) *The Great Terror: Stalin's Purge of the Thirties,* Macmillan: NY.

Eguia, Jon (2006) "Voting Blocs, Coalitions and Parties," New York University, mimeo.

Evans, Richard J. (2006) *Third Reich in Power*, Penguin Books, New York.

Jackson, Matthew and Boaz Moselle (2002) "Coalition and Party Formation in a Legislative Voting Game," *Journal of Economic Theory* 103: 49-87.

Jehiel, Philipp and Benny Moldovanu (1999) "Resale Markets in the Assignment of Property Rights," *Review of Economic Studies* 66(4): 971-991.

Gomes, Armando and Philippe Jehiel (2005) "Dynamic Processes of Social and Economic Interactions: On the Persistence of Inefficiencies," *Journal of Political Economy*, 113(3), 626-667

Gorlizki, Yoram and Oleg Khlevniuk (2004). *Cold Peace: Stalin and the Ruling Circle, 1945-1953.* Oxford: Oxford University Press.

Greenberg, Joseph and Shlomo Weber (1993) "Stable Coalition Structures with a Unidimensional Set of Alternatives," *Journal of Economic Theory,* 60: 62-82.

Hart, Sergiu and Mordechai Kurz (1983) "Endogenous Formation of Coalitions," *Econometrica,* 52:1047-1064.

Khlevniuk, Oleg (2006) *Politburo*, Yale Univ. Press.

Konishi Hideo and Debraj Ray (2001) "Coalition Formation as a Dynamic Process," New York University, mimeo.

Le Breton, Michel, Ignacio Ortuno-Ortin, and Shlomo Weber (2008) "Gamson's Law and Hedonic Games," *Social Choice and Welfare*, 50, 57-67.

Marriotti, Marco (1997) "A Model of Agreements in Strategic Form Games" *Journal of Economic Theory*, 74, 196-217.

Maskin, Eric (2003) "Bargaining, Coalitions, and Externalities," Presidential Address to the Econometric Society, Princeton University, mimeo.

Moldovanu, Benny (1992) "Coalition-Proof Nash Equilibrium and the Core in Three-Player Games," *Games and Economic Behavior*, 9, 21-34.

Moldovanu, Benny and Eyal Winter (1995) "Order Independent Equilibria," *Games and Economic Behavior* 9(1):21-34.

Peleg, Bezalel (1980) "A Theory of Coalition Formation in Committees," *Journal of Mathematical Economics*, 7, 115-134.

Pepinsky, Thomas (2007) "Durable Authoritarianism as a Self-Enforcing Coalition," Yale University, mimeo.

Perry, Motty and Philip Reny (1994) "A Noncooperative View of Coalition Formation and the Core," *Econometrica* 62(4):795-817.

Powell, Robert (1999) *In the Shadow of Power.* Princeton: Princeton University Press.

Pycia, Marek (2006) "Bargaining and Coalition Formation," Pennsylvania State University, mimeo.

Ray, Debraj (1989) "Credible Coalitions and the Core" *International Journal of Game Theory*, 18(2): 185-87.

Ray, Debraj (2007) *A Game-Theoretic Perspective on Coalition Formation*, manuscript.

Ray, Debraj and Rajiv Vohra (1997) "Equilibrium Binding Agreements," *Journal of Economic Theory*, 73, 30-78

Ray, Debraj and Rajiv Vohra (1999) "A Theory of Endogenous Coalition Structures," *Games and Economic Behavior*, 26: 286-336.

Ray, Debraj and Rajiv Vohra (2001) "Coalitional Power and Public Goods," *Journal of Political Economy*, 109, 1355-1384.

Riker, William H. (1962) *The Theory of Political Coalitions*, Yale University Press, New Haven.

Seidmann, Daniel and Eyal Winter (1998) "Gradual Coalition Formation," *Review of Economic Studies*, 65, 793-815.

Skaperdas, Stergios (1998) "On the Formation of Alliances in Conflict and Contests", *Public Choice*, 96, 25-42.

Slantchev, Branislav (2005) "Territory and Commitment: The Concert of Europe as Self-Enforcing Equilibrium," *Security Studies*, Vol. 14-4, 565-606.

Tan, Guofu and Ruqu Wang (1999) "Endogenous Coalition Formation in Rivalry", Queen's University, mimeo.