NBER WORKING PAPER SERIES

COALITION FORMATION IN POLITICAL GAMES

Daron Acemoglu
Georgy Egorov
Konstantin Sonin

Working Paper 12749
http://www.nber.org/papers/w12749

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2006

We thank Attila Ambrus, Irina Khovanskaya, Victor Polterovich, Muhamet Yildiz and seminar participants at the Canadian Institute of Advanced Research, the New Economics School, and University of Pennsylvania PIER conference for useful comments. Acemoglu gratefully acknowledges financial support from the National Science Foundation. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Coalition Formation in Political Games
Daron Acemoglu, Georgy Egorov, and Konstantin Sonin
NBER Working Paper No. 12749
December 2006
JEL No. C71,D71,D74

ABSTRACT

We study the formation of a ruling coalition in political environments. 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 sub-coalitions should be able to secede and become the new ruling coalition. We first present an axiomatic approach that captures these notions and determines a (generically) unique ruling coalition. We then construct a simple dynamic game that encompasses these ideas and prove that the sequentially weakly dominant equilibria (and the Markovian trembling hand perfect equilibria) of this game coincide with the set of ruling coalitions of the axiomatic approach. We also show the equivalence of these notions to the core of a related non-transferable utility cooperative game. In all cases, the nature of the ruling coalition is determined by the power constraint, which requires that the ruling coalition be powerful enough, and by the enforcement constraint, which imposes that no sub-coalition of the ruling coalition that commands a majority is self-enforcing. The key insight that emerges from this characterization is that the coalition is made self-enforcing precisely by the failure of its winning sub-coalitions to be self-enforcing. This is most simply illustrated by the following simple finding: with a simple majority rule, while three-person (or larger) coalitions can be self-enforcing, two-person coalitions are generically not self-enforcing. Therefore, the reasoning in this paper suggests that three-person juntas or councils should be more common than two-person ones. In addition, we provide conditions under which the grand coalition will be the ruling coalition and conditions under which the most powerful individuals will not be included in the ruling coalition. We also use this framework to discuss endogenous party formation.

Daron Acemoglu
Department of Economics
MIT, E52-380B
50 Memorial Drive
Cambridge, MA 02142-1347
and NBER
daron@mit.edu

Georgy Egorov
Littauer Center
1805 Cambridge Street
Cambridge, MA 02138
gegorov@fas.harvard.edu

Konstantin Sonin
New Economic School
47 Nakhimovsky prosp.
Moscow, 117418 RUSSIA
ksonin@nes.ru

1 Introduction

Consider a society in which each individual possesses some amount of military power ("guns") and can form a "coalition" with other individuals to fight against the remaining individuals. A group (coalition) that has sufficient power becomes the "ruling coalition"; it determines the allocation of resources in the society, e.g., how a pie of size 1 will be distributed. A group with more guns can eliminate (win against) a group with less guns. However, once the elimination has taken place, a subgroup within the winning group can engage in further rounds of eliminations in order to reduce the size of the ruling coalition and receive more for each of its members. What types of ruling coalitions do we expect to form? Will there generally be multiple "equilibrium" or "stable" ruling coalitions? Will more powerful individuals necessarily obtain more of the resources?

Consider an alternative scenario in which a group of individuals needs to form a committee to reach a decision. A group that has a (power-weighted) majority can impose the selection of a particular committee, thus potentially influencing the collective choice. However, once the committee forms, a subgroup within the committee may sideline some of the members and impose its own rule. What types of committees do we expect to emerge?

The key feature in both of these examples is that the structure and nature of the ruling coalition depends not only the power of different groups (coalitions), but also whether, once formed, a particular group will be stable, i.e., *self-enforcing*. This tension between the power and the stability of a group is a common feature in many situations where coalitions have to form in order to make collective choices or determine the allocation of resources.

To investigate these questions systematically, we consider a society consisting of a finite number of individuals, each with an exogenously given level of political power.[1] A group's power is the sum of the power of its members. The society has a fixed resource, e.g., a pie of size 1. The distribution of this resource among the individuals is determined by a (self-enforcing) *ruling coalition*. For concreteness, let us suppose that a ruling coalition distributes this resource among its members according to their political power. A ruling coalition needs to satisfy two requirements. First, it needs to have total power more than $\alpha \in [1/2, 1)$ times the power of all the individuals in society, so that it is a *winning* coalition. This is captured by the *power constraint*. Second, it should have no subcoalition that would be willing to secede and become the new ruling coalition, so that it is self-enforcing or stable. This second requirement is captured by the *enforcement constraint*.

We show that a subcoalition will be self-enforcing if its own winning subcoalitions are not self-

[1] Throughout, we will work with a society consisting of individuals. Groups that have solved their internal collective action problem and have well-defined preferences can be considered as equivalent to individuals in this game.

enforcing. Intuitively, any subcoalition that is both winning and self-enforcing will secede from the original coalition and obtain more for its members. Subcoalitions that are not self-enforcing will prefer not to secede because some of their members will realize that they will be left out of the ultimate ruling coalition at the next round of secession (elimination).

One of the simple but interesting implications of these interactions is that under majority rule, ruling coalitions are generically (in a sense to be made precise below) not two-person coalitions, *duumvirates*, but can be three-person coalitions, *triumvirates*.[2]

Example 1 Consider two agents A and B with powers $\gamma_A > 0$ and $\gamma_B > 0$ and assume that the decision-making rule requires power-weighted majority (i.e., $\alpha = 1/2$). If $\gamma_A > \gamma_B$, then starting with a coalition of agents A and 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, such a coalition, a duumvirate, is 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), in turn, implies that (A, B, C) is itself self-enforcing. To see this, suppose, for example, that a subcoalition of (A, B, C), (A, B) considers seceding from the original coalition. 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. Consequently, a triumvirate can be self-enforcing and become the ruling coalition. This example also shows that contrary to approaches with unrestricted side-payments (e.g., Riker, 1962), the ruling coalition will not generally be the minimal winning coalition (which is (A, B) in this example).

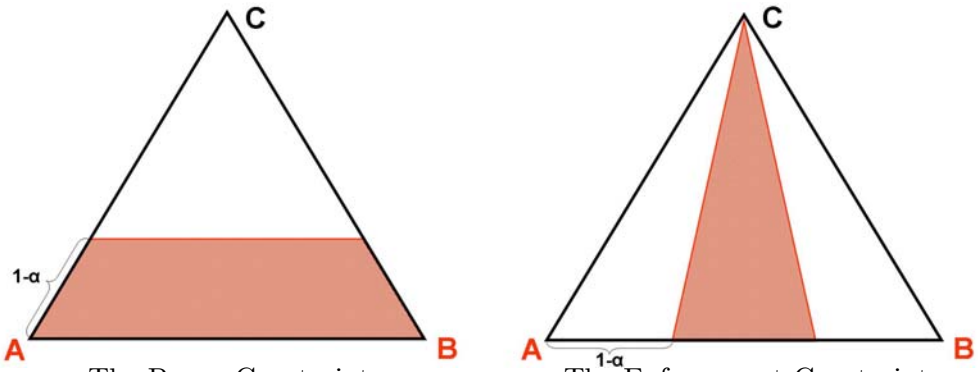
Next, consider a society consisting of four individuals. To illustrate the main ideas, suppose that we have $\gamma_A = 3, \gamma_B = 4, \gamma_C = 5$ as well as an additional individual, D, with power $\gamma_D = 10$. D's power is insufficient to eliminate the coalition (A, B, C) starting from the initial coalition

[2] Duumvirate and triumvirate are, respectively, the terms given to two-man and three-man executive bodies in Ancient Rome.

(A, B, C, D). 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 resist D's offer to secede and eliminate the third. 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, thus B, C and D will eliminate the weakest individual, A, and become the ruling coalition.

This example highlights the central roles of the power and the enforcement constraints, which are illustrated diagrammatically in Figure 1 for a society with three members, (A, B, C). The two dimensional simplex in the figure represents the powers of the three players (with their sum normalized to 1 without loss of generality). The shaded area in the first panel is the set of all coalitions where the subcoalition (A, B) is winning and shows the power constraint, which is parallel to the AB facet of the simplex.[3]



The Power Constraint The Enforcement Constraint

Figure 1

The enforcement constraint (for a subcoalition), on the other hand, defines the area, where, if other players are eliminated, the subcoalition still remains self-enforcing. The second panel depicts the enforcement constraint for the subcoalition (A, B) when player C is eliminated for a game with $\alpha > 1/2$. When $|N| = 3$, the enforcement constraint defines a cone.[4] In the case where $\alpha = 1/2$, this cone becomes a straight line perpendicular to the AB facet.

[3] More generally, if X is a coalition in a society N, then its power constraint corresponds to a hyperplane in a $(|N| - 1)$-dimensional simplex, which is parallel to a $(|X| - 1)$-dimensional facet that contains all vertices from X, and to a $(|N| - |X| - 1)$-dimensional facet containing all other vertices.

[4] More generally, for $N > 3$ and a proper coalition X, the enforcement constraint defines a quasi-cone, that is, a union of all segments that connect any point from the $(|N| - |X| - 1)$-dimensional facet containing all vertices from $N \setminus X$ to the set of self-enforcing points on the $(|X| - 1)$-dimensional facet containing all vertices from X.

Using this figure, we can determine the distributions of powers for which the subcoalition (A, B) can emerge as the ruling coalition within (A, B, C). First, it needs to be powerful enough, i.e., lie in the shaded area in the first panel. Second, it needs to be self-enforcing, i.e., lie in the cone of enforcement in the second panel. Clearly, when $\alpha = 1/2$, only a segment of the line where the powers of A and B are equal can satisfy these constraints, which captures the result in Example 1 that, generically, a two-person coalition cannot become a ruling coalition under majority rule. More generally, a coalition X can threaten the stability of the grand coalition only if both the power and the enforcement constraints are satisfied.

Our first major result is that an axiomatic approach to the determination of the ruling coalition using these two notions, together with two other technical axioms, is sufficient to single out a unique ruling coalition in generic games. We achieve this by defining a mapping from the set of coalitions of the society into itself that satisfies the above-mentioned axioms. When applied to the entire society, this mapping gives the *ruling coalition*.

That this axiomatic approach and the notion of ruling coalition capture important aspects of the process of coalition formation in political games is reinforced by our analysis of a simple dynamic game of coalition formation. In particular, we consider a dynamic game where at each stage a subset of the agents forms a coalition and "eliminates" those outside the coalition. The game ends when an ultimate ruling coalition, which does not wish to engage in further elimination, emerges. This ultimate ruling coalition divides the resources among its members according to their power. The important assumptions here are as follows. First, a player who is eliminated at any point cannot join future coalitions. This is in line with the motivating examples given above. Second, there is no possibility of commitment to the division of the resources once the ruling coalition is established. This no-commitment assumption is natural in political games, since it is impossible to make commitments or write contracts on future political decisions.[5] We establish the existence and generic uniqueness of sequentially weakly dominant and Markov trembling hand perfect equilibria of this dynamic game.[6] We also show that these equilibrium outcomes coincide with the ruling coalition derived from the axiomatic approach.

[5] See Acemoglu and Robinson (2006) for a discussion. Browne and Franklin (1973), Browne and Frendreis (1980), Schoffield and Laver (1985), and Warwick and Druckman (2001) provide empirical evidence consistent with the notion that ruling coalitions share resources according to the powers of their members. For example, these papers find a linear relationship between parties' shares of parliamentary seats (a proxy for their political power) and their shares of cabinet positions ("their share of the pie"). Ansolabehere et al. (2005) find a similar relationship between cabinet positions and voting weights (which are even more closely related to political power in our model) and note that: "The relationship is so strong and robust that some researchers call it 'Gamson's Law' (after Gamson, 1961, who was the first to predict such a relationship)."

[6] In fact, we establish the more general result that all *agenda-setting games* (as defined below) have a sequentially weakly dominant equilibrium and a Markov trembling hand perfect equilibrium.

Finally, we show that the same solution emerges when we model the process of coalition formation as a non-transferable utility cooperative game incorporating the notion that only self-enforcing coalitions can implement allocations that give high payoffs to their members.

At some level, it may not be surprising that all these approaches lead to the same result, since they capture the same salient features of the process of collective decision-making—the power and the enforcement constraints. Nevertheless, the three approaches model these features in very different ways. We therefore find it reassuring that they all lead to the same conclusions.

Our substantive results relate to the structure of ruling coalitions in this environment:

1. There always exists a ruling coalition and it can be computed by induction on the number of players.

2. Despite the simplicity of the environment, the ruling coalition 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.[7]

3. Relatedly, the most powerful individual will be excluded from the ruling coalition, unless he is powerful enough to win by himself or weak enough so as to be part of smaller self-enforcing coalitions.

4. Again somewhat paradoxically, an increase in α, i.e., an increase in the degree of supermajority, does not necessarily lead to larger ruling coalitions, 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.

5. 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 makes it non-self-enforcing.

6. Nevertheless, *ruling* coalitions are (generically) continuous in the distribution of power across individuals in the sense that a ruling coalition remains so when the powers of the players are perturbed by a small amount.

7. Coalitions of certain sizes are more likely to emerge as the ruling coalition. For example, with majority rule, the ruling coalition cannot (generically) consist of two individuals. Moreover,

[7] In terms of our first motivating example, this implies that individuals may wish to "give up their guns" voluntarily as a "commitment" not to overpower members of certain coalitions and thus be accepted as part of these coalitions.

again under majority rule, coalitions where members have relatively equal powers are self-enforcing only when the coalition's size is $2^k - 1$ where k is an integer.

Our paper is related to a number of different literatures. The first is the social choice literature. The difficulty of determining the social welfare function of a society highlighted by Arrow's (im)possibility theorem is related to the fact that the core of the game defined over the allocation of resources is empty (Arrow, 1951, Austen-Smith and Banks, 1999). Our approach therefore focuses on a weaker notion than the core, whereby only "self-enforcing" coalitions are allowed to form. Our paper therefore contributes to the collective choice literature by considering a different notion of aggregating individual preferences and establishes that such aggregation is possible.[8]

Our work is also related to models of bargaining over resources, both generally and in the context of political decision-making. In political economy (collective choice) context, two different approaches are worth noting. The first is given by the legislative bargaining models (e.g., Baron and Ferejohn, 1989, Calvert and Dietz, 1996, Jackson and Boaz, 2002), which characterize the outcomes of bargaining among a set of players by assuming specific game-forms approximating the legislative bargaining process in practice. Our approach differs from this strand of the literature, since we do not impose any specific bargaining structure. The second strand includes Shapley and Shubik (1954) on power struggles in committees and the paper by Aumann and Kurz (1977), which looks at the Shapley value of a bargaining game to determine the distribution of resources in the society.[9] Our approach is different since we focus on the endogenously-emerging ruling coalition rather than bargaining among the entire set of agents in a society or in an exogenously-formed committee.

At a more abstract level, our approach is a contribution to the literature on equilibrium coalition formation, which combines elements from both cooperative and noncooperative game theory (e.g., Peleg, 1980, Hart and Kurz, 1983, Aumann and Myerson, 1988, Greenberg and Weber, 1993, Chwe, 1994, Bloch, 1996, Ray and Vohra, 1997, 1999, 2001, Seidmann and Winter, 1998, Konishi and Ray, 2001, Maskin, 2003).[10] The most important difference between our approach and the

[8] 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 non-cooperative games, but impose that only stable coalitions can form. The main difference is that our basic approach is axiomatic or cooperative. We then provide a non-cooperative foundation for our approach, but do not explicitly use coalition-proofness as a refinement in the non-cooperative game. Another difference is that coalition-proof Nash equilibria typically fail to exist (though coalitionally rationalizable outcomes do exist, see Ambrus, 2006), while we show existence and generic uniqueness of our equilibrium concept in all three approaches we adopt.

[9] See also the literature on weighted majority cooperative games, which also model situations in which different individuals have different "weights" or "powers," e.g., Isbell (1956), Peleg (1968), Peleg and Rosenmuller (1992).

[10] Like some of these papers, our approach can be situated within the "Nash program" since our axiomatic approach

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.[11] 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 (see, for example, Ray and Vohra, 1999, and Maskin, 2003). A second important difference is that most of these works assume the possibility of binding commitments (see again Ray and Vohra, 1997, 1999), while we suppose that players have no commitment power. In addition, many previous approaches have proposed equilibrium concepts for cooperative games by restricting the set of coalitions that can block an allocation. Myerson (1991, ch. 9) and Osborne and Rubinstein (1994, ch. 14) give comprehensive discussions of many of these approaches. Our paper is also a contribution to this literature, since we propose a different axiomatic solution concept. To the best of our knowledge, neither the axiomatic approach nor the specific cooperative game form nor the dynamic game we analyze in this paper have been considered in the previous literatures on cooperative game theory or coalition formation.

The rest of the paper is organized as follows. Section 2 introduces the basic political game and contains a brief discussion of why it captures certain salient features of political decision-making. Section 3 provides our axiomatic treatment of this game. It introduces the concept of ruling coalition and proves its existence and generic uniqueness. Section 4 considers a dynamic game of coalition formation and a number of equilibrium concepts for this type of extensive-form games. It then establishes the equivalence between the ruling coalition of Section 3 and the equilibria of this extensive-form game. Section 5 introduces the cooperative game and establishes the equivalence between the core allocations of this game and the ruling coalitions. Section 6 contains our main results on the nature and structure of ruling coalitions in political games. Section 7 considers a number of extensions such as endogenous party formation and voluntary redistribution of power within a coalition. Section 8 concludes, and the Appendices contain the proofs not provided in the text as well as a number of examples to further motivate some of our equilibrium concepts.

is supported by an explicit extensive form game (Nash, 1953). See Serrano (2005) for a recent survey of work on the Nash program.

[11] This is a distinctive and general feature of political games. In presidential systems, the political contest is winner-take-all by design, while in parliamentary systems, parties left out of the governing coalition typically have limited say over political decisions. The same is a fortiori true in dictatorships.

2 The Political Game

We are interested in coalition-formation among a finite set of individuals. Let $\mathcal{I}$ be a collection of individuals. We refer to a *finite* subset N of $\mathcal{I}$ as a *society*.[12] The society has some resource to be distributed among these individuals. Each individual has strictly increasing preferences over his share of the resource and does not care about how the rest of the resource is distributed. The distribution of this scarce resource is the key political/collective decision. This abstract formulation is general enough to nest collective decisions over taxes, transfers, public goods or any other collective decisions.

Our focus is on how differences in the powers of individuals map into political decisions. We define a *power* mapping,

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

which determines the power of each individual in $\mathcal{I}$ (here $\mathbb{R}_{++} = \mathbb{R}_{+} \setminus \{0\}$). In particular, we refer to $\gamma_i = \gamma(i)$ as the political *power* of individual $i \in \mathcal{I}$. In addition, we denote the set of all possible power mappings by $\mathcal{I}^*$ and a power mapping γ restricted to some society $N \subset \mathcal{I}$ by $\gamma|_N$. For every set Y denote the set of its *non-empty* subsets by $P(Y)$. For any society $N \in P(\mathcal{I})$, any $X \in P(N)$ is called a *coalition* within N, or, for short, a coalition. Throughout, $|X|$ denotes the number of individuals in the set $X \in P(N)$, which is finite in view of the fact that a society N is always finite. The value

$$\gamma_X = \sum_{i \in X} \gamma_i \tag{1}$$

is called the *power* of coalition X (it is well-defined since X is finite). We define $\gamma_{\varnothing} = 0$.

We assume that collective decisions require a (super)majority in terms of power. In particular, let $\alpha \in [1/2, 1)$ be a number characterizing the degree of supermajority necessary for a coalition to implement any decision. The link between α and supermajority or majority rules is based on the following definition.

Definition 1 Suppose X and Y are coalitions such that $Y \in P(X)$. Coalition Y is *winning within* X if

$$\gamma_Y > \alpha \gamma_X.$$

Coalition $Y \subset N$ is called winning if it is winning within the society N.

[12] The reason for distinguishing between $\mathcal{I}$ and a society N is that, for some of our results, we imagine different subsets of a given collection $\mathcal{I}$ as distinct societies and make predictions for the ruling coalitions of these societies.

Clearly, $\gamma_Y > \alpha\gamma_X$ is equivalent to $\gamma_Y > \alpha\gamma_{X\setminus Y}/(1-\alpha)$. This illustrates that when $\alpha = 1/2$ a coalition Y to be winning within X needs to have a majority (within X) and when $\alpha > 1/2$, it needs to have a supermajority. Trivially, if Y_1 and Y_2 are winning within X, then $Y_1 \cap Y_2 \neq \varnothing$.

Given this description, for any society N we define *an abstract political game* as a triple $\Gamma = \Gamma(N, \gamma|_N, \alpha)$. We refer to Γ as an abstract game to distinguish it from the extensive-form and cooperative games to be introduced below. In particular, for game Γ, we do not specify a specific extensive form or strategic interactions, but proceed axiomatically.

We assume that in any political game, the decision regarding the division of the resource will be made by some ruling coalition. In particular, we postulate that the payoff to each player i is entirely determined by the ruling coalition $X \in P(N)$; we denote this payoff by $w_i(X)$. We impose the following natural restrictions on $w_i(X)$.

Payoffs The payoff function $w_i(X) : P(N) \to \mathbb{R}$ for any $i \in N$ satisfies the following:

1. $\forall i \in N$, $\forall X, Y \in P(N)$, if $i \in X$ and $i \notin Y$, then $w_i(X) > w_i(Y)$. [This means that each player prefers to be part of a ruling coalition.]
2. $\forall i \in N$, $\forall X, Y \in P(N)$, if $i \in X$ and $i \in Y$ then $w_i(X) > w_i(Y) \iff \gamma_i/\gamma_X > \gamma_i/\gamma_Y$ ($\iff \gamma_X < \gamma_Y$). [This means that each player prefers to be in a ruling coalition where his relative weight is higher.]
3. $\forall i \in N$, $\forall X, Y \in P(N)$, if $i \notin X$ and $i \notin Y$ then $w_i(X) = w_i(Y) = w_i^-$. [This means that each player is indifferent between two ruling coalitions which he is not part of, while the last equality defines w_i^- as the payoff of individual i when he is not a member of the ruling coalition for future reference.]

The restrictions on $w_i(X)$ are quite natural and capture the idea that the player's payoffs depend positively on the player's relative strength in the ruling coalition. The most important assumption introduced so far is that a coalition cannot commit to an arbitrary distribution of resources among its members. Instead, the allocation of resources is determined by the payoff functions $\{w_i(\cdot)\}_{i\in N}$. For example, a coalition consisting of two individuals with powers 1 and 10 cannot commit to share the resource equally if it becomes the ruling coalition. This assumption will play an important role in our analysis. We view this as the essence of political-economic decision-making: political decisions are made by whichever group has political power at the time, and ex ante commitments to future political decisions are generally not possible (see the discussion and references in footnote 5).

A specific example of payoff function $w_i(\cdot)$ that satisfies the requirements 1–3 results from the division of a fixed resource between members of the ruling coalition proportional to their power. In particular, for any $X \in P(N)$ let the share obtained by individual $i \in N$ of this fixed resource be

$$w_i(X) = \frac{\gamma_{X\cap\{i\}}}{\gamma_X} = \begin{cases} \frac{\gamma_i}{\gamma_X} & \text{if } i \in X \\ 0 & \text{if } i \notin X \end{cases} . \tag{2}$$

Evidently, for any $X \in P(N)$, $\sum_{i\in N} w_i(X) = 1$.

Since restrictions 1–3 define the division of the resource given the ruling coalition uniquely, the outcome of any game can be represented by its ruling coalition alone. More formally, fix a society $N \in P(\mathcal{I})$ and let $\mathcal{G}$ be the set of all possible games of the form $\Gamma = \Gamma(N, \gamma|_N, \alpha)$, i.e.,

$$\mathcal{G} = \{(N, \gamma|_N, \alpha) \mid N \in P(\mathcal{I}), \gamma \in \mathcal{I}^*, \alpha \in [1/2, 1)\}.$$

The outcome mapping, Φ, is a correspondence determining one or several subsets of $N \in P(\mathcal{I})$ as potential ruling coalitions for any game $\Gamma \in \mathcal{G}$, i.e.,

$$\Phi : \mathcal{G} \rightrightarrows P(\mathcal{I}).$$

Our main focus is to characterize the properties of this outcome correspondence. We wish to understand what types of ruling coalitions will emerge from different games, what the size of the ruling coalition will be, when it will include the more powerful agents, when it will be large relative to the size of the society (i.e., when it will be "dictatorial," and when it will be "inclusive").

We will sometimes impose the following assumption to obtain sharper results (though this assumption is not necessary for most of our results; see below).

Assumption 1 The power mapping $\gamma \in I^*$ is *generic* in the sense that for any $X, Y \in P(\mathcal{I})$, $\gamma_X = \gamma_Y$ implies $X = Y$. Similarly, we 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 this assumption is without much loss of generality since for any society N the set of vectors $\{\gamma_i\}_{i\in N} \in \mathbb{R}^{|N|}_{++}$ that fail to satisfy Assumption 1 is a set of Lebesgue measure 0 (in fact, it is a union of a finite number of hyperplanes in $\mathbb{R}^{|N|}_{++}$). For this reason, when a property holds under Assumption 1, we will say that it holds *generically*.

3 Axiomatic Analysis

We begin with an axiomatic analysis. Our approach is to impose some natural restrictions on the outcome mapping Φ that capture the salient features related to the formation of ruling coalitions. In particular, we would like the outcome mapping to determine the ruling coalitions for an arbitrary society. Such a ruling coalition must be winning (according to Definition 1) and self-enforcing, i.e. must be able to withstand challenges from its subcoalitions that satisfy the enforcement constraint (i.e., from subcoalitions that are self-enforcing).

Let us fix an abstract game $\Gamma(N, \gamma|_N, \alpha)$, that is, a power mapping $\gamma : \mathcal{I} \to \mathbb{R}_{++}$, a parameter $\alpha \in [1/2, 1)$, and a society $N \in P(\mathcal{I})$. Define the correspondence

$$\phi : P(\mathcal{I}) \rightrightarrows P(\mathcal{I})$$

by $\phi(N) = \Phi(\Gamma(N, \gamma|_N, \alpha))$ for any $N \in P(\mathcal{I})$. We will use ϕ instead of Φ whenever this causes no confusion.

Fix an abstract game $\Gamma = (N, \gamma, \alpha)$. Then in the spirit of the power and the enforcement constraints, we adopt the following axioms on ϕ.

Axiom 1 ***(Inclusion)*** For any $X \in P(N)$, $\phi(X) \neq \varnothing$ and if $Y \in \phi(X)$ then $Y \subset X$.

Axiom 2 ***(Power)*** For any $X \in P(N)$ and $Y \in \phi(X)$, $\gamma_Y > \alpha\gamma_X$.

Axiom 3 ***(Enforcement)*** For any $X \in P(N)$ and $Y \in \phi(X)$, $Y \in \phi(Y)$.

Axiom 4 ***(Group Rationality)*** For any $X \in P(N)$, for any $Y \in \phi(X)$ and any $Z \subset X$ such that $\gamma_Z > \alpha\gamma_X$ 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 2 Coalition $X \in P(\mathcal{I})$ is *self-enforcing* if $X \in \phi(X)$.

All four axioms are natural. Axiom 1, inclusion, implies that ϕ maps into subcoalitions of the coalition it operates upon and that it is defined, i.e., $\phi(X) \neq \varnothing$. (Note that this does not rule out $\phi(X) = \{\varnothing\}$, which is instead ruled out by Axiom 2.) Intuitively, this axiom implies that new players cannot join a coalition, which is the key feature of the political game we study in this paper. Axiom 2, the power axiom, requires that the winning coalition lies within X and has sufficient power according to Definition 1. Axiom 3, the enforcement axiom, simply requires that

any coalition $Y \in \phi(X)$ (for some $X \in P(N)$) should be self-enforcing according to Definition 2. Axiom 4 requires that if two coalitions $Y, Z \subset X$ are 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). This is an individual rationality or "group rationality" type axiom: since $\alpha \geq 1/2$, any two winning self-enforcing coalitions intersect, and we merely require players from such intersection to have rational preferences between these two coalitions (recall that $w_i(\cdot)$ is such that player i prefers a coalition where his relative power is greater).

The main result of the axiomatic analysis is the following theorem.

Theorem 1 *Fix a collection of players $\mathcal{I}$, a power mapping γ, and $\alpha \in [1/2, 1)$. Then:*

1. *There exists a unique mapping ϕ that satisfies Axioms 1–4. Moreover, when Assumption 1 holds, ϕ is single-valued.*

2. *This mapping ϕ may be obtained by the following inductive procedure:*

   For any $k \in \mathbb{N}$, let $P_k(\mathcal{I}) = \{X \in P(\mathcal{I}) : |X| = k\}$. Clearly, $P(\mathcal{I}) = \cup_{k \in \mathbb{N}} P_k(\mathcal{I})$. If $X \in P_1(\mathcal{I})$, then let $\phi(X) = \{X\}$. If $\phi(Z)$ has been defined for all $Z \in P_{k'}(\mathcal{I})$ and for all $k' < k$, then define $\phi(X)$ for $X \in P_k(\mathcal{I})$ as

$$\phi(X) = \underset{A \in \mathcal{M}(X) \cup \{X\}}{\operatorname{argmin}} \gamma_A, \tag{3}$$

   where

$$\mathcal{M}(X) = \{Z \in P(X) \setminus \{X\} : \gamma_Z > \alpha \gamma_X \text{ and } Z \in \phi(Z)\}. \tag{4}$$

   Proceeding inductively defines $\phi(X)$ for all $X \in P(\mathcal{I})$.

Proof. We begin with properties of the set $\mathcal{M}(X)$ defined in (4) and the mapping $\phi(X)$ defined in (3) (Step 1). We then prove that this mapping $\phi(X)$ satisfies Axioms 1–4 (Step 2). We next prove that this is the unique mapping satisfying Axioms 1–4 (Step 3). Finally, we establish that when 1 holds, ϕ is a 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 (4) satisfies $|Z| < |X|$ and thus ϕ has already been defined for Z. The argmin set in (3) is also well defined, because it selects the minimum of a finite number of elements (this number is less than $2^{|X|}$; X is a subset of N, 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 $X \in P(\mathcal{I})$. Axiom 1 is satisfied, because either $\phi(X) = \{X\}$ (if $|X| = 1$) or is given by (3), so $\phi(X)$ contains only subsets of X and $\phi(X) \neq \varnothing$. 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 (4).

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 $\gamma_Z > \alpha\gamma_X$ 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 (4). Then, since

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

we must have $\gamma_Z > \gamma_Y$, completing the proof that Axiom 4 holds.

Step 3: Now let us prove that Axioms 1–4 define a unique mapping ϕ. Suppose that there are two such mappings: ϕ and $\phi' \neq \phi$. Axioms 1 and 2 immediately imply that if $|X| = 1$, then $\phi(X) = \phi'(X) = \{X\}$. This is because Axiom 2 implies that $\phi(X) \neq \varnothing$ and Axiom 1 implies that $\phi(X)$ is non-empty; 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 $X \in P(\mathcal{I})$, $|X| = k$, 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 $|Y| < k$ (since $Y \subset X$). By Axioms 2 and 3, $Y \in \phi(X)$ implies that $\gamma_Y > \alpha\gamma_X$ 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 $Z \in \phi'(X)$, $\gamma_Y > \alpha\gamma_X$, $Y \in \phi'(Y)$ 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 $\gamma_Z > \alpha\gamma_X$ and $Z \in \phi'(Z)$. Moreover, since $\gamma_Z < \gamma_Y$, we have $Z \neq X$ and therefore $|Z| < k$ (since $Z \subset X$). This again yields $Z \in \phi(Z)$ by hypothesis. Since $Y \in \phi(X)$, $Z \in \phi(Z)$, $\gamma_Z > \alpha\gamma_X$, $\gamma_Z > \gamma_Y$, Axiom 4 implies that $Z \in \phi(X)$. Now, since $Z \in \phi(X)$, $Y \in \phi(Y)$, $\gamma_Y > \alpha\gamma_X$, $\gamma_Z < \gamma_Y$, 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 1 holds. If $|X| = 1$, then $\phi(X) = \{X\}$ and the conclusion follows. If $|X| > 1$, then $\phi(X)$ is given by (3); since under Assumption 1 there does not exist $Y, Z \in P(N)$ 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 and of Theorem 1. ■

At the first glance, Axioms 1–4 may appear relatively mild. Nevertheless, they are strong enough to pin down a unique mapping ϕ. This reflects the fact that the requirement of self-enforcement places considerable structure on the problem.

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

The fact that ϕ is determined uniquely implies the following corollary.

Corollary 1 *Take any collection of players $\mathcal{I}$, $\alpha \in [1/2, 1)$, and a power mapping $\gamma \in \mathcal{I}^*$. Let ϕ be the unique mapping satisfying Axioms* 1–4. *Coalition N is self-enforcing, that is, $N \in \phi(N)$, if and only if there exists no coalition $X \subset N$, $X \neq N$, which is winning within N and self-enforcing. Moreover, if N is self-enforcing, then $\phi(N) = \{N\}$.*

Proof. If $|N| = 1$, this trivially follows from the inductive procedure described in Theorem 1. If $|N| > 1$, $\phi(N)$ is given by (3). Note that for any $Y \in \mathcal{M}(N)$, we have $\gamma_Y < \gamma_N$. Therefore, $N \in \phi(N)$ if and only if $\mathcal{M}(N) = \varnothing$. The definition of $\mathcal{M}(N)$, (4), implies that $\mathcal{M}(N) = \varnothing$ if and only if there does not exist $X \subset N$, $X \neq N$, such that $\gamma_X > \alpha\gamma_N$ and $X \in \phi(X)$. It follows from (3) that in this case $\phi(N) = \{N\}$, proving both claims in Corollary 1. ■

While Theorem 1 proves the existence and uniqueness of a mapping ϕ satisfying Axioms 1–4 and describes the inductive procedure for constructing such mapping, Corollary 1 provides a recursive method of checking whether a particular coalition is self-enforcing. In particular, the result of Corollary 1 justifies our geometric representation in the Introduction: a coalition that includes a winning and self-enforcing subcoalition cannot be self-enforcing, and thus to determine the set of self-enforcing coalitions, we need to depict the power and enforcement constraints of all proper subcoalitions.

To illustrate the results of Theorem 1 and Corollary 1, let us now return to Example 1.

Example 1 (continued) Again, consider 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 1 is satisfied and $\phi(\{A, B, C\}) = \{A, B, C\}$.

Indeed, in this case, any two-person coalition is stronger than any single person and no coalitions containing the same number of members have equal power. Furthermore, under Assumption 1, $\phi(\{A,B\}) \neq \{A,B\}$, $\phi(\{A,C\}) \neq \{A,C\}$ and $\phi(\{B,C\}) \neq \{B,C\}$. Therefore, $\phi(\{A,B,C\})$ cannot be a doubleton, since there exists no two-person coalition X that can satisfy Axiom 3. Moreover, $\phi(\{A,B,C\})$ could not be a singleton, since, in view of the fact that $\gamma_A + \gamma_B > \gamma_C$, no singleton could satisfy Axiom 2. The only possible value for $\phi(\{A,B,C\})$ that does not violate Axioms 2 and 3 is $\{A,B,C\}$; it is straightforward to check that Axiom 4 is satisfied, too. We can also see that ϕ would not be single valued if Assumption 1 were not satisfied. Suppose, for example, that $\gamma_A = \gamma_B = \gamma_C$. In this case, $\phi(\{A,B,C\}) = \{\{A,B\},\{B,C\},\{A,C\}\}$.

Our next task is to characterize the mapping ϕ and determine the structure and properties of ruling coalitions. Before doing this, however, we will present a dynamic game and then a cooperative game, which will further justify our axiomatic approach.

4 A Dynamic Game of Coalition Formation

In this section, we introduce a dynamic game of coalition formation. We then discuss several equilibrium concepts for dynamic games of this kind, and show that for reasonable equilibrium concepts, when Assumption 1 holds there will exist a unique equilibrium coalition, coinciding with the ruling coalition defined in the previous section.

4.1 The Basic Game Form

Consider an abstract game $\Gamma(N, \gamma|_N, \alpha)$, that is, a society $N \in P(\mathcal{I})$ consisting of a finite number of individuals, a distribution of power $\{\gamma_i\}_{i\in N}$, and an institutional rule $\alpha \in [1/2, 1)$. We will now describe a related extensive-form game by $\hat{\Gamma} = \hat{\Gamma}(N, \gamma|_N, \alpha)$. This game $\hat{\Gamma}$ is different from Γ, since it refers to a particular extensive form game (described next).

Let $\{w_i(\cdot)\}_{i\in N}$ be as described in the previous section. Moreover, let $\varepsilon > 0$ be an arbitrarily small number such that for any $i \in N$ and any $X, Y \in P(N)$, if $w_i(X) > w_i(Y)$, then $w_i(X) > w_i(Y) + \varepsilon$. (Such $\varepsilon > 0$ exists, since N is a finite set.) In particular, for any $X \in P(N)$ such that $i \in X$, we have:

$$w_i(X) - w_i^- > \varepsilon. \tag{5}$$

Then the extensive form of the game $\hat{\Gamma}(N, \gamma|_N, \alpha)$ is as follows.

1. At each stage, $j = 0, 1, \ldots$, the game starts with an intermediary coalitions by $N_j \subset N$ (with $N_0 = N$).

2. Nature randomly picks agenda setter $i_{j,q} \in N_j$ for $q = 1$ (i.e., a member of the coalition N_j).

3. Agenda setter $i_{j,q}$ proposes a coalition $X_{j,q} \in P(N_j)$.

4. All players in $X_{j,q}$ vote over this proposal; let $v_{j,q}(i, X_{0,q}) \in \{\tilde{y}, \tilde{n}\}$ be the vote of player $i \in X_{j,q}$. Let $\text{Yes}\{X_{j,q}\}$ be the subset of $X_{j,q}$ voting in favor of this proposal, i.e.,

$$\text{Yes}\{X_{j,q}\} = \{i \in X_{j,q} : v_{j,q}(i, X_{0,q}) = \tilde{y}\}.$$

   Then, if

$$\sum_{i \in \text{Yes}\{X_{j,q}\}} \gamma_i > \alpha \sum_{i \in N_j} \gamma_i,$$

   i.e., if $X_{j,q}$ is winning within N_j (according to Definition 1), then we proceed to step 5; otherwise we proceed to step 6.

5. If $X_{j,q} = N_j$, then we proceed to step 7 and the game ends. Otherwise players from $N_j \setminus X_{j,q}$ are eliminated, players from $X_{j,q}$ add $-\varepsilon$ to their payoff, and the game proceeds to step 1 with $N_{j+1} = X_{j,q}$ (and j increases by 1).

6. If $q < |N_j|$, then next agenda setter $i_{j,q+1} \in N_j$ is randomly picked by nature such that $i_{j,q+1} \neq i_{j,r}$ for $1 \leq r \leq q$ (i.e., it is picked among those who have not made a proposal at stage j) and the game proceeds to step 3 (with q increased by 1). Otherwise, we proceed to step 7.

7. N_j becomes the ultimate ruling coalition (URC) of this terminal node, and each player $i \in N_j$ adds $w_i(N_j)$ to his payoff.

This game form implies that coalitions that emerge during the game form a sequence $N_0 \supset N_1 \supset \ldots \supset N_{\bar{j}}$ where $\bar{j}$ is the number of coalitions (excluding the initial one) that emerges during the game. Summing over the payoffs at each node, the payoff of each player i in game $\hat{\Gamma}$ is given by

$$U_i = w_i(N_{\bar{j}}) - \varepsilon \sum_{1 \leq j \leq \bar{j}} I_{N_j}(i), \tag{6}$$

where $I_X(\cdot)$ is the indicator (characteristic) function of set X. This payoff function captures the idea that individuals' overall utility in the game is related to their share w_i and to the number of rounds of elimination in which the individual is involved in (the second term in (6)). The number of players eliminated equals $|N| - |N_{\bar{j}}|$, and there is a total of $\bar{j}$ rounds of elimination.

With a slight abuse of terminology, we refer to j above as "the round of elimination." Without loss of generality, we assume that this is a game of perfect information, in particular after each time voting takes place each player's vote become common knowledge; this is convenient because the players would then share the same information sets.

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 (even temporary) coalition is formed. Its role for us is to rule out some "unreasonable" equilibria that arise in dynamic voting games. Example 5 in Appendix B illustrates that when $\varepsilon = 0$, there exist equilibria in which the outcome may depend on the behavior of players that will be eliminated for sure or on the order of moves chosen by Nature.

Note that $\hat{\Gamma}$ is a finite game; it ends after no more than $|N|(|N|+1)/2$ iterations because the size of a coalition as a function of the voting stage j defines a decreasing sequence bounded from below by 0. The coalition that forms after the last elimination is the URC. Consequently, the extensive-form game $\hat{\Gamma}$ necessarily has a subgame perfect Nash equilibrium (SPNE, see below). However, as the next example shows, there may be many SPNEs and some of those are highly unintuitive because they involve clear inefficiencies in voting (Appendix B presents a more striking example, Example 4, with multiple Markovian SPNEs).

Example 2 Consider $N = \{A, B, C, D\}$, with $\gamma_A = 3$,$\gamma_B = 4$, $\gamma_C = 5$ and $\gamma_D = 10$ (as in Example 1 above), and suppose that $\alpha = 1/2$. From Theorem 1, it can be seen that $\phi(\{A, B, C, D\}) = \{A, B, C\}$. As we will show later (Theorem 2), there is a Sequentially Weakly Dominant Equilibrium (which is also a SPNE) in which A, B, or C proposes this coalition, A, B and C vote for in favor of this coalition, and if D is the first to make a proposal, he proposes $\{A, B, C, D\}$ which is rejected by the rest. However, there exists another equilibrium, where all players make random proposals and everybody votes against any proposal. This is indeed an equilibrium, because no deviation by a single player can lead to acceptance of any proposal. Consequently, both $\{A, B, C\}$ and $\{A, B, C, D\}$ can emerge as subgame perfect equilibrium URCs, even though the former is not a reasonable outcome, since A, B, and C have enough votes to eliminate D.

In voting games, equilibria like the one involving $\{A, B, C, D\}$ as the URC in this example are typically eliminated by focusing on weakly dominant (or weakly undominated) strategies. However, this refinement loses its power in dynamic games, since even unreasonable actions are typically undominated because they may give relatively high payoffs when other players choose unreasonable actions in the future.

Because these standard equilibrium concepts do not give a satisfactory refinement, in the next section, we introduce the concept of *sequentially weakly dominant equilibrium*, which combines the ideas of backward induction and equilibrium in weakly dominant strategies.[13] We then demonstrate the existence of such equilibria for a broad class of agenda-setting games (defined below).

4.2 Sequentially Weakly Dominant Equilibria

In this subsection, we introduce the notion of *Sequential Weakly Dominant Equilibrium* (SWDE) inductively for finite extensive-form games. Consider a general n-person T-stage game, where each individual can take an action at every stage. Let the action profile of each individual be

$$a^i = \left(a_1^i, \ldots, a_T^i\right) \text{ for } i = 1, \ldots, n,$$

with $a_t^i \in A_t^i$ and $a^i \in A^i = \prod_{t=1}^{T} A_t^i$. Let $h_t = (a_1, \ldots, a_t)$ be the history of play up to stage t (not including stage t), where $a_s = \left(a_s^1, \ldots, a_s^n\right)$, so h_0 is the history at the beginning of the game, and let H_t be the set of histories h_t for $t : 0 \leq t \leq T-1$. We denote the set of all potential histories up to date t by $H^t = \bigcup_{s=0}^{t} H_s$. Let t-continuation action profiles be

$$a^{i,t} = \left(a_t^i, a_{t+1}^i, \ldots, a_T^i\right) \text{ for } i = 1, \ldots, n,$$

with the set of continuation action profiles for player i denoted by $A^{i,t}$. Symmetrically, define t-truncated action profiles as

$$a^{i,-t} = \left(a_1^i, a_2^i, \ldots, a_{t-1}^i\right) \text{ for } i = 1, \ldots, n,$$

with the set of t-truncated action profiles for player i denoted by $A^{i,-t}$. We also use the standard notation a^i and a^{-i} to denote the action profiles for player i and the action profiles of all other players (similarly, A^i and A^{-i}). The payoff functions for the players depend only on actions, i.e., player i's payoff is given by

$$u^i\left(a^1, \ldots, a^n\right).$$

We also define the restriction of the payoff function u^i to a continuation play $\left(a^{1,t}, \ldots, a^{n,t}\right)$ as $u^i\left(a^{1,-t}, \ldots, a^{n,-t} \ \vdots \ a_t^1, \ldots, a_t^n \ \vdots \ a^{1,t+1}, \ldots, a^{n,t+1}\right)$. In words, this function specifies the utility of player i if players played action profile $\left(a^{1,-t}, \ldots, a^{n,-t}\right)$ up to and including time $t-1$, played the action profile $\left(a_t^1, \ldots, a^n\right)$ at time t and are restricted to the action profile $a^{1,t}, \ldots, a^{n,t}$ from

[13] An alternative would have been to modify the game, so that all voting is sequential. It can be proved that the same results as here apply in this case *irrespective* of the order in which voting takes place. We prefer the notion of sequentially weakly dominant equilibrium, since it is more intuitive, creates greater continuity with static voting games and has applications in a broad class of agenda-setting games (see Appendix B).

t onwards. Symmetrically, this payoff function can also be read as the utility from continuation action profile $\left(a^{1,t+1},\ldots,a^{n,t+1}\right)$ given that up to time t, the play has consisted of the action profile $\left(a^{1,-t},\ldots,a^{n,-t}\ \vdots\ a_t^1,\ldots,a_t^n\right)$.

A (possibly mixed) strategy for player i is

$$\sigma^i : H^{T-1} \to \Delta\left(A^i\right),$$

where $\Delta(X)$ denotes the set of probability distributions defined over the set X, and for any $h \in H^T$ actions in the support of $\sigma^i(h)$ are feasible.

Denote the set of strategies for player i by Σ^i. A t-truncated strategy for player i (corresponding to strategy σ^i) specifies plays only until time t (including time t), i.e.,

$$\sigma^{i,-t} : H^{t-1} \to \Delta\left(A^{i,-t}\right).$$

The set of truncated strategies is denoted by $\Sigma^{i,-t}$. A t-continuation strategy for player i (corresponding to strategy σ^i) specifies plays only after time t (including time t), i.e.,

$$\sigma^{i,t} : H^{T-1} \setminus H^{t-2} \to \Delta\left(A^{i,t}\right),$$

where $H^{T-1} \setminus H^{t-2}$ is the set of histories starting at time t.

With a slight abuse of notation, we will also use the same utility function defined over strategies (as actions) and write

$$u^i\left(\sigma^{i,t},\sigma^{-i,t} \mid h^{t-1}\right)$$

to denote the continuation payoff to player i after history h^{t-1} when he uses the continuation strategy $\sigma^{i,t}$ and other players use $\sigma^{-i,t}$. We also use the notation $u^i\left(\sigma^{1,t},\ldots,\sigma^{n,t}\ \vdots\ \sigma^{1,t+1},\ldots,\sigma^{n,t+1} \mid h^{t-1}\right)$ as the payoff from strategy profile $\left(\sigma^{1,t},\ldots,\sigma^{n,t}\right)$ at time t restricted to the continuation strategy profile $\left(\sigma^{1,t+1},\ldots,\sigma^{n,t+1}\right)$ from $t+1$ onwards, given history h^{t-1}. Similarly, we use the notation

$$u^i\left(a^{i,t},a^{-i,t} \mid h^{t-1}\right)$$

for the payoff to player i when he chooses the continuation action profile $a^{i,t}$ and others choose $a^{-i,t}$ given history h^{t-1}. We start by providing the standard definitions of Nash equilibria and subgame perfect Nash equilibria.

Definition 3 A strategy profile $\left(\hat{\sigma}^1,\ldots,\hat{\sigma}^n\right)$ is a *Nash Equilibrium* if and only if

$$u^i\left(\hat{\sigma}^i,\hat{\sigma}^{-i}\right) \geq u^i\left(\sigma^i,\hat{\sigma}^{-i}\right) \text{ for all } \sigma^i \in \Sigma^i \text{ and for all } i = 1,\ldots,n.$$

Definition 4 A strategy profile $\left(\hat{\sigma}^1, ..., \hat{\sigma}^N\right)$ is a *Subgame Perfect Nash Equilibrium* if and only if

$$u^i\left(\hat{\sigma}^{i,t}, \hat{\sigma}^{-i,t} \mid h^{t-1}\right) \geq u^i\left(\sigma^{i,t}, \hat{\sigma}^{-i,t} \mid h^{t-1}\right) \text{ for all } h^{t-1} \in H^{t-1},$$
$$\text{for all } t, \text{ for all } \sigma^{-i} \in \Sigma^{-i} \text{ and for all } i = 1, \ldots, n.$$

Towards introducing weakly dominant strategies, let us take a small digression and consider a one stage game with actions $\left(a^1, ..., a^n\right)$.

Definition 5 We say that $\left(\hat{\sigma}^1, \ldots, \hat{\sigma}^n\right)$ is a *weakly dominant equilibrium* if

$$u^i\left(\hat{\sigma}^i, \sigma^{-i}\right) \geq u^i\left(\sigma^i, \sigma^{-i}\right) \text{ for all } \sigma^i \in \Sigma^i, \text{ for all } \sigma^{-i} \in \Sigma^{-i} \text{ and for all } i = 1, \ldots, n.$$

Naturally, such an equilibrium will often fail to exist. However, when it does exist, it is arguably a more compelling strategy profile than a strategy profile that is only a Nash equilibrium. Let us now return to the general T-stage game. A weakly dominant strategy equilibrium in this last stage of the game is defined similar to Definition 5.

Definition 6 Take any $t : 1 \leq t \leq T+1$. Strategy profile $\left(\hat{\sigma}^1, \ldots, \hat{\sigma}^n\right)$ is a h^{t-1}-*sequentially weakly dominant equilibrium* if $\left(\hat{\sigma}^1, \ldots, \hat{\sigma}^n\right)$ is a h^t-sequentially weakly dominant equilibrium for any h^t that may occur after h^{t-1} and

$$\begin{aligned} u^i\left(\hat{\sigma}^{i,t}, \sigma^{-i,t} \ \vdots \ \hat{\sigma}^{1,t+1}, \ldots, \hat{\sigma}^{N,t+1} \mid h^{t-1}\right) &\geq u^i\left(\sigma^{i,t}, \hat{\sigma}^{-i,t} \ \vdots \ \hat{\sigma}^{1,t+1}, \ldots, \hat{\sigma}^{N,t+1} \mid h^{t-1}\right) \\ \text{for all } \sigma^{i,t} &\in \Sigma^{i,t}, \text{ for all } \sigma^{-i,t} \in \Sigma^{-i,t}, \text{ and for all } i = 1, \ldots, N. \end{aligned}$$

Definition 6 is inductive, but this induction is finite. Indeed, if history h^{t-1} leads to a terminal node, then the first condition is satisfied automatically, because the history will not be recorded any further. Put differently, we first hypothesize that there exists a h^t-sequentially weakly dominant equilibrium, impose that it will be played from time $t+1$ onwards and then look for a weakly dominant strategy profile at stage t of the game.

Definition 7 Strategy profile $\left(\hat{\sigma}^1, \ldots, \hat{\sigma}^n\right)$ of a finite extensive-form game is a *Sequentially Weakly Dominant Equilibrium (SWDE)* if it is a h^0-sequentially weakly dominant equilibrium.

4.3 Characterization of Sequentially Weakly Dominant Equilibria

In this section, we characterize the SWDE of $\hat{\Gamma}$. Before doing this, recall that for any extensive-form game $\hat{\Gamma} = \hat{\Gamma}\left(N, \gamma|_N, \alpha\right)$, there is a corresponding abstract game $\Gamma = \Gamma\left(N, \gamma|_N, \alpha\right)$. Recall that $\mathcal{G}$ denotes the set of all such abstract games; we will interchangeably use it to denote the

set of all extensive form games as described in this section. Theorem 1 established that there is a unique mapping $\phi : P(\mathcal{I}) \rightrightarrows P(\mathcal{I})$ that satisfies Axioms 1–4 for any power mapping γ and any $\alpha \in [1/2, 1)$, and moreover, this mapping is single valued when Assumption 1 holds Now, for any extensive-form game $\hat{\Gamma} = \hat{\Gamma}(N, \gamma|_N, \alpha)$ and any equilibrium σ there is a probability distribution over terminal nodes (because players or Nature may randomize), which gives rise to a probability distribution of ultimate ruling coalitions (URC); denote the (essential) support of this probability distribution by $\Phi_\sigma\left(\hat{\Gamma}(N, \gamma|_N, \alpha)\right)$ or, for short, by $\phi_\sigma(N)$. Thus we use the notation $Y \in \phi_\sigma(N)$ to designate that coalition Y arises as URC with positive probability.

The main result in this section is that under Assumption 1 if σ is any SWDE, then $\phi_\sigma(N)$ is entirely concentrated in $\phi(N)$. If Assumption 1 does not hold, then for any ϕ satisfying Axioms 1–4, there exists an SWDE σ such that $\phi_\sigma(N)$ is entirely concentrated in $\phi(N)$.

The next theorem establishes both the existence of a pure strategy SWDE and the above equivalence result.

Theorem 2 *For an extensive-form game $\hat{\Gamma}(N, \gamma|_N, \alpha)$, denote the corresponding abstract game by $\Gamma(N, \gamma|_N, \alpha)$ and let $M \in \phi(N)$, where ϕ is the mapping defined in Theorem 1. Then:*

1. *There exists a pure strategy SWDE where the ultimate ruling coalition (URC) is M and is reached after at most one stage of elimination with probability* 1. *The payoff to each $i \in N$ in this SWDE is given by*

$$U_i(N) = w_i(X) - \varepsilon I_M(i) I_{\{M \neq N\}}. \tag{7}$$

2. *If Assumption 1 holds, then in* any *SWDE (in pure or mixed strategies) the only possible URC is M (i.e., $\Pr(\phi_\sigma(N) = \{M\}) = 1$), it is necessarily reached after at most one stage of elimination, and payoffs are given by* (7).

Proof. See Appendix C. ■

This theorem establishes two important results. First, a pure strategy SWDE exists for any game $\hat{\Gamma}(N, \gamma|_N, \alpha)$ and the URC is reached in the first stage of elimination (unless N is itself the URC). The existence of a pure strategy SWDE for this class of games is a noteworthy fact by itself, since SWDE is a demanding equilibrium concept and many games will not have such an equilibrium (see, e.g., Theorem 6). Second, the SWDE URC coincides with $\phi(N)$, that is, with the ruling coalition of $\Gamma(N, \gamma|_N, \alpha)$, which was derived axiomatically in Section 3. This equivalence allows us to use the inductive algorithm to determine the URC of the dynamic game. Finally, when Assumption 1 holds, the URC is the same in all SWDEs.

While the equivalence between pure strategy SWDE and our axiomatic approach is reassuring, in Appendix A, we introduce the concept of Markov Trembling-Hand Perfect Equilibrium (MTHPE), which is a slight refinement of the standard Trembling-Hand Perfect Equilibrium. We then prove that the same results apply with MTHPE as well. In this process, we also establish a number of more general results about the existence and structure of MTHPE in a broad class of agenda-setting games, which are of independent interest.

5 A Cooperative Game

In this section, we present a non-transferable utility cooperative game and establish that the (strong) core of this game coincides with the ruling coalition derived in Section 3.[14] This exercise is useful both because it links our equilibrium concept to those in the cooperative game theory literature and also because it provides another justification for our axiomatic approach. It is noteworthy that while the cooperative game theory approach is related to the axiomatic approach in Section 3, there are also crucial differences. First, the axiomatic approach involves no strategic interactions; in contrast (and similar to our noncooperative approach in Section 4), the emphasis here will be on strategic interactions among players and how there emerges a coalition structure implementing a payoff allocation such that no other coalition can implement an alternative feasible allocation making all of its members better off. Second, the cooperative approach incorporates the dynamic interactions that were also present in the previous section.

A non-transferable utility cooperative game is represented by $\tilde{\Gamma}_N = \tilde{\Gamma}\left(N, \gamma|_N, \alpha, v_N\left(\cdot\right)\right)$, where

$$v_N : P\left(N\right) \rightrightarrows \mathbb{R}_+^{|N|}$$

is a mapping from the set of coalitions to the set of allocations this coalition can enforce. Notice that the range of the mapping is not $\mathbb{R}_+^{|N|}$, but $P\left(\mathbb{R}_+^{|N|}\right)$, since typically a given coalition can enforce more than a single vector of utilities.

We define the mapping v_N inductively. To do this, we first introduce the concept of strong core (taking as given the definition of a feasible allocation). We then formally define the set of feasible allocations. Throughout this section, we assume, without loss of generality, that $w_i^- = 0$ for any $i \in N$ (this can always be achieved by a linear transformation).

We denote the core allocations for a non-transferable utility game $\tilde{\Gamma}_N = \tilde{\Gamma}\left(N, \gamma|_N, \alpha, v_N\left(\cdot\right)\right)$ by $C\left(\tilde{\Gamma}_N\right) \subset \mathbb{R}_+^{|N|}$. Moreover, for any vector $x \in \mathbb{R}_+^{|N|}$, we denote its ith component by x_i. Then:

[14] By "strong core" we refer to an allocation that cannot be strictly improved upon for all members of a blocking coalition; see Definition 8. To reduce terminology, refer to this as the "core."

Definition 8 A vector $x \in R_+^{|N|}$ is in the (strong) *core* for the game $\tilde{\Gamma}_N = (N, \gamma|_N, \alpha, v_N(\cdot))$, i.e., $x \in C\left(\tilde{\Gamma}_N\right)$, if and only if it is a feasible allocation and there exists no $Z \in P(N)$ for which there is $z \in v_N(Z)$ such that $z_i > x_i$ for all $i \in Z$.

Given this definition, we define feasible allocations as follows:

Definition 9 A vector $x \in R_+^{|N|}$ is a *feasible allocation* if either

1. $x_i = 0$ for all $i \in N$, or
2. $x_i = w_i(N)$ for all $i \in N$, or
3. there exists a subcoalition $Y \subset N$, $Y \neq N$, such that $x|_Y \in C\left(\tilde{\Gamma}_Y\right)$, where $\tilde{\Gamma}_Y = \tilde{\Gamma}(Y, \gamma|_Y, \alpha, v_Y(\cdot))$, while $x_i = 0$ for all $i \notin Y$.

This definition states that feasible allocations include those where all individuals receive zero payoff; those in which all individuals in the society share the resource according to their powers (which could be referred to as the "status quo" allocation); and those where a coalition Y distributes the resource among its members (according to the payoff functions $\{w_i(\cdot)\}_{i \in N}$ introduced above). In this latter case, however, not all coalitions are feasible. A coalition Y can only distribute the resource among its members if it is in the core of the same game restricted to a society identical to Y. Therefore, this definition is "recursive" in nature; it makes reference to core allocations, defined in Definition 8.

For a feasible allocation x, let $x^+ = \{i \in N : x_i > 0\}$, and $x^0 = \{i \in N : x_i = 0\}$. Then define the mapping v_N as follows:

$$v_N(X) = \begin{cases} \left\{x \in \mathbb{R}_+^{|N|} : x \text{ is feasible}\right\} & \text{if } \gamma_X > \alpha\gamma_N \\ \left\{x \in \mathbb{R}_+^{|N|} : x_i = 0 \ \forall\ i \in N\right\} \cup \\ \left\{x \in \mathbb{R}_+^{|N|} : x_i = w_i(N) \ \forall\ i \in N\right\} & \text{if } (1-\alpha)\gamma_N \leq \gamma_X \leq \alpha\gamma_N \\ \left\{x \in \mathbb{R}_+^{|N|} : x_i = 0 \ \forall\ i \in N\right\} & \text{if } \gamma_X < (1-\alpha)\gamma_N \end{cases} . \tag{8}$$

Notice that $v_N(X)$ is a subset of $\mathbb{R}_+^{|N|}$, meaning that coalition X may enforce multiple allocations. To motivate Definition 9 and the payoff vector in (8), we can reason as follows. Let us suppose that the allocation that gives $x_i = w_i(N)$ to each $i \in N$ is the "status quo" allocation. Then, an α-majority is needed to overrule this allocation, but $1-\alpha$ votes are sufficient to keep the status quo. That every coalition may enforce allocation $x_i = 0$ for all $i \in N$ is a requirement of the standard

definition of non-transferable utility cooperative games. Another requirement is that when $X \subset Y$, we should have $v_N(X) \subset v_N(Y)$. It is straightforward to check that (8) satisfies both requirements.

The most important feature of this non-transferable utility cooperative game is that even a winning coalition does not have complete discretion over the allocation; instead, it may only choose a core allocation for some smaller coalition Y, which is in some sense equivalent to excluding the players in $N \setminus Y$ from "sharing the pie" and making players in Y play the cooperative game $\tilde{\Gamma}(Y, \gamma|_Y, \alpha, v_Y(\cdot))$. If an allocation is not in the core of this reduced game, then it cannot be imposed, since coalition Y would rather choose a different allocation. This feature introduces the dynamic aspect mentioned above in the context of the cooperative game approach.

Example 3 To illustrate how this game works, consider a society consisting of three individuals A, B, C with powers $(\gamma_A, \gamma_B, \gamma_C) = (3, 4, 5)$, let $\alpha = 1/2$. Allocation $(3/12, 4/12, 5/12)$ is a feasible allocation; to see that it is in the core we need to check whether any coalition that has a majority of votes can improve for all its members. Take, for instance, coalition $\{B, C\}$; it is winning and therefore can implement any feasible allocation. Both B and C would be better off if they could give A nothing and share the resource between themselves. However, allocation $(0, 4/9, 5/9)$ is not feasible because the core of the game $\tilde{\Gamma}\left(\{B, C\}, \gamma|_{\{B,C\}}, \alpha, v_{\{B,C\}}(\cdot)\right)$ is a singleton with $x_B = 0$ and $x_C = 1$ (C constitutes a majority alone and can implement this allocation). Allocation $(0, 0, 1)$ is feasible and may be enforced by coalition $\{B, C\}$. But, naturally it makes player B worse off than he would have been with the allocation $(3/12, 4/12, 5/12)$. Likewise we can show that no other coalition can improve for all its members. Therefore, $(3/12, 4/12, 5/12)$ is in the core, and it is easy to verify that no other feasible vector lies in the core.

Our main result in this section is:

Theorem 3 *1. For any $\tilde{\Gamma}_N = \tilde{\Gamma}(N, \gamma|_N, \alpha, v_N(\cdot))$ with v_N defined by (8), the core $C\left(\tilde{\Gamma}_N\right)$ is nonempty.*

2. Take abstract game $\Gamma(N, \gamma|_N, \alpha)$ and the outcome set $\phi(N)$ corresponding to it (characterized in Theorem 1). Then, for any $x \in C\left(\tilde{\Gamma}_N\right)$, $x^+ \in \phi(N)$, and for $i \in x^+$, $x_i(i) = w_i(x^+)$; vice versa, for any $M \in \phi(N)$ there exists a unique $x \in C\left(\tilde{\Gamma}_N\right)$ such that $x^+ \in \phi(N)$ (so there is a one-to-one correspondence between core allocations and equilibrium ruling coalitions). Moreover, when Assumption 1 holds, the core $C\left(\tilde{\Gamma}_N\right)$ is a singleton and $x^+ = \phi(N)$, where x the unique element of the core.

Proof. See Appendix C. ■

This theorem therefore establishes the equivalence between the axiomatic approach in Section 3, the dynamic game in the previous section, and the cooperative game in this section, even though each of these three different approaches models the determination of the ruling coalition in a very different manner. The main idea underlying this result is that only self-enforcing and winning coalitions can implement payoff vectors that are attractive for their own members. In the framework of non-transferable utility cooperative games, this is captured by two features: first, only winning coalitions can implement feasible payoff vectors (Definition 8); second, in addition to zero payoffs and status quo allocation, only payoff vectors that correspond to core allocations for a smaller game are feasible (Definition 9). These two features introduce the power and enforcement constraints that were also essential in the axiomatic and the dynamic game approaches. In view of this, the finding that the set of core allocations here correspond to the ruling coalitions is perhaps not surprising, though still reassuring.

6 The Structure of Ruling Coalitions

In this section, we present several results on the structure of ruling coalitions. Given the equivalence results in the previous two sections, without loss of generality we focus on ruling coalitions of abstract games $\Gamma = \Gamma(N, \gamma|_N, \alpha)$. In addition to Assumption 1, consider:

Assumption 2 For no $X, Y \in P(\mathcal{I})$ such that $X \subset Y$ the equality $\gamma_Y = \alpha\gamma_X$ is satisfied.

Essentially, Assumption 2 guarantees that a small perturbation of a non-winning coalition Y does not make it winning. Similar to Assumption 1, this assumption fails only in a set of Lebesgue measure 0. Together, Assumptions 1 and 2 simplify the analysis in this section, and we assume that they both hold throughout the section, without explicitly stating this in every proposition.

6.1 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 will facilitate further proofs. Fix a society N and $\alpha \in [1/2, 1)$ and consider the set of power mappings restricted to society N, $\gamma|_N$. Clearly, each mapping 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 1 and 2 by $\mathfrak{A}(N)$, the subset of $\mathfrak{A}(N)$ for which $\Phi\left(N, \{\gamma_i\}_{i\in N}, \alpha\right) = N$ (i.e., the subset of power distributions for which coalition N is stable) by $\mathfrak{S}(N)$ and let $\mathfrak{N}(N) = \mathfrak{A}(N) \setminus \mathfrak{S}(N)$.

Theorem 4 *1. The set of power allocations that satisfy Assumptions 1 and 2, $\mathfrak{A}(N)$, its subsets for which coalition N is self-enforcing, $\mathfrak{S}(N)$, and its subsets for which coalition N is not self-enforcing, $\mathfrak{N}(N)$, are open sets in $\mathbb{R}_{++}^{|N|}$. The set $\mathfrak{A}(N)$ is also dense in $\mathbb{R}_{++}^{|N|}$.*

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

Proof. (**Part 1)** The set $\mathfrak{A}(N)$ 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 $\mathfrak{A}(N)$ is an open set; it is dense because hyperplanes have dimension lower than $|N|$. The proofs for $\mathfrak{S}(N)$ and $\mathfrak{N}(N)$ are by induction. The base follows immediately since $\mathfrak{S}(N) = \mathbb{R}_{++}$ and $\mathfrak{N}(N) = \varnothing$ 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 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 $A \subset \mathfrak{A}(N)$. Both $\mathfrak{S}(N) \cap A$ and $\mathfrak{N}(N) \cap A$ are open in A in the topology induced by $\mathfrak{A}(N)$ (and, in turn, by $\mathbb{R}_{++}^{|N|}$) by definition of induced topology. Also, $(\mathfrak{S}(N) \cap A) \cap (\mathfrak{N}(N) \cap A) = \varnothing$ and $(\mathfrak{S}(N) \cap A) \cup (\mathfrak{N}(N) \cap A) = A$, which, given that A is connected, implies that either $\mathfrak{S}(N) \cap A$ or $\mathfrak{N}(N) \cap A$ is empty. Hence, A lies either entirely within $\mathfrak{S}(N)$ or $\mathfrak{N}(N)$. This completes the proof. ■

An immediate corollary of Theorem 4 is that if the distribution of powers in two different games are "close," then these two games will have the same ruling coalition. To state and prove this proposition, endow the set of mappings γ, $\mathcal{I}^*$, with the sup-metric, with distance given by $\rho(\gamma, \gamma') = \sup_{i\in\mathcal{I}^*} |\gamma_i - \gamma_i'|$. Define a δ-neighborhood of γ as $\{\gamma' \in \mathcal{I}^* : \rho(\gamma, \gamma') < \delta\}$.

Proposition 1 *Fix a society N, $\alpha \in [1/2, 1)$ and a power mapping $\gamma : N \to \mathbb{R}_{++}$. Then:*

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

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

Proof. We prove this Proposition by induction. If $N = 1$, it is trivial: for any γ and α, $\Phi(N, \gamma, \alpha) = \{N\}$. Now assume that we have proved this result for all societies with $|N| < n$; take any society N with $|N| = n$. We take advantage of the inductive procedure for determining $\Phi(N, \gamma, \alpha)$, which is described in Theorem 1. Indeed, since we assume that Assumptions 1 and 2 hold, then the set $\mathcal{M}(N)$, as defined by (4), is the same for $\Gamma(N, \gamma, \alpha)$, $\Gamma(N, \gamma', \alpha)$, and $\Gamma(N, \gamma, \alpha')$, provided that δ is sufficiently small (we use induction to get that self-enforcing coalitions remain self-enforcing after perturbation). Moreover, if δ is small, then $\gamma_X > \gamma_Y$ is equivalent to $\gamma'_X > \gamma'_Y$. Therefore, (3) implies that $\Phi(N, \gamma, \alpha) = \Phi(N, \gamma', \alpha) = \Phi(N, \gamma, \alpha')$. This completes the proof. ■

In addition, we prove that the mapping ϕ is "robust" in the sense that the inclusion of sufficiently weak player(s) does not change the ruling coalition in a society.

Proposition 2 *Consider a game $\Gamma = \Gamma(N \cup M, \gamma|_{N \cup M}, \alpha)$ with arbitrary disjoint finite sets M and N. Then exists $\delta > 0$ such that for all M such that $\gamma_M < \delta$, $\phi(N) = \phi(N \cup M)$.*

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

Essentially, Proposition 2 says that mapping ϕ is "continuous at zero", in the sense that adding a group of agents with limited powers to the society (which is equivalent to changing their power from 0 to small positive values) does not change the ruling coalition. Indeed, while we did not define ϕ for societies where some of the members have 0 power, it is natural to think that they do not have an impact in votings or in resource allocation. Here, we prove that agents (or groups of agents) with sufficiently small power have no impact either.

6.2 Size of Ruling Coalitions

The next question we address is how many players may be included in the ruling coalition. We start with the case of majority rule and then consider supermajority rules. These result implies that relatively little can be said about the structure of ruling coalitions without putting some more structure.

Proposition 3 *If $\alpha = 1/2$, the following statements are true.*

1. *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 $\gamma|_N$ such that $|\phi(N)| = m$. In particular, for any $m \neq 2$ there exists a self-enforcing coalition of size m.*

2. *There is no self-enforcing coalition of size 2.*

Proof. (**Part 1)** Given Proposition 2, 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. Coalition X either includes γ_m or not. If it includes γ_m and is not proper, 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 γ_m is stronger than the rest, and thus coalition M is non-self-enforcing. If it 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-enforcing. This proves that M is self-enforcing.

(**Part 2)** If $|X| = 2$ and Assumption 1 holds, then one of the players (denote him 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. ■

The first part of Proposition 3 may be generalized for $\alpha > 1/2$. Moreover, in that case, any size (including 2) of self-enforcing coalitions is possible.

Proposition 4 *Take any collection of players $\mathcal{I}$, any power mapping $\gamma : \mathcal{I} \to \mathbb{R}_{++}$, and suppose that $\alpha > 1/2$. Then for any n and any m such that $1 \leq m \leq n$ there exists a set of players N, $|N| = n$, with powers $\gamma|_N$ such that $|\phi(N)| = m$. In particular, there exists a self-enforcing coalition of size m.*

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

These results show that one can say relatively little about the size and composition of the equilibrium ruling coalition without taking the specifics of the distribution of powers among the individuals into consideration. For the case when power distribution is nearly uniform, we have the following sequence of results.

Proposition 5 *Fix a society N, $\alpha \in [1/2, 1)$ and a power mapping $\gamma : N \to \mathbb{R}_{++}$. Then:*

1. *Let $\alpha = 1/2$ and suppose that for any two coalitions $X, Y \in P(N)$ 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 $k_m = 2^m - 1$, $m \in \mathbb{Z}$.*

2. *Let $\alpha = 1/2$. Consider set of players $\mathcal{I} = \mathbb{N}$ and define*

$$\gamma_n = \gamma(n) = \frac{2^n - 1}{2^n} \tag{9}$$

for each $n \in \mathbb{N}$. Let $N_n = \{1, \ldots, n\} \in P(\mathcal{I})$. Then in the game $\hat{\Gamma}(N_n, \gamma|_{N_n}, \alpha)$ the equilibrium ruling coalition has size m (i.e. $|\phi(N_k)| = m$) where $m = \max\{z \in N_n \mid z = 2^k - 1 \text{ for } k \in \mathbb{N}\}$.

3. *For the condition $\forall X, Y \in P(N) : |X| > |Y| \Rightarrow \gamma_X > \gamma_Y$ to hold, it is sufficient to require that there exists some $\lambda > 0$ such that*

$$\sum_{j=1}^{|N|} \left| \frac{\gamma_j}{\lambda} - 1 \right| < 1. \tag{10}$$

4. *More generally, 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 $k_{m,\alpha} = \left[\frac{k_{m-1,\alpha}}{\alpha}\right] + 1$ for $m > 1$, where $[z]$ denotes the integer part of z*

5. *For any mapping $\gamma|_N$ that satisfies Assumptions 1 and 2 there exists $\delta > 0$ such that $\max_{i,j \in N} \frac{\gamma_i}{\gamma_j} < 1 + \delta$ implies that $|X| > \alpha |Y|$ ($|X| < \alpha |Y|$, resp.) whenever $\gamma_X > \alpha \gamma_Y$*

($\gamma_X < \alpha\gamma_Y$, resp.). In particular, coalition $X \in P(N)$ is self-enforcing if and only if $|X| = k_{m,\alpha}$ for some m (where $k_{m,\alpha}$ is defined in Part 3).

Proof. (Part 1) Let us check that the condition in Part 4 is satisfied. Take any $X \subset Y \subset N$. Obviously, $|X| \gtrless \frac{1}{2}|Y| \iff |X| \gtrless |Y \setminus X| \implies \gamma_X \gtrless \gamma_{Y\setminus X} \iff \gamma_X \gtrless \frac{1}{2}\gamma_Y$. Now let us check that k_m's in Part 1 and in Part 4 are equal. Indeed, $k_1 = 2^1 - 1 = 1$ and if $k_{m-1} = 2^{m-1} - 1$ then $k_m = 2^m - 1 = [2k_{m-1}] + 1$. By induction, we get that Part 1 follows as a special case of Part 4.

(Part 2) The statement follows immediately from Part 1.

(Part 3) Assume the contrary, 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. Mathematically,

$$\sum_{j\in X'} \gamma_j \leq \sum_{j\in Y'} \gamma_j.$$

This implies

$$\sum_{j\in X'} \frac{\gamma_j}{\lambda} \leq \sum_{j\in Y'} \frac{\gamma_j}{\lambda}$$

and thus

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

Rearranging, we have

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

However, X' and Y' do not intersect, and therefore this violates (10). This contradiction completes the proof of Part 3.

(Part 4) 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 $\alpha\left(\left[\frac{k_{m-1}}{\alpha}\right] + 1\right) > \alpha\frac{k_{m-1}}{\alpha} = k_{m-1}$ (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 $k_{m-1} = \alpha\frac{k_{m-1}}{\alpha} \geq \alpha\left[\frac{k_{m-1}}{\alpha}\right] = \alpha(k_m - 1) \geq \alpha|N|$, 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 not self-enforcing, completing the proof.

(Part 5) This follows immediately from Part 4 and Proposition 4. ■

Consequently, while it is impossible to make any general claims about the size of coalitions without specifying more details about the distribution of power within the society, we are able to provide a tight characterization of the structure of the ruling coalition when individuals are relatively similar in terms of their power.

6.3 Fragility of Self-Enforcing Coalitions

Here, we show that while 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, such as adding or losing a member of the ruling coalition. The next proposition establishes that under simple majority rule, self-enforcing coalitions are fragile in the sense that addition or subtraction of a single agent from these coalitions or, more generally, a union of two disjoint self-enforcing coalitions leads to a non-self enforcing coalition.

Proposition 6 *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.*

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.*

Proof. (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 $X \cup Y$. This implies that $X \cup Y$ is not the minimal winning self-enforcing coalition, and so it is not the ruling coalition 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 deleting: suppose that it is wrong, and the coalition is self-enforcing. Then, by Part 1, adding this person back will result in an non-self-enforcing coalition. This is a contradiction which completes the proof of Part 2. ■

6.4 Power and the Structure of Ruling Coalitions

One might expect that an increase in α—the supermajority requirement—may increase the size of the ruling coalition and also turn otherwise non-self-enforcing coalitions into self-enforcing ones

(e.g., coalition $(3,4)$ is not self-enforcing when $\alpha = 1/2$, but becomes self-enforcing when $\alpha > 4/7$). Nevertheless, this is generally not the case.

Proposition 7 *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 $X \in \Phi(N, \gamma, \alpha)$ and $X' \in \Phi(N, \gamma, \alpha')$, $|X| > |X'|$ and $\gamma_X > \gamma_{X'}$.*

Proof. The following example is sufficient to establish this result: 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. ■

Intuitively, higher α turns certain coalitions that were otherwise non-self-enforcing into self-enforcing coalitions as expected, 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. This proposition therefore establishes that greater power or "agreement" requirements in the form of supermajority rules do not necessarily lead to larger ruling coalitions.

The next proposition establishes the complementary result that an increase in the power of an individual can remove him out of the ruling coalition. To state this result, let us use the notation $j \in \Phi(N, \gamma, \alpha)$ to denote a situation in which $X \in \Phi(N, \gamma, \alpha)$ such that $j \in X$.

Proposition 8 *There exist 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 $j \in \Phi(N, \gamma, \alpha)$, but $j \notin \Phi(N, \gamma', \alpha)$. Moreover, this remains correct if we require j to be the strongest individual in both cases, i.e. $\gamma'_i = \gamma_i < \gamma_j < \gamma'_j$ for all $i \neq j$.*

Proof. Take $\alpha = 1/2$, five players A, B, C, D, E with $\gamma_A = \gamma'_A = 2$, $\gamma_B = \gamma'_B = 10$, $\gamma_C = \gamma'_C = 15$, $\gamma_D = \gamma'_D = 20$, $\gamma_E = 21$, and $\gamma'_E = 40$. Then $\Phi(N, \gamma, \alpha) = \{A, D, E\}$, while $\Phi(N, \gamma', \alpha) = \{B, C, D\}$, so player E, who is the most powerful player in both cases, belongs to $\Phi(N, \gamma, \alpha)$ but not to $\Phi(N, \gamma', \alpha)$. ■

Proposition 8 shows that being more powerful may be a disadvantage, even for the most powerful player. This 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 9 *Suppose that $N = \{1, \ldots, |N|\}$, $\alpha \in [1/2, 1)$, and $\gamma|_N$ is such that $\gamma_1, \ldots, \gamma_{|N|}$ is an increasing sequence. Consider the game $\Gamma(N, \gamma|_N, \alpha)$. If $\gamma_{|N|} \in \left(\frac{\alpha}{1-\alpha} \sum_{j=2}^{|N|-1} \gamma_j, \frac{\alpha}{1-\alpha} \sum_{j=1}^{|N|-1} \gamma_j \right)$, then either coalition N is self-enforcing or the most powerful individual, $|N|$, is not a part of the ruling coalition.*

Proof. Inequality $\gamma_{|N|} > \frac{\alpha}{1-\alpha}\sum_{j=2}^{n-1}\gamma_j$ implies that any coalition that includes $|N|$, but excludes even the weakest player will not be self-enforcing. The inequality $\gamma_{|N|} < \frac{\alpha}{1-\alpha}\sum_{j=2}^{n-1}\gamma_j$ 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. This completes the proof. ■

6.5 Self-Enforcing Coalitions When $N = 3$

In this subsection, we illustrate the structure of equilibrium ruling coalitions more explicitly for the case $N = 3$. This representation also shows that even in this most simple environment, an increase in α might make it less likely that larger coalitions emerge as the ruling coalition.

We use the geometric representation already introduced in the Introduction. Generally, the geometric representation of an N player game uses the $(N-1)$-dimensional simplex to depict all potential power allocations, which are represented by points $(\gamma_1, ..., \gamma_N)$ with $\gamma_i \geq 0$ and $\sum_i \gamma_i = 1$ (where this last equality is without loss of generality). As discussed in the Introduction, there are two kinds of constraints that define the set of all self-enforcing coalitions: "power constraints" $\left\{\sum_{j\in K}\gamma_j \geq \alpha\right\}$ which are always parallel to be respective $(K-1)$-dimensional facet, and "enforcement constraints," which correspond to (quasi-)cones.

Figure 2 shows the evolution of the set of self-enforcing coalitions as α changes from 1/2 (simple majority) to 1 (unanimous voting rule) for the case with $N = 3$. The set of power configurations such that the grand coalition is the ruling coalition is shaded. The first panel corresponds to the case $\alpha = 1/2$. For any point $(\gamma_1, \gamma_2, \gamma_3)$ outside the shaded triangle, there is some member i who has power $\gamma_i > 1/2$. The second panel corresponds to the case when α becomes larger than 1/2; it demonstrates that the set of self-enforcing coalitions, while remaining a union of a finite number of convex sets, may have non-convex connected components. Interestingly, the "central coalitions", i.e. those close to $(1/3, 1/3, 1/3)$, which were self-enforcing when $\alpha = 1/2$, cease to be self-enforcing when α increases. The reason is that with $\alpha > 1/2$, there is a range of 2-person self-enforcing coalitions; the fact that they are self-enforcing makes 3-person coalitions containing them non-self-enforcing (Figure 2). When α is large enough, but still less than 2/3, the set of self-enforcing coalitions coalition set becomes a union of a finite number of convex connected components (namely, of three trapezoids). When $\alpha = 2/3$, the trapezoids become triangles.

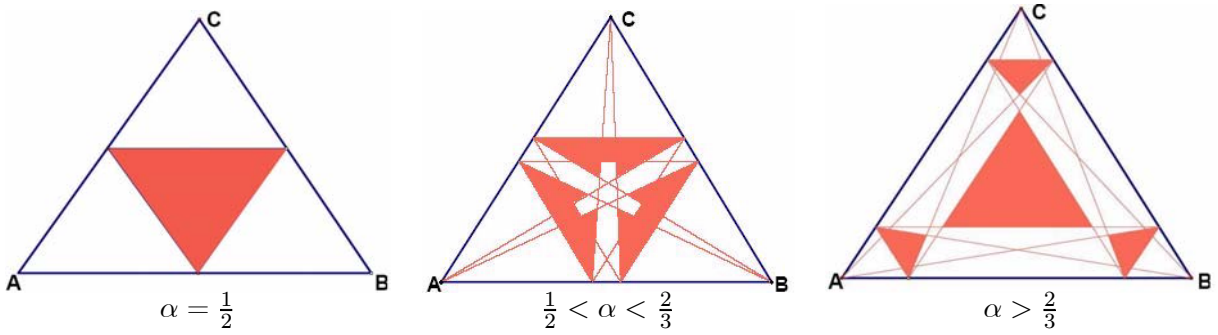


Figure 2

The third panel shows that when $\alpha > 2/3$ (and this generalizes straightforwardly to $\alpha > (N-1)/N$ for an arbitrary N), there is a new part to the self-enforcing coalition set around. This demonstrates that the self-enforcing coalition set is non-monotonic in α: the coalitions close to $(1/3, 1/3, 1/3)$ are self-enforcing again. This new set of self-enforcing coalitions increases with α and eventually grows to cover all points when α approaches 1, but for all α such that $2/3 < \alpha < 1$, it is a joint of four triangles as shown in the third panel. Obviously, points in the "middle" set of self-enforcing coalitions are more stable than other self-enforcing coalitions: even if there is a random shock that eliminates some players, the remainder is a self-enforcing coalition.

7 Extensions

In this section, we discuss two extensions of our basic framework. First, we show how this framework can be applied for thinking about endogenous party formation, and second we show how reallocation of power within a group can be beneficial for all group members (which is a form of "giving up the guns" mentioned in footnote 7).

7.1 Party Formation

If there were no enforcement constraints, the *minimal winning coalition* would always emerge as the ruling coalition.[15] However, we have seen that the ultimate ruling coalition is not necessarily the minimal winning coalition. This is because the minimal winning coalition might not be self-enforcing (e.g., as is coalition $(3, 4)$ in $(3, 4, 5)$), and thus cannot form a ruling coalition. What prevents the formation of this coalition is the fact that its members do not to trust each other. If somehow they could enter into binding agreements, the minimum winning coalition could emerge

[15] $X \in P(N)$ is a minimal winning coalition within N if $\gamma_X > \alpha\gamma_N$ and $\gamma_X \leq \gamma_Z$ for any $Z \in P(N)$ such that $\gamma_Z > \alpha\gamma_N$.

as the ultimate willing coalition. In this subsection, we think of party formation as a way of forming binding agreements among a subset of agents. In particular, we allow some of the players to form permanent alliances, effectively merging into a single member with combined power. Another way is to allow members of a coalition freely transfer (shares of) their power to each other to make the coalition self-enforcing, which is explored in the next subsection.

More specifically, consider the party-formation game $\hat{\Gamma}'$, which is identical to the game $\hat{\Gamma}$ in Section 4, except that there is now a first stage before $\hat{\Gamma}$ is played, in which a subset of agents can form a binding coalition, a "party," and this party plays the game as a single agent. The result of this party-formation game will be that the minimal winning coalition will form a party and will guarantee power for its members.

Proposition 10 *Suppose Assumption 1 holds and let X be the minimal winning coalition, i.e., $X = \{Z \in P(X) : \gamma_Z > \alpha\gamma_N$ and $\nexists Y$ with $\gamma_Y \in (\alpha\gamma_N, \gamma_Z)\}$. Then, in the party-formation game $\hat{\Gamma}'(N, \gamma|_N, \alpha)$, the URC is X.*

Proof. The proof follows the steps of the proof of Theorem 2 and is omitted. ■

7.2 Power Exchange

We have seen that individuals can be made worse off by having more power. This naturally raises the question of whether individuals would like to relinquish their power (for example, give up their guns in the context of fighting preceding political decision-making). We now investigate this issue under the assumption that $\alpha = 1/2$. Our main result is that any minimal winning coalition X can redistribute power in such a way that it becomes self-enforcing, and each member of X is better off than he would have been without power redistribution.

The next result demonstrates that when the size of the minimal winning coalition exceeds 2 (i.e., if $|X| \geq 3$), X can redistribute power to become self-enforcing, with each member becoming strictly better off than they would have been in the initial allocation.

Theorem 5 *Suppose that $\alpha = 1/2$, the grand coalition N is ruling, and X is a minimal winning coalition in N. Then, provided that $|X| \geq 3$, there exists a redistribution of power among the members of X such that X becomes the ruling coalition and implements a payoff $\widehat{w}_i$ for each $i \in N$, such that $\widehat{w}_i > w_i(N)$ for all $i \in X$.*

Proof. We will make our argument in terms of agents' power $(\gamma_1, ..., \gamma_N)$ and then use our assumptions about payoff functions to prove the claim. Without loss of generality, assume that $\sum_{i \in N} \gamma_i = 1$, and denote $W = \sum_{i \in N \setminus X} \gamma_i$. We will prove the claim in the theorem in two steps.

(Step 1) Suppose that there exists $k \in X$ such that coalition $X \setminus \{k\}$ is not self-enforcing.

Consider the parametrized family $(\gamma_i^\beta)_{i \in X}$ of distributions of power in coalition X : $\gamma_k^\beta = \gamma_k + W\beta$ and

$$\gamma_i^\beta = \gamma_i + \frac{\gamma_i}{\sum_{i \in X \setminus \{k\}} \gamma_i} W(1-\beta).$$

When $\beta = 1$, k alone forms a winning coalition, since

$$\gamma_k + W \geq \sum_{i \in X \setminus \{k\}} \gamma_i.$$

(Otherwise, coalition $X \setminus \{k\}$ is winning which contradicts the minimality of X.) We claim that there exists some β such that with $\left(\gamma_i^\beta\right)_{i \in X}$ X becomes a self-enforcing coalition and since it is the minimal winning coalition, it becomes the ruling coalition.

Let β_0 be determined by

$$\gamma_k + W\beta_0 = \sum_{i \in X \setminus \{k\}} \gamma_i + W(1-\beta_0).$$

(Such β_0 exists since $\gamma_k < \sum_{i \in N \setminus \{k\}} \gamma_i$ by assumption). Since $0 < \beta_0 < 1$, $\gamma_i^{\beta_0} > \gamma_i$ for any $i \in X$.

Now let ζ be a positive number such that

$$\begin{aligned} \gamma_k^{\beta_0} - \zeta &< \sum_{i \in X \setminus \{k\}} \left(\gamma_i^{\beta_0} + \frac{\zeta}{|X|-1}\right) \\ \gamma_k^{\beta_0} - \zeta &> \sum_{i \in X \setminus \{k,j\}} \left(\gamma_i^{\beta_0} + \frac{\zeta}{|X|-1}\right) \end{aligned}$$

for any $j \in X \setminus \{k\}$. Then coalition $\left(\widehat{\gamma}_k = \gamma_k^{\beta_0} - \zeta, \left(\widehat{\gamma}_i = \gamma_i^{\beta_0} + \frac{\zeta}{|X|-1}\right)_{i \neq k}\right)$ is a ruling coalition, implementing a payoff vector with higher payoffs for each member of the minimal winning coalition.

(Step 2) Now suppose that for any $k \in X$, coalition $X \setminus \{k\}$ is self-enforcing. Then any coalition $X \setminus \{k, j\}$ is not self-enforcing (adding one player to a self-enforcing coalition makes it non-self-enforcing). Assume without loss of generality that $\gamma_1 = \max_{i \in X} \gamma_i$, and $\gamma_M = \min_{i \in X} \gamma_i$. Coalition $X \setminus \{1, M\}$ is not self-enforcing. Therefore, there exists a coalition $Y \subset X \setminus \{1, M\}$ such that Y is self-enforcing and

$$\gamma(Y) > \gamma((X \setminus \{1, M\}) \setminus Y).$$

Since $\gamma(X \setminus \{1\}) - \gamma(X \setminus \{1, M\}) = \gamma_M$, if we divide γ_M proportionally between the members of Y coalition, the resulting new coalition Y' will make coalition $X \setminus \{1, M\}$ not self-enforcing and $\widehat{\gamma}_i > \gamma_i$ for all $i \in X$.

Now using the result from Step 1, it suffices to show that

$$\gamma_1 + (W - \gamma_M) > \sum_{i \in X \setminus \{1,M\}} \gamma_i + \gamma_M = \sum_{i \in X \setminus \{1\}} \gamma_i. \tag{11}$$

(This would imply that $\gamma_1 + (W - \gamma_M) > \sum_{i \in X \setminus \{1,j\}} \gamma_i + \gamma_M$ for any $j \in \{2, \ldots, M\}$, which would mean that 1 cannot form a self-enforcing coalition with less than all $M - 1$ members of $X \setminus \{1\}$.)

Inequality (11) is equivalent to

$$1 - \gamma_M > 2 \sum_{i \in X \setminus \{1\}} \gamma_i$$

(we add $\sum_{i \in X \setminus \{1\}} \gamma_i$ to both sides of (11), and use the fact that $\sum_{i \in X} \gamma_i + W = 1$), which is in turn equivalent to

$$\frac{1}{2} > \sum_{i \in X \setminus \{1\}} \gamma_i + \frac{\gamma_M}{2}. \tag{12}$$

To prove that (12), note that if $\frac{1}{2} < \sum_{i \in X \setminus \{1\}} \gamma_i + \frac{\gamma_M}{2}$, then $\sum_{i \in N \setminus X} \gamma_i + \frac{\gamma_M}{2} < \frac{1}{2}$ and

$$\sum_{i \in X} \gamma_i = \left(\sum_{i \in X \setminus \{1\}} \gamma_i + \frac{\gamma_M}{2} \right) + \frac{\gamma_M}{2} > \left(\sum_{i \in N \setminus X} \gamma_i + \frac{\gamma_M}{2} \right) + \frac{\gamma_M}{2} = \sum_{i \in N \setminus X} \gamma_i + \gamma_M.$$

Since $\sum_{i \in N \setminus X} \gamma_i + \gamma_M > \frac{1}{2}$, the above inequality yields that $(N \setminus X, \gamma_M)$ rather than X is the minimal winning coalition, which is a contradiction. Therefore, $\gamma_i' > \gamma_i$ for any $i \in X$. Using assumptions 1–3 about the payoff functions completes the proof of Theorem 5. ■

We also conjecture (but have not yet proved) that Theorem 5 holds for any $\alpha > 1/2$.

8 Conclusion

The central question of political economy is how collective choices are made among a group of individuals with conflicting preferences (and potentially different "powers"). We study this question in the context of a game of endogenous coalition formation. We assume that each individual is endowed with a level of political power, which may be derived from his or her specific skills or access to resources (guns, money etc.). The ruling coalition consists of a subset of the individuals in the society and decides the distribution of resources. The main innovation of our approach is that we also require ruling coalitions to be *self-enforcing*, in the sense that none of the subcoalitions of this ruling coalition should be able to secede and become the new ruling coalition.

We first model these issues using an axiomatic approach based on four axioms. The two important axioms are that the ruling coalition should be powerful enough (the *power constraint*) and that the ruling coalition should be self-enforcing (the *enforcement constraint*). We prove that

there exists a unique mapping, which is generically single-valued, that satisfies these axioms. This provides an axiomatic way of characterizing the ruling coalitions for any game.

We support this notion by showing that the result of our axiomatic analysis also follows from the "reasonable equilibria" of a dynamic game of coalition formation and also as the unique core allocation of a related non-transferable cooperative game. In particular, we construct a simple dynamic game that captures the same notions that a ruling coalition should have a certain amount of power and should be self-enforcing (stable). As with other dynamic voting games, this game possesses many subgame perfect equilibria. We propose the notion of *sequentially weakly dominant equilibrium* as an equilibrium concept for this and related games (which referred to as *agenda-setting games*). We prove that agenda-setting games always have sequentially weakly dominant equilibria and Markov trembling and perfect equilibria. Moreover, in our dynamic game, both concepts generically yield a unique equilibrium allocation.

After establishing these results on the existence of equilibria and ruling coalitions in related axiomatic, noncooperative and cooperative games, we present a series of results on the structure of ruling coalitions. In particular, we establish the following results:

- Despite the simplicity of the environment, the ruling coalition can be of any size relative to the society, and may include or exclude more powerful individuals in the society. Consequently, the equilibrium payoff of an individual is not monotone in his power.

- Self-enforcing coalitions are generally "fragile," especially under majority rule. For example, under majority rule, adding or subtracting one player from a self-enforcing coalition makes it non-self-enforcing. Despite this type of fragility of self-enforcing coalitions, we also show that the ruling coalition is "continuous," in the sense that two games where the powers of the players are sufficiently close will have the same ruling coalition.

- Somewhat paradoxically, an increase in α—that is an increase in the degree of supermajority necessary to make decisions—does not necessarily lead to larger ruling coalitions. Also, the most powerful individual will be often excluded from the ruling coalition, unless he is powerful enough to win by himself or weak enough so as to be part of smaller self-enforcing coalitions.

- Coalitions of certain sizes are more likely to emerge as the ruling coalition. For example, with majority rule, i.e., $\alpha = 1/2$, the ruling coalition cannot (generically) consist of two individuals. Moreover, again when $\alpha = 1/2$, coalitions where members have roughly the same power exist only when the coalition's size is $2^k - 1$ where k is an integer.

There are a number of natural areas for future study. A similar approach blending axiomatic foundations and dynamic games can be adopted to analyze the structure of ruling coalitions in a more general class of political games, where there are multiple resources to be distributed (or multiple policies over which individuals disagree). Our results on general agenda-setting games suggest that the approach here might be extended to this more general setting. Another interesting area for future research would be to investigate what types of coalitions will form when there is some randomness in the environment, for example, if the powers or preferences of different individuals may change by a small amount after the coalition is formed. Such an approach would allow us to determine the extent to which a coalition is "robust" and also to quantify what "price" the coalition is willing to pay for robustness by including individuals that may not be necessary for obtaining a majority.

Appendix A: Markov Trembling-Hand Perfect Equilibria

We now introduce the notion of Markov Trembling-Hand Perfect Equilibria (MTHPE) and establish both a number of general results about MTHPE and also prove that the results in Section 4 apply with MTHPE in the same way as they do with SWDE. With the terminology in subsection 4.2, we have:

Definition 10 A continuation strategy $\sigma^{i,t}$ is *Markovian* if

$$\sigma^{i,t}\left(h^{t-1}\right)=\sigma^{i,t}\left(\tilde{h}^{t-1}\right)$$

for all $h^{t-1},\tilde{h}^{t-1}\in H^{t-1}$ such that for any $a^{i,t},\tilde{a}^{i,t}\in A^{i,t}$ and any $a^{-i,t}\in A^{-i,t}$ we have

$$u^{i}\left(a^{i,t},a^{-i,t}\mid h^{t-1}\right)\geq u^{i}\left(\tilde{a}^{i,t},a^{-i,t}\mid h^{t-1}\right)$$

implies that

$$u^{i}\left(a^{i,t},a^{-i,t}\mid \tilde{h}^{t-1}\right)\geq u^{i}\left(\tilde{a}^{i,t},a^{-i,t}\mid \tilde{h}^{t-1}\right).$$

Definition 11 We say that a strategy profile $\left(\hat{\sigma}^{1},...,\hat{\sigma}^{n}\right)$ of an extensive-form game in agent-strategic form is *Markov Trembling-Hand Perfect Equilibrium (MTHPE)* if there exists a sequence of totally mixed Markovian strategy profiles $\left\{\left(\hat{\sigma}^{1}(m),\ldots,\hat{\sigma}^{n}(m)\right)\right\}_{m\in\mathbb{N}}$ (meaning that continuation strategy $\sigma^{i,0}(m)$ is Markovian for all $i=1,\ldots,n$ and all $m\in N$) such that $\left(\hat{\sigma}^{1}(m),\ldots,\hat{\sigma}^{n}(m)\right)\rightarrow\left(\hat{\sigma}^{1},\ldots,\hat{\sigma}^{n}\right)$ as $m\rightarrow\infty$ and

$$u^{i}\left(\hat{\sigma}^{i},\hat{\sigma}^{-i}(m)\right)\geq u^{i}\left(\sigma^{i},\hat{\sigma}^{-i}(m)\right)\text{ for all }\sigma^{i}\in\Sigma^{i},\text{ for all }m\in M\text{ and for all }i=1,\ldots,n.$$

We say that a strategy profile $\left(\hat{\sigma}^{1},...,\hat{\sigma}^{n}\right)$ of an extensive-form game in strategic form is *Markov Trembling-Hand Perfect Equilibrium (MTHPE)* if it is MTHPE is corresponding agent-strategic form game.

Note that MTHPE is defined directly in the agent-strategic form in order to avoid standard problems that arise when trembling hand perfection is defined on the strategic form (e.g., Selten, 1975, Osborne and Rubinstein, 1994). After characterizing the SWDEs of our game, we will also characterize the MTHPEs for game $\hat{\Gamma}$ and show their equivalence.

We next show that for our extensive form game the MTHPE refinement leads to the same equilibrium URC and payoffs as the SWDE. Our main results are contained Theorem 7, which is the equivalent of Theorem 2 for MTHPE under Assumption 1. In addition, we will establish that an MTHPE exists in a more general class of political games, which we refer to as *agenda-setting games*.[16] We will also show that in this class of games, every MTHPE is an SWDE, so an SWDE exists. Nevertheless, these two equilibrium concepts do not always coincide. First, a MTHPE always exists, while SWDE may not. Second, there may exist SWDEs that are not MTHPE (see Theorem 6 and Appendix B).

Let us first define general agenda-setting games, which include most voting games as a special case, and establish the existence of MTHPE and SWDE for these games.

Definition 12 A finite perfect-information game Γ in extensive form with a set of players $N\cup\{\text{Nature}\}$ is called an *agenda-setting game* if and only if at each stage ξ either

1. only one player (possibly Nature) moves, or
2. there is voting among the players in $X\subset N$. Voting means that
   (a) each player $i\in X$ has two actions, say $a_{i}^{y}(\xi)$ and $a_{i}^{n}(\xi)$;
   (b) those in $N\setminus X$ have no action or only one action at this stage;

[16] Another trembling hand refinement used in the literature, *truly perfect equilibrium*, is stronger than our notion of MTHPE. A truly perfect equilibrium requires strategies from σ to be best responses to all fully mixed profiles in some neighborhood of σ rather than to one sequence of profiles in the standard case and to one sequence of Markovian profiles in the case of MTHPE. However, this equilibrium concept fails to exist in many games, including our dynamic game of coalition formation (except in some special cases).

(c) there are only two equivalence classes of subgames following node ξ (where equivalence classes of subgames include subgames that are continuation payoff identical), say $y(\xi)$ and $n(\xi)$;

(d) for each player $i \in X$ and for any other players' actions held fixed, the action $a_i^y(\xi)$ does not decrease the probability of moving into the equivalence classes of subgames $y(\xi)$.

This definition states that any game in which one of the agents makes a proposal and others vote in favor or against this proposal is an agenda-setting game. Moreover, any perfect-information game where players move sequentially is an agenda-setting game, as is any game with consecutive votings over pairs of alternatives. Clearly, our dynamic game here is an agenda-setting game. We prove the following general result about equilibria in agenda-setting games.

Theorem 6 *1. Any finite extensive-form game has a MTHPE (possibly in mixed strategies).*

2. Any agenda-setting game has a MTHPE and a SWDE in pure strategies.

3. In an agenda-setting game, any MTHPE is a SWDE.

4. There exist games where a MTHPE is not a SWDE.

5. There exist agenda-setting games where a SWDE is not a MTHPE.

Proof. (Part 1) Consider a perturbed game where each player $i \in N$ is confined to play each action at each stage with probability $\eta_{k,t}^i \geq \underline{\eta} > 0$, where $\underline{\eta}$ is a small number, to each of its finite number of actions in each stage. By the standard fixed point theorem argument, this perturbed game has a Nash equilibrium; moreover, the fixed point theorem applies if we restrict our attention to Markovian strategies only. Therefore, the perturbed game has a Markov Perfect Equilibria $\left(a^1\left(\underline{\eta}\right), \ldots, a^n\left(\underline{\eta}\right)\right)$. Because the action space has finite dimensions and is thus compact, we can choose a sequence $\underline{\eta}^1, \underline{\eta}^2, \ldots$ which converges to 0 such that $\left(a^1\left(\underline{\eta}^k\right), \ldots, a^n\left(\underline{\eta}^k\right)\right)$ has a limit. This limit would be a trembling-hand perfect equilibrium in Markovian strategies, i.e. an MTHPE.

(Part 2) Let us prove that an agenda-setting game has a MTHPE in pure strategies, then the result for SWDE will immediately follow from Part 3. We do this by induction on the number of stages. The base is trivial; indeed, consider a one-shot agenda-setting game. If this stage is one-player move, then, evidently, the action which maximizes his utility constitutes a pure strategy MTHPE, since it is Markovian and trembling-hand perfect. If the single stage is voting, then each player i weakly prefers one of the outcomes to another. It is trivial to check that voting for a weakly preferred outcome is an MTHPE.

Now proceed with the induction step. Suppose that we have proved the existence of pure strategy MTHPE in all agenda-setting games with number of stages less than T; take an agenda-setting game with T stages. Consider its first stage. Suppose that there is one player i making a choice between k actions $a_1, \ldots, a_k$. In each of k corresponding subgames there exists (by induction) a pure strategy MTHPE; we can choose the same MTHPE for isomorphic subgames. Therefore, there exist sequences of strategy profiles $\sigma_j^1, \sigma_j^2, \ldots$ for each $j \in \{1, \ldots, k\}$ which converge to a pure strategy MTHPE for each such j and which are Nash equilibria in constrained game; moreover, we can require that if two subgames (for $j = j_1, j_2$ are isomorphic, then $\sigma_{j_1}^n = \sigma_{j_2}^n$ for any n. Now, consider player i at first stage. For any n, there is j_n such that a_{j_n} is weakly better than other actions. Since there are a finite number of actions, there is action j such that a_j is weakly better than other actions for infinitely many values of n. It is now straightforward to prove that a_j, along with the chosen MTHPE, forms a pure strategy MTHPE of the whole game. Now suppose that the initial stage is a voting. Then there are two subgames; take a pure strategy MTHPE in each (if they are isomorphic, take the same MTHPE), and we can similarly construct two sequences of strategy profiles $\sigma_y^1, \sigma_y^2, \ldots$ and $\sigma_n^1, \sigma_n^2, \ldots$ Each player i has two actions, y and n, in stage 1. Consider one of players (i) and take any m. If in the subgames strategies σ_y^m and σ_n^m are played, respectively, then player i weakly prefers to choose one of the actions, y or n, to another one. There is an action $a(i)$ which is weakly preferred for infinite number of m's. Now consider two subsequences of sequences $\sigma_y^1, \sigma_y^2, \ldots$ and $\sigma_n^1, \sigma_n^2, \ldots$ which are formed by values of m for which $a(i)$ is weakly better for player i. Take another player, j, and repeat the procedure; then we will get action $a(j)$ and two subsequences of the previous subsequences. Since the number of players is finite, we can proceed in this way, and for any player i find action $a(i)$. It is now evident

that the MTHPEs chosen for the stages starting from the second one and actions $a(i)$ for each player i at the first stage, form a pure strategy MTHPE.

(Part 3) Next take any strategy profile σ that forms a MTHPE. This is proved by backward induction on the number of stages in the game. Suppose that the Lemma has been proved for games with $q' < q$ stages. Consider an agenda-setting game with q stages and take any MTHPE in it. By induction, this MTHPE, when truncated to any of the game's proper subgames, forms a SWDE. Consider its first stage.

Suppose that only one player i moves at this stage and denote his expected utility (in this MTHPE) from making action a at first stage by u_i^a. If action a^* is an action played with a non-zero probability in equilibrium then $u_i^{a^*} \geq u_i^a$ for any other feasible action a (otherwise there would exist a payoff-improving deviation). Hence, all actions played in a MTHPE with a non-zero probability yield the same expected utility for player i, and this utility is maximum possible over the set of feasible actions. Hence, this MTHPE is a SWDE.

Now consider the other situation where the first stage is a voting stage. Consider a profile σ' consisting of fully mixed strategies and suppose that it is η-close to σ for a small η. Depending on how other players vote, three mutually exclusive situations are possible: proposal is accepted regardless of how player i votes, it is rejected regardless of how he votes, and player i is pivotal; let μ^+, μ^-, and μ^p be the respective probabilities of these events. By definition, $\mu^+ + \mu^- + \mu^p = 1$, and by assumption $\mu^p > 0$. Voting for the proposal yields $(\mu^+ + \mu^p) u_i^{+\prime} + \mu^- u_i^{-\prime}$ in expectation, voting against it yields $\mu^+ u_i^{+\prime} + (\mu^- + \mu^p) u_i^{-\prime}$ where $u_i^{+\prime}$ and $u_i^{-\prime}$ are i's utilities from acceptance and rejection of the proposal if profile σ' is played. Thus, if $u_i^{+\prime} > u_i^{-\prime}$ then player i's sole best response is voting for the proposal, and if $u_i^{+\prime} < u_i^{-\prime}$ it is voting against it. If η is sufficiently small then $u_i^+ > u_i^-$ implies $u_i^{+\prime} > u_i^{-\prime}$, and thus by definition of MTHPE player i must support the proposal in equilibrium with probability one. Similar reasoning applies to the case $u_i^+ < u_i^-$.

Now take any player i who participates in voting. If $u_i^+ > u_i^-$, then he votes for the proposal in this MTHPE. This is a weakly dominant strategy for him (given continuation strategies of himself and other players). Similarly, if $u_i^+ < u_i^-$ then the strategy he plays in this MTHPE is weakly dominant. If, $u_i^+ = u_i^-$ or the player is never pivotal, any strategy is weakly dominant. Therefore, for any player, the strategy he plays in this MTHPE is weakly dominant, and thus this MTHPE is a SWDE. This completes the induction step.

(Part 4) Consider a one-stage game with two players making simultaneous moves with payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} l & r \end{array} \\ \begin{array}{c} L \\ R \end{array} & \begin{array}{cc} (1,1) & (0,0) \\ (0,0) & (1,1) \end{array} \end{array}.$$

This game does not have SWDE, because it is one-stage and in that only stage neither of the players has a weakly dominant strategy. It is straightforward to check, however, that both (L,l) and (R,r) are MTHPEs of this game.

(Part 5) This follows from Example 7 in Appendix B. ■

In addition to the existence results, which are of interest in and of themselves, Theorem 6 establishes that while MTHPE and SWDE are not subsets of one another in general, but for agenda-setting games an MTHPE is always a SWDE (so within this class of games SWDE is a stronger concept). This implies, in particular, existence of a SWDE in an agenda-setting game, which then follows from existence of an MTHPE.

Theorem 7 *Any extensive-form game $\hat{\Gamma}(N, \gamma|_N, \alpha)$ has at least one pure strategy MTHPE. Moreover, suppose that Assumption 1 holds. Then any MTHPE (in pure or mixed strategies) is a SWDE, and, in particular, the only ultimate ruling coalition (URC) is given by $\phi(N)$ as defined in Theorem 1. The URC is reached after one stage of elimination, and the payoff of each $i \in N$ is given by* (7).

Proof. For any N, game $\hat{\Gamma}(N, \gamma|_N, \alpha)$ is an agenda-setting game as it satisfies Definition 12. By Theorem 6 a pure strategy MTHPE always exists and, moreover, any MTHPE is a SWDE. The rest of the Theorem 7 follows immediately from the characterization of SWDEs in Theorem 2. ■

9 Appendix B: Examples

SPNE and MPE in the Dynamic Game

Example 4 Let $\alpha = 1/2$, $|N| = 4$, $N = \{A, B, C, D\}$, players' powers be given by $\gamma_A = 4$, $\gamma_B = 5$, $\gamma_C = 6$, $\gamma_D = 8$, and players' utilities from coalitions be given by (2). Take some MTHPE σ^* of the dynamic game $\hat{\Gamma}(N, \gamma|_N, \alpha)$ and consider the following modification σ_X, where X is one of the six three-player coalitions of players from N. Before any elimination, if X is proposed, player i votes $\tilde{y}$ if $i \in X$ and votes $\tilde{n}$ if $i \notin X$; if any proposal other than X is made, all players vote $\tilde{n}$. If player $i \in X$ is the agenda-setter, he proposes coalition X, while if $i \notin X$ is the agenda-setter, the proposal is N. After the first elimination has taken place, the corresponding part of σ^* is played.

Using the fact that σ^* is MTHPE, it is easy to see that strategies from σ_X are Markovian, it is also straightforward to check that σ_X forms a SPNE. Before any elimination occurred, it is not profitable to deviate at the stage of voting. Indeed, if the proposal is X, a deviation by player $i \notin X$ does not change the outcome of voting, while if player $i \in X$ makes a one-shot deviation, the URC will be either X or N; the latter outcome being worse for such i than X, because $w_i(X) > w_i(N)$. If the proposal under consideration is some other proposal $Y \neq X$, then none of the players is pivotal, and a deviation will not change the outcome of the voting or the subsequent equilibrium strategies (because strategies are Markovian). It is also straightforward to check that any deviation is not profitable at the stage where proposals are made (mainly because no proposal other than X may be accepted in the subsequent voting). After the first elimination, strategies from σ^* are played. Given all this, σ_X is both a SPNE and a MPE. Evidently, the URC in σ_X is X with probability 1.

So, for any of the six three-player coalitions, we have constructed a strategy profile which is a SPNE and, moreover, a MPE. Therefore, none of these refinements help us get a unique prediction of the outcome of the dynamic game (even though Assumption 1) is satisfied.

Dynamic Game Without ε

Example 5 Let $\alpha = 1/2$, $|N| = 4$, $N = \{A, B, C, D\}$, players' powers be given by $\gamma_A = 2$, $\gamma_B = 4$, $\gamma_C = 7$, $\gamma_D = 10$, and players' utilities from coalitions be given by (2). Theorem 2 establishes that if $\varepsilon > 0$, the only possible URC that may emerge in SWDE is $\{B, C, D\}$ (it is straightforward to check that any other winning coalition is not self-enforcing). Denote the pure strategy SWDE profile constructed in the proof of Theorem 2 by σ^*; note that it does not depend on ε. Because all utilities are continuous in ε, we immediately get that even if $\varepsilon = 0$, σ^* forms a SWDE.

Consider strategy profile σ' which coincides with σ^* for all players and histories except the following. If the first proposer $i_{0,1} \in \{A, D\}$, then he proposes $X_{0,1} = \{A, D\}$. If the first proposal is $X_{0,1} = \{A, D\}$, then players A and D cast vote $\tilde{y}$, while players B and C cast vote $\tilde{n}$. Let us show that this strategy profile forms a SWDE.

Suppose the first proposal is $X_{0,1} = \{A, D\}$. If it is accepted, then B and C are eliminated, and subsequently D eliminates A. Therefore, $u_A^- = u_B^- = u_C^- = 0$, $u_D^- = 1$ (because $\varepsilon = 0$), while if it is rejected, then, as follows from the construction of σ^*, $\{B, C, D\}$ will be the URC, resulting in $u_A^- = 0$, $u_B^- = 4/21$, $u_C^- = 1/3$, $u_D^- = 10/21$. It is evident that voting $\tilde{y}$ is weakly dominant for A and D, and voting $\tilde{n}$ is weakly dominant for B and C. Now suppose that A or D is the first to make proposal. If he makes proposal $\{A, D\}$ then, as we have just proved, it will be the URC. If he makes any other proposal, then, as again follows from construction of σ^*, $\{B, C, D\}$ will be the URC. Therefore, proposing $\{A, D\}$ is weakly dominant for A and strictly dominant for D. Finally, in any other node players' behavior is optimal because σ^* coincides with σ' in the subsequent subgame. Therefore, σ' is indeed a SWDE.

In SWDE σ', the URC is $\{B, C, D\}$ if the first proposer (chosen by Nature) is B or C, and $\{D\}$ if the first proposer is A or D (indeed, once proposal $\{A, D\}$ is accepted, player A will be eliminated). Hence, this equilibrium, σ', features at least two unappealing properties. First, the outcome may depend on the order of proposals. Second, players that will be eliminated anyway (A in this example) may have have a non-trivial effect on the outcome, depending on how they vote when they are indifferent. Introducing small organizational costs $\varepsilon > 0$ allows us to get rid of these effects and get equilibria where the ultimate ruling coalition that emerges in equilibrium is uniquely determined.

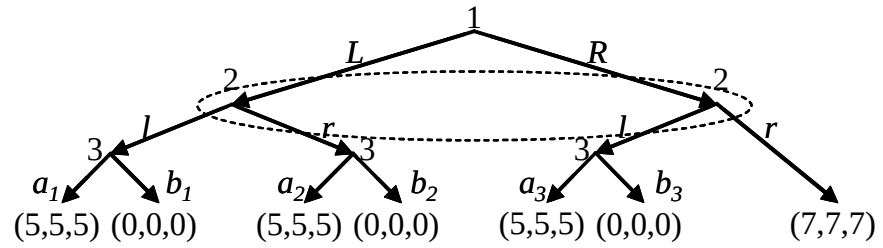


Figure 3: A Game With Herding in Trembling-Hand Perfect Equilibrium.

It may seem at first glance that σ' is not a MTHPE, because player A will not vote for proposal $\{A, D\}$ but it actually is one. Without going into much detail, A will vote for $\{A, D\}$ if it is much more likely that D will make a mistake and not propose $\{D\}$ after elimination of B and C than that B, C, and D will make a mistake and fail to eliminate A if proposal $\{A, D\}$ is not accepted.

THPE vs. MTHPE

Example 6 Consider a game of three players with extensive form and payoffs as shown on Figure 3. The first two players vote, and if both vote for the 'right', all three players receive first-best; if one of them votes for the 'left' then the third player chooses between 'moderate' and 'bad'. All players receive the same in all terminal nodes, so there is no strategic conflict between them.

Equilibrium $(R, r, (a_1, a_2, a_3))$ is trembling-hand perfect, but so is $(L, l, (a_1, a_2, a_3))$ where efficiency is not achieved because of 'herding' in voting (note that neither L not l are dominated strategies: for instance, L is best response to second player playing l and third player playing (a_1, b_2, b_3)). Indeed, take some η and consider

$$\sigma^n = \left(\left(1-\eta^3\right) L + \eta^3 R, \left(1-\eta^3\right) l + \eta^3 r, \left(\left(1-\eta^2\right) a_1 + \eta^2 b_1, (1-\eta)\, a_2 + \eta b_2, (1-\eta)\, a_3 + \eta b_3\right)\right).$$

Evidently, player 3 (and all his agents in agent-strategic form) are better off choosing a_1 over b_1, a_2 over b_2, and a_3 over b_3. Now consider payoffs of player 1 choosing L or R. If he chooses L, he gets $u_L = 5\left(\left(1-\eta^3\right)\left(1-\eta^2\right) + \eta^3\left(1-\eta\right)\right) = 5 - 5\eta^2 - 5\eta^4 + 5\eta^5$. If he chooses R, he gets $u_R = 5\left(\left(1-\eta^3\right)(1-\eta)\right) + 7\eta^3 = 5 - 5\eta + 2\eta^3 + 5\eta^4$. Hence, For small η, player 1 should put all weight to L, and a similar argument would show that player 2 should put all weight to l. This proves that $(L, l, (a_1, a_2, a_3))$ is also a trembling-hand perfect equilibrium.

The effect that Example 6 emphasizes would not be the case if fully mixed profiles σ^n were required to be Markovian, which is what our definition of MTHPE imposes. Indeed, it is a natural restriction to require that in the three subgames where player 3 moves and payoffs are identical, his mixed action profile σ^n should lead to identical place. In that case, the increase of utility of player 1 due to the possibility of player 2 playing r instead of l would be not be offset by worse development in the subgame if he still plays l.

SWDE vs. MTHPE

Example 7 Consider a game of two players with extensive form depicted on Figure 4. This is an agenda-setting game, because at each stage only one player has a (non-trivial) move. It game has exactly one MTHPE (R, r). However, there are two SWDEs: (R, r) and (L, r). The latter is not MTHPE, because if there is a non-zero chance that player 2 will play l, player 1 is better off putting all weight to R.

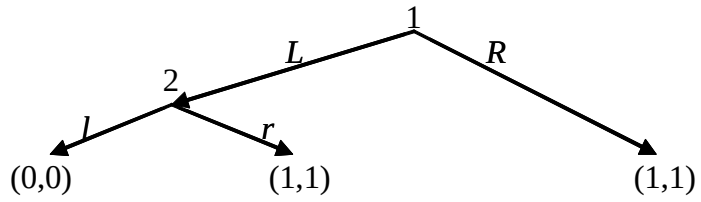


Figure 4: A Game With SWDE which is not MTHPE

Appendix C: Proofs of Theorems 2 and 3

Proof of Theorem 2.

(Part 1 of Theorem) The first part of the Theorem (existence of pure strategy SWDE with in the required properties) is proved by induction on the size of coalition $|N|$.

The base of the induction is straightforward. If $|N| = 1$, then from Theorem 1, $\phi(N) = \{N\}$. It is also straightforward that in the unique SWDE, the single player i will propose coalition $X_{0,1} = N$ and then vote in favor of it, i.e., $v_{0,1}(i) = \tilde{y}$. Therefore, when $|N| = 1$, there exists a pure strategy SWDE, with the URC reached after at most one stage elimination and coincides with an element of $\phi(N)$.

Suppose next that we have proved this result for all coalition sizes less than n, and we will prove it for any coalition N with $|N| = n$.

First, we introduce the following notation. Let v be a one-to-one map between $P(N)$ and the set $\{1, \ldots, 2^{|N|} - 1\}$; let ν_X denote the number corresponding to $X \subset N$, and assume that the set M corresponds to 1 (i.e., $\nu_M = 1$). Define a single-valued mapping $\phi^* : P(N) \to P(N)$ by

$$\phi^*(X) = \underset{A \in \phi(X)}{\operatorname{argmin}} \nu_A. \tag{13}$$

Then, $\nu_M = 1$ implies that $\phi^*(N) = M$, since $M \in \phi(X)$. Also, Corollary 1 implies that $\phi^*(X) = X$ if and only if X is self-enforcing, i.e., if and only if $X \in \phi(X)$.

Consider $j = 0$, i.e., the stage of the game before any elimination has taken place. Suppose that the first q $(0 \leq q \leq |N| - 1)$ proposals made by players $i_{0,1}, \ldots, i_{0,q}$ were rejected, and Nature has picked player $i_{0,q+1}$ as the next agenda-setter. Denote $A_q = N \setminus \{i_{0,1}, \ldots, i_{0,q}\}$ for $1 \leq q \leq |N|$ and $A_{-1} = N$. For any player i, let us define the set of minimal self-enforcing coalitions including this player:

$$\chi_i = \underset{A \in \{X : X \subset N, \gamma_X > \alpha\gamma_N, \phi^*(X) = X, X \ni i\} \cup \{N\}}{\operatorname{argmin}} \gamma_A, \tag{14}$$

and a selection from the set,

$$\chi_i^* = \underset{A \in \chi_i}{\operatorname{argmin}} \nu_A. \tag{15}$$

Similarly, for any $X \subset N$, let as define the corresponding sets

$$\chi_X = \begin{cases} \underset{\{A : |\exists i \in X : \chi_i^* = A\}}{\operatorname{argmin}} \gamma_A & \text{if } X \neq \varnothing; \\ \{N\} & \text{otherwise,} \end{cases} \tag{16}$$

and

$$\chi_X^* = \underset{A \in \chi_X}{\operatorname{argmin}} \nu_A. \tag{17}$$

Notice that for any $i \in N$, χ_i^* either equals N or is a winning coalition that satisfies $\phi^*(\chi_i^*) = \chi_i^*$ (or both). The same is true for χ_X^* for any $X \subset N$. Moreover, if $Y \subset X \subset N$, then $\gamma_{\chi_X^*} \leq \gamma_{\chi_Y^*}$.

Next, for a non-empty $X \subset N$, let

$$\psi(X) = \begin{cases} \phi^*(X) & \text{if } X \neq N; \\ \\ N & \text{otherwise.} \end{cases} \tag{18}$$

With $\psi(X)$ defined this way, for any $X \subset N$, $\psi(\chi^*_X) = \chi^*_X$.

We now construct a pure strategy SWDE profile.

Consider the following profile: if $j = 0$, then player $i_{0,q}$ (i.e., the qth proposer, for $1 \leq q \leq |N|$), offers $X_{0,q} = \chi^*_{i_{0,q}}$. For any proposal X made by the qth proposer, player $i \in X$ casts the following vote:

$$v_{0,q}(i, X) = \begin{cases} \tilde{n} & \begin{array}{c} \text{if } i \notin \psi(X) \text{ or} \\ i \in \chi^*_{A_q},\ \gamma_{\psi(X)} \geq \gamma_{\chi^*_{A_q}}, \text{ and } X \neq \chi^*_{A_{q-1}} \end{array}; \\ \\ \tilde{y} & \text{otherwise.} \end{cases} \tag{19}$$

This voting rule implies that a player votes against a proposal that leaves him out of the URC along the continuation equilibrium path and also against a proposal that is not better than a proposal that will be made if the current one is rejected (unless the current proposal is $\chi^*_{A_{q-1}}$).

After the first elimination has taken place, the subgame coincides with $\hat{\Gamma}(N_1, \gamma|_{N_1}, \alpha)$, where $|N_1| < |N|$. By the induction hypothesis, it has a pure strategy SWDE that satisfies the necessary conditions, in particular, it has a pure strategy SWDE that leads to $\phi^*(N_1)$ (after at most one round of elimination).

Now denote the pure strategy profile for the entire game described above by σ^*_{N,ϕ^*}. We will show that it forms a SWDE. The first step is the following lemma.

Lemma 1 *Suppose that the current history h is such that there has been no eliminations (i.e., $j = 0$), the first q proposals ($0 \leq q \leq |N|$) were made by players $i_{0,1}, \ldots, i_{0,q}$ and rejected, and h terminates in the node after the qth rejection (if $q = 0$, $h = \{\varnothing\}$). Then:*

1. *Profile σ^*_{N,ϕ^*} is an h-SWDE.*
2. *The play of σ^*_{N,ϕ^*} after history h will result in coalition $\chi^*_{A_{q-1}}$ (where recall that $A_q = N \setminus \{i_{0,1}, \ldots, i_{0,q}\}$) as the URC with probability 1.*
3. *Continuation utility of player $i \in N$ is given by*

$$u_i\left(\sigma^*_{N,\phi^*} \mid h\right) = w_i\left(\chi^*_{A_{q-1}}\right) - \varepsilon I_{\chi^*_{A_{q-1}}}(i)\, I_{\left\{\chi^*_{A_{q-1}} \neq N\right\}}. \tag{20}$$

Proof. (Part 1 of Lemma) This part of the lemma is also proved by induction. If $q = |N|$, then the game has reached a terminal node, and the URC is N. Player i receives (20), because $\chi_\varnothing = N$. Moreover, since there are no more actions left, profile σ^*_{N,ϕ^*} is h-SWDE.

The induction step goes as follows. Suppose that we have proved the Lemma for the number of rejections greater than or equal to q. Given this, we will establish it for $q - 1$ ($1 \leq q \leq |N|$).

Consider qth voting over proposal X made by player $i_{0,q}$ and suppose that it is accepted. If $X \neq N$ and is accepted, player $i \in X$ will receive payoff

$$u_i^+ = w_i(\psi(X)) - \varepsilon I_{\{X \neq N\}}\left(1 + I_{\{\phi^*(X) \neq X\}} I_{\phi^*(X)}(i)\right). \tag{21}$$

This follows from the induction hypothesis of Theorem 2, which implies that when $X \neq N$, $\psi(X) = \phi^*(X)$, and the continuation payoff of player i in the game $\hat{\Gamma}(X, \gamma|_X, \alpha)$ is given by $w_i(X) - \varepsilon I_{\phi^*(X)}(i)\, I_{\{\phi^*(X) \neq X\}}$, and we also subtract ε when coalition X is not the ultimate one, because $i \in X$. If $X = N$, the game ends immediately and (21) again applies.

If proposal X is rejected, player i receives, by induction of this Lemma, payoff given by (20), so

$$u_i^- = w_i\left(\chi^*_{A_q}\right) - \varepsilon I_{\chi^*_{A_q}}(i)\, I_{\left\{\chi^*_{A_q} \neq N\right\}}. \tag{22}$$

Notice that both (21) and (22) hold regardless of the distribution of votes.

If $i \notin \psi(X)$, then

$$u_i^+ \leq w_i(\psi(X)) = w_i^-.$$

If, in addition, $i \notin \chi^*_{A_q}$, then

$$u_i^- = w_i\left(\chi^*_{A_q}\right) = w_i^-,$$

and if $i \in \chi^*_{A_q}$, then

$$u_i^- \geq w_i\left(\chi^*_{A_q}\right) - \varepsilon > w_i^-$$

(this follows because, by construction, ε is small and satisfies (5)). These expressions imply that in both cases $u_i^+ \leq u_i^-$, and voting $\tilde{n}$ is weakly dominant for player i.

Next, consider the case $i \in \psi(X)$. If all three conditions $i \in \chi^*_{A_q}$, $\gamma_{\psi(X)} \geq \gamma_{\chi^*_{A_q}}$, and $X \neq \chi^*_{A_{q-1}}$ are satisfied, then

$$\begin{aligned} u_i^- &= w_i\left(\chi^*_{A_q}\right) - \varepsilon I_{\chi^*_{A_q}}(i)\, I_{\left\{\chi^*_{A_q} \neq N\right\}} \\ &\geq w_i(\psi(X)) - \varepsilon I_{\{X \neq N\}} \\ &\geq u_i^+, \end{aligned}$$

because $X \neq N$ implies $\psi(X) \neq N$, which together with $\gamma_{\psi(X)} \geq \gamma_{\chi^*_{A_q}}$ implies $\chi^*_{A_q} \neq N$ (and $I_{\chi^*_{A_q}}(i) = 1$). This yields that for player i it is weakly dominant to vote $\tilde{n}$.

Now assume that at least one of these three conditions (i.e., $i \in \chi^*_{A_q}$, $\gamma_{\psi(X)} \geq \gamma_{\chi^*_{A_q}}$, and $X \neq \chi^*_{A_{q-1}}$) does not hold. If $i \notin \chi^*_{A_q}$, then $w_i(\psi(X)) > w_i\left(\chi^*_{A_q}\right)$, and (again, since ε satisfies (5))

$$\begin{aligned} u_i^+ &\geq w_i(\psi(X)) - \varepsilon \\ &> w_i^- \\ &= u_i^-. \end{aligned}$$

If $i \in \chi^*_{A_q}$, but $\gamma_{\psi(X)} < \gamma_{\chi^*_{A_q}}$, then again $w_i(\psi(X)) > w_i\left(\chi^*_{A_q}\right)$ and $u_i^+ > u_i^-$. Finally, if $i \in \chi^*_{A_q}$, $\gamma_{\psi(X)} \geq \gamma_{\chi^*_{A_q}}$, but $X = \chi^*_{A_{q-1}}$, we have

$$\psi(X) = \psi\left(\chi^*_{A_{q-1}}\right) = \chi^*_{A_{q-1}}.$$

Furthermore, since $A_q \subset A_{q-1}$,

$$\gamma_{\psi(X)} = \gamma_{\chi^*_{A_{q-1}}} \leq \gamma_{\chi^*_{A_q}},$$

and therefore $\gamma_{\psi(X)} = \gamma_{\chi^*_{A_q}}$, so $w_i(\psi(X)) = w_i\left(\chi^*_{A_q}\right)$. If $X = N$, then $\psi(X) = N$ and $\gamma_{\chi^*_{A_q}} = N$, so $u_i^+ = u_i^-$. If $X \neq N$ and $X = \chi^*_{A_{q-1}}$, then $\phi^*(X) = X$, and

$$u_i^+ = w_i(\psi(X)) - \varepsilon = w_i\left(\chi^*_{A_q}\right) - \varepsilon = u_i^-.$$

In all these cases we have $u_i^+ \geq u_i^-$, and therefore voting $\tilde{y}$ is weakly dominant.

Now denote the precursor history to history h by h'. The above argument has established that profile σ^*_{N,ϕ^*} is a h'-SWDE, because both in the case the proposal is accepted and in the case it is rejected, this profile is an SWDE for the corresponding history (by induction of Theorem 2 or by induction of this Lemma).

Next consider history h'', after which player $i_{0,q}$ makes a proposal. Denote player i's utility if $i_{0,q}$ makes proposal X at this stage by $u_i(X)$. Similarly, denote his utility if proposal X is made and is accepted by $u_i^+(X)$. We next prove that $u_{i_{0,q}}\left(\chi^*_{i_{0,q}}\right) \geq u_{i_{0,q}}(X)$ for any $X \in P(N)$. First, notice that regardless of $i_{0,q}$'s proposal, the continuation equilibrium play will lead to some URC, Y, satisfying $\psi(Y) = Y$ (if proposal Y is rejected, then the URC will be $\chi^*_{A_q}$, as we have just proved, and if it is accepted it will be either N if

$Y = N$ or $\phi^*(N)$ if $Y \neq N$, which follows from the induction hypothesis in Theorem 2). By definition of $\chi^*_{i_{0,q}}$, $w_{i_{0,q}}\left(\chi^*_{i_{0,q}}\right) \geq w_{i_{0,q}}(Y)$ for any such Y.

Consider the following cases.

(i) $\chi^*_{i_{0,q}}$ is accepted if proposed. Then $\chi^*_{i_{0,q}}$ becomes the URC after at most one round of elimination. This implies that player $i_{0,q}$ receives

$$u_i\left(\chi^*_{i_{0,q}}\right) = u_i^+\left(\chi^*_{i_{0,q}}\right) = w_i\left(\chi^*_{i_{0,q}}\right) - \varepsilon I_{\left\{\chi^*_{i_{0,q}} \neq N\right\}},$$

since $i_{0,q} \in \chi^*_{i_{0,q}}$ and the analysis above has established that he cannot receive higher utility if he proposes some $X \neq \chi^*_{i_{0,q}}$.

(ii) $\chi^*_{i_{0,q}}$ is rejected if proposed. Then consider $i_{0,q}$ proposing some $X \in P(N)$. If X is also rejected when proposed, then $u_{i_{0,q}}\left(\chi^*_{i_{0,q}}\right) = u^-_{i_{0,q}} = u_{i_{0,q}}(X)$. If X is accepted, then the URC is $\psi(X)$. However, for X to be accepted given profile σ^*_{N,ϕ^*}, at least one member of coalition $\chi^*_{A_q}$ must vote $\tilde{y}$ (because $\chi^*_{A_q}$ is a winning coalition). Then (19) suggests, either $\gamma_{\psi(X)} < \gamma_{\chi^*_{A_q}}$ or $X = \chi^*_{A_{q-1}}$ (or both). In the first case, we obtain

$$\gamma_{\psi(X)} < \gamma_{\chi^*_{A_q}} \leq \gamma_{\chi^*_{i_{0,q}}}.$$

This implies that $i_{0,q} \notin \psi(X)$ (otherwise $\chi^*_{i_{0,q}} \notin \chi_{i_{0,q}}$ because $\psi(X)$ has lower total power, see (14)). But then

$$\begin{aligned} u_{i_{0,q}}(X) &= u^+_{i_{0,q}}(X) \\ &\leq w_{i_{0,q}}(\psi(X)). \end{aligned}$$

However, we also have

$$\begin{aligned} u_{i_{0,q}}\left(\chi^*_{i_{0,q}}\right) &= u^-_{i_{0,q}} \\ &= w_{i_{0,q}}\left(\chi^*_{A_q}\right) - \varepsilon I_{\chi^*_{A_q}}(i_{0,q})\, I_{\left\{\chi^*_{A_q} \neq N\right\}}. \end{aligned}$$

Consequently,

$$u_{i_{0,q}}(X) \leq u_{i_{0,q}}\left(\chi^*_{i_{0,q}}\right). \tag{23}$$

In the second case, if $X = \chi^*_{A_{q-1}}$, we have $X = \chi^*_i$ for some $i \in A_{q-1}$. However, $i \neq i_{0,q}$ (otherwise X could not be accepted when $\chi^*_{i_{0,q}}$ is rejected), and thus $i \in A_q$. Then, (16) and (17) imply $\chi^*_{A_{q-1}} = \chi^*_{A_q}$ (because ν reaches its minimum on the same coalition both for $\chi_{A_{q-1}}$ and χ_{A_q}), so $X = \chi^*_{A_q}$. Then,

$$\begin{aligned} u_{i_{0,q}}(X) &= u^+_{i_{0,q}}(X) \\ &= w_{i_{0,q}}(X) - \varepsilon I_X(i_{0,q})\, I_{\{X \neq N\}} \\ &= w_{i_{0,q}}\left(\chi^*_{A_q}\right) - \varepsilon I_{\chi^*_{A_q}}(i_{0,q})\, I_{\left\{\chi^*_{A_q} \neq N\right\}} \\ &= u^-_i \\ &= u_{i_{0,q}}\left(\chi^*_{i_{0,q}}\right), \end{aligned}$$

as implied by (22).

Consequently, in both cases (i) and (ii), (23) holds, and therefore proposing $\chi^*_{i_{0,q}}$ is weakly dominant for player $i_{0,q}$. This establishes that the profile σ^*_{N,ϕ^*} is h''-SWDE.

To prove that σ^*_{N,ϕ^*} is h-SWDE we need to go one more step back, before Nature chooses the qth proposer. This is straightforward, since for Nature all moves are weakly dominant. This completes the induction step and the proof of Part 1.

(Part 2 of Lemma) We need to prove that after history h, $\chi^*_{A_{q-1}}$ will be chosen along the equilibrium path, and this is regardless of the qth proposer $i_{0,q}$ that Nature chooses (provided that $i_{0,q} \in A_{q-1}$, as required). In equilibrium, $\chi^*_{i_{0,q}}$ is proposed by player $i_{0,q}$ (irrespective of the identity of this player is).

Consider two cases.

(i) $\chi^*_{i_{0,q}}$ is accepted (which this happens with probability 1, because strategies are pure), then there is at least one player $i \in \chi^*_{A_q}$ who has voted $\tilde{y}$ ($\chi^*_{A_q}$ is a winning coalition, and if all its members vote $\tilde{n}$, the proposal cannot be accepted). For $i \in \chi^*_{A_q}$ to vote $\tilde{y}$, either $\gamma_{\chi^*_{i_{0,q}}} < \gamma_{\chi^*_{A_q}}$ or $\chi^*_{i_{0,q}} = \chi^*_{A_{q-1}}$ (or both) must hold (see (19)), and note that $\psi\left(\chi^*_{i_{0,q}}\right) = \chi^*_{i_{0,q}}$. In the first case, $\chi_{A_{q-1}}$ is clearly a singleton consisting of $\chi^*_{i_{0,q}}$ only (because the other sets in argmin in (16) have total power greater than $\chi^*_{i_{0,q}}$), and thus $\chi^*_{A_{q-1}} = \chi^*_{i_{0,q}}$. In the latter case, this equality holds automatically. In both cases the URC is $\chi^*_{i_{0,q}} = \chi^*_{A_{q-1}}$, and there is are no more than one round of elimination (because either $\chi^*_{i_{0,q}} = N$ or $\phi^*\left(\chi^*_{i_{0,q}}\right) = \chi^*_{i_{0,q}}$).

(ii) $\chi^*_{i_{0,q}}$ is rejected. In this case, there is at least one player $i \in \chi^*_{A_q}$ who has voted $\tilde{n}$ (otherwise it would be accepted because $\chi^*_{A_q}$ is winning within N), which implies $\gamma_{\chi^*_{i_{0,q}}} \geq \gamma_{\chi^*_{A_q}}$ and $\chi^*_{i_{0,q}} \neq \chi^*_{A_{q-1}}$ (recall (19)). The first inequality implies that $\gamma_{\chi^*_{A_{q-1}}} \geq \gamma_{\chi^*_{A_q}}$ (indeed, $\chi^*_{A_{q-1}}$ either belongs to $\chi^*_{A_q}$ or equals $\chi^*_{i_{0,q}}$), and $A_q \subset A_{q-1}$ implies $\gamma_{\chi^*_{A_{q-1}}} \leq \gamma_{\chi^*_{A_q}}$, so $\gamma_{\chi^*_{A_{q-1}}} = \gamma_{\chi^*_{A_q}}$. Consequently, $\chi_{A_q} \subset \chi_{A_{q-1}}$. The inequality $\chi^*_{i_{0,q}} \neq \chi^*_{A_{q-1}}$ implies that

$$\chi^*_{i_{0,q}} \neq \underset{A \in \chi_{A_{q-1}}}{\operatorname{argmin}}\ \nu_A,$$

therefore,

$$\chi^*_{A_{q-1}} = \underset{A \in \chi_{A_{q-1}}}{\operatorname{argmin}}\ \nu_A \in \chi_{A_q}.$$

But then $\chi_{A_q} \subset \chi_{A_{q-1}}$ implies $\chi^*_{A_{q-1}} = \chi^*_{A_q}$. Since $\chi^*_{i_{0,q}}$ is rejected, induction of Lemma 1 yields that $\chi^*_{A_q}$ is the URC with probability 1, but then the same is true for $\chi^*_{A_{q-1}}$. Moreover, this URC is achieved after at most one round of elimination, since the qth proposal $\chi^*_{i_{0,q}}$ is rejected.

The previous two steps have proved the second part of the lemma both for the case where $\chi^*_{i_{0,q}}$ is accepted in equilibrium and where it is rejected and completes the proof of Part 2.

(Part 3 of Lemma) Since there is at most one round of elimination (one if $\chi^*_{i_{0,q}} \neq N$ and zero otherwise), the continuation utility of player i is given by (20) by definition, and this completes the proof of Lemma 1. ■

We now return to complete the induction step in the proof that σ^*_{N,ϕ^*} forms an SWDE, leading to $M \in \phi(N)$ as the URC with probability 1.

We first use Lemma 1 for $q = 0$, i.e. at the initial node of the game $\hat{\Gamma}(N, \gamma|_N, \alpha)$. Lemma 1 implies that σ^*_{N,ϕ^*} is an h-SWDE where h is empty history, so σ^*_{N,ϕ^*} is an SWDE. The URC is $\chi^*_{A_{-1}} = \chi^*_N$ with probability 1, and it is achieved after at most one round of elimination.

To prove that $\chi^*_N = M$, take any $i \in M$. The fact that $\nu_M = 1$ immediately yields to results: first, $\phi^*(N) = \phi^*(M) = M$; second $M \in \chi_i$ (this is true both if $M = N$ and if $M \neq N$) as $M \in \phi(N)$. Consequently, there exists no coalition Y that is winning within N and satisfies $Y \in \phi(Y)$ such that $\gamma_Y < \gamma_M$.

Moreover, by definition $M = \chi^*_i$, $M \in \chi_N$, and, finally, $M = \chi^*_N$. So, M is the URC with probability 1 and is reached after at most one round of elimination, and players' utilities are indeed given by (7). Hence, this completes the induction proof about for $|N| = n$ and thus establishes that there exists a pure strategy SWDE leading to $M \in \phi(N)$ as the URC with probability 1, with payoff specified in (7), completing the proof of the first part.

(Part 2 of Theorem) First suppose that Assumption 1 holds. We will establish that in any SWDE, the URC $M = \phi(N)$ is achieved with probability 1 after at most one stage of elimination. (Here we use $\phi(X)$ to denote the single element of $\phi(X)$ for any $X \in P(N)$ whenever it does not cause confusion.)

The proof is again by induction on $|N|$. The base, $|N| = 1$, follows immediately, since the only possible URC in any terminal node is $\phi(N) = N$, and therefore in any SWDE it is achieved with certainty (and without elimination).

Now suppose that we have proved that the URC is $M = \phi(N)$ and is achieved with probability 1 after at most one stage of elimination for all coalition sizes strictly less than n. We will then prove it for any coalition N with $|N| = n$. Take any strategy profile σ^* that forms an SWDE.

Consider any node (on or off equilibrium path) before the first elimination took place (i.e., at $j = 0$). Denote the corresponding history by h. Part 1 of the Theorem implies that any URC X that may emerge with a non-zero probability in a SWDE (i.e., any $X \in \phi_{\sigma^*}(N)$) starting from this history must satisfy $\psi(X) = X$ (where, like in (18), $\psi(X) = \phi(X)$ if $X \neq N$ and $\psi(X) = X = N$ otherwise). In particular, if no elimination ever occurs, then the URC is N which, by definition, satisfies $\psi(X) = X$. If an elimination occurs (i.e., some proposal $Y \neq N$ is accepted), then by induction, $\phi(Y)$ becomes the URC. But we have that

$$\phi(\phi(Y)) = \phi(Y) \text{ and } \phi(Y) \neq N,$$

so in this case the URC $\phi(Y)$ must satisfy

$$\psi(\phi(Y)) = \phi(Y).$$

Next, as an intermediate result, we prove that if $Y \in \phi_{\sigma^*}(N)$, then $\phi(Y)$ must be winning within N, i.e., $\gamma_{\phi(Y)} > \alpha\gamma_N$. Suppose, to obtain a contradiction, that this is not the case for coalition Y, i.e., $\gamma_{\phi(Y)} \leq \alpha\gamma_N$. Take the maximum possible number of voting stages, q, where this may happen. Such a maximum exists in view of the fact that we have a finite game. Then by hypothesis the proposal $Y \neq N$ is being accepted with positive probability at the qth voting stage (if all previous proposals are rejected), so $\gamma_Y > \alpha\gamma_N$. Also by hypothesis $\gamma_{\phi(Y)} \leq \alpha\gamma_N$. Moreover, since q is the maximum voting stage where this can happen, if $\phi(Y)$ is rejected, then the URC will be a winning coalition with probability 1. Since $\gamma_Y > \alpha\gamma_N$ and $\gamma_{\phi(Y)} \leq \alpha\gamma_N$, there exists some player $i \in Y \setminus \phi(Y)$ voting $\tilde{y}$ in response to the proposal of Y with a positive probability and who is pivotal for at least one configuration of votes (otherwise the proposal Y could never be accepted). For the configuration of votes where i is pivotal, his vote of $\tilde{y}$ gives him utility

$$w_i(\phi(Y)) - \varepsilon = w_i^- - \varepsilon$$

(in that case, i gets $-\varepsilon$ because he is a member of $Y \neq N$, and then he is eliminated). A vote of $\tilde{n}$ would, on the other hand, give at least least w_i^- in expectation. To see this, let $\phi_{\sigma^*}(N) = \{Z\}$. If $i \in Z$, i receives

$$w_i(Z) - 2\varepsilon > w_i^- - \varepsilon.$$

If $i \notin Z$, that he receives $w_i^- > w_i^- - \varepsilon$. Therefore, $\tilde{n}$ can lead to strictly higher payoff than $\tilde{y}$. This implies that at the qth voting stage, i cannot vote $\tilde{y}$ with a positive probability in SWDE σ^*. Since $i \in Y \setminus \phi(Y)$ was arbitrary, this yields a contradiction and implies that Y with $\gamma_{\phi(Y)} \leq \alpha\gamma_N$ cannot emerge as the URC after any history h. This yields a contradiction and establishes the intermediate result. Consequently, any $X \in \phi_{\sigma^*}(N)$ must satisfy $\psi(X) = X$. This also implies that for any coalition Y with $\gamma_Y < \gamma_{\phi(N)}$, we have $Y \notin \phi_{\sigma^*}(N)$ (and Assumption 1 implies that there are no coalitions with $\gamma_Y = \gamma_{\phi(N)}$).

The next step is to prove that if $M = \phi(N)$ is proposed (before any elimination), then M will necessarily be the URC. Suppose not. Then there exists some history h and player $i \in M$ who casts vote $\tilde{n}$ with a positive probability (and who may be pivotal for some distribution of votes) at history h. For the equilibrium distribution of other player's votes, the vote of $\tilde{n}$ yields a payoff

$$u_i^d < w_i(M) - \varepsilon I_{\{M \neq N\}}$$

for player i. This inequality follows because either M is not the URC, and from the previous paragraph, the URC Z must have

$$\gamma_Z > \gamma_M = \gamma_{\phi(N)},$$

and thus yielding lower payoff to player i than $w_i(M) - \varepsilon I_{\{M \neq N\}} - \varepsilon k_h$, or because M is the URC but will will be reached with more than one round of elimination. Since $\tilde{n}$ is chosen with a positive probability as part of SWDE σ^*, voting $\tilde{y}$ must yield no more than u_i^d, and therefore, even if i voted $\tilde{y}$, M would not have been reached after at most one round of elimination.

Suppose next that i switches to voting $\tilde{y}$, but there is some other player $i' \in M$ who also casts vote $\tilde{n}$ with a positive probability. With a similar reasoning, voting $\tilde{n}$ yields a payoff of

$$u_{i'}^d < w_{i'}(M) - \varepsilon I_{\{M \neq N\}}$$

to player i'. Again since voting $\tilde{n}$ to M for $i' \in M$ in SWDE σ^* must be no worse than voting $\tilde{y}$. This implies that even if i' were to switch to $\tilde{y}$, M will still have a positive chance of not being reached, or reached after

more than one round of elimination. Now proceeding in the same manner through all the players in M, we can conclude that, since σ^* is a SWDE, even if all $i \in M$ vote $\tilde{y}$, M should not be reached after no more than one round of elimination with probability 1. This last observation yields a contradiction, however, because of the following facts: (i) M is a winning coalition, so if all $i' \in M$ vote $\tilde{y}$ to the proposal of M, M will emerge as the URC after at most one round of elimination with probability 1; (ii) when M is reached after at most one round of elimination, each $i \in M$ will obtain

$$w_i(M) - \varepsilon I_{\{M \neq N\}} > u_i^d,$$

implying that voting $\tilde{n}$ to the proposal of M could not be part of any SWDE σ^*. Consequently, we have established that when $M = \phi(N)$ is proposed at $j = 0$ (before any elimination), then M will be the URC with probability 1 under σ^*.

An identical argument applies for any stage $j \geq 0$, where all $i \in M$ have not been eliminated. Therefore, whenever $M = \phi(N)$ is proposed, M will be the URC with probability 1, and there cannot be any other URC.

To complete the proof, it remains to show that M will be offered with probability 1 before any elimination has occurred (i.e., at $j = 0$). Suppose that at $j = 0$ Nature has picked player $i \in M$ as the proposer. It is straightforward to see that in this case M will emerge as the URC after at most one round of elimination. Suppose not. Since we know from the previous paragraph that once M is proposed, it will be accepted with probability 1 after at most one round of elimination, giving payoff $w_i(M) - \varepsilon I_{\{M \neq N\}}$ to player i, it must be that i makes a proposal $\chi_i \neq M$ with a non-zero probability, and receives at least this payoff in expectation. However, since M will be the URC with probability 1 under σ^* and it is not reached in one round of elimination, when the proposal χ_i is accepted, it must be that the payoff to player i must take a value less than or equal to $w_i(M) - \varepsilon I_{\{M \neq N\}} - \varepsilon < w_i(M) - \varepsilon I_{\{M \neq N\}}$. Therefore, $\chi_i \neq M$ (not proposing M) for $i \in M$ cannot be part of any SWDE σ^*. This argument establishes that whenever $i \in M$ is picked by Nature at $j = 0$, the proposal of M will be made and will be accepted with probability 1. Now suppose that Nature picks $i_{0,q} \notin M$ as the proposer before any $i \in M$. Note that at stage $j = 0$, such $i_{0,q} \notin M$ can be picked as the proposer at most $|N| - |M|$ times. Thus, after at most $|N| - |M|$ votes against the proposals, some $i' \in M$ will be picked as the proposer. Then the same argument as above establishes that in any SWDE σ^* all $i \in M$ must vote $\tilde{n}$ with probability 1 to the proposal by any $i_{0,q} \notin M$. This follows since voting $\tilde{y}$ will lead to some payoff $u_i^d < w_i(M) - \varepsilon I_{\{M \neq N\}}$ for $i \in M$. knowing that ultimately Nature will pick some $i' \in M$ at stage $j = 0$, and in any SWDE σ^*, i' will propose M and all $i \in M$ will vote $\tilde{y}$ with probability 1 to this proposal, voting $\tilde{n}$ to the proposal by $i_{0,q} \notin M$ has a payoff $w_i(M) - \varepsilon I_{\{M \neq N\}}$ to each $i \in M$. Therefore, in any SWDE σ^*, all proposals by $i_{0,q} \notin M$ at stage $j = 0$ are rejected with probability 1 and M is proposed and emerges as the URC with probability 1, yielding each player i's a payoff given by (7). This completes the proof of the induction step and proves that under Assumption 1, any SWDE σ^* has $M = \phi(N)$ as the URC with probability 1, it is reached after at most one round of elimination, and players' payoffs are given by (7). This completes the proof of Theorem 2. ■

Proof of Theorem 3.

The proof is by induction on $|N|$. The base, $|N| = 1$, is trivial: in this case, $\phi(N) = 1$, whereas the two feasible allocations are $x_i = 0$ and $x_i = 1$, where i is the only player. Evidently, the latter is the core allocation, because both may be enforced by coalition N. This proves the base of the induction; now suppose that we have proved Theorem 3 for all coalition sizes less than n; let us prove it for any coalition N with $|N| = n$.

For any $X \subset N$, let x^X be the allocation given by

$$x_i^X = \begin{cases} w_i(X) & \text{if } i \in X; \\ \\ 0 & \text{otherwise.} \end{cases}$$

(Note that such allocation x^X may well be not feasible.)

First, take any core allocation x, then $x^+ \subset N$ is the set of players receiving positive payoff. Clearly, $x^+ \neq \varnothing$, for otherwise coalition N would deviate to x^N (it is feasible because of Definition 9), and all members of N would be better off. Moreover, x^+ is a winning coalition ($\gamma_{x^+} > \alpha\gamma_N$), for otherwise $\gamma_{x^0} \geq (1-\alpha)\gamma_N$,

and then non-empty coalition x^0 would be able to deviate to the same allocation x^N, making each member of x^0 better off.

Given this, Definition 9 implies that either $x_i = w_i(N)$ for all $i \in N$, or, by the induction hypothesis, $x^+ \in \phi(Y)$ for some $Y \subset N$, $Y \neq N$ (so x^+ is self-enforcing), and $x_i = w_i(x^+)$.

Now suppose, to obtain a contradiction, that $x^+ \notin \phi(N)$. Take coalition $X \in \phi(N)$. We have that $\gamma_{x^+} > \gamma_X$, because x^+ is winning within N and is either self-enforcing or equal to N. But then the winning coalition X can make all its members better off by choosing x^X, which is feasible because either $X = N$, or $X \neq N$ and

$$X \in C\left(\tilde{\Gamma}\left(X, \gamma|_X, \alpha, v_X(\cdot)\right)\right)$$

by induction. In particular, in this case, player $i \in X \cap x^+$ would obtain $w_i(X) > w_i(x^+)$, since $\gamma_{x^+} > \gamma_X$, while player $i \in X \setminus x^+ = X \cap x^0$ would receive $w_i(X) > 0 = w_i^-$. Therefore, coalition X has a feasible deviation that makes all its members better off, contradicting the hypothesis that $x^+ \notin \phi(N)$ may be a core allocation. This contradiction establishes that x is a core allocation. Moreover, if $x^+ = N$, Definition 9 implies that $x = x^N$, and $x_i = w_i(N)$. If, on the other hand, $x^+ \neq N$, then $x|_Y$ is a core allocation in the game $\tilde{\Gamma}(Y, \gamma|_Y, \alpha, v_X(\cdot))$ for some $Y \in P(N)$ such that $x^+ \subset Y$. But then by induction for any $i \in x^+$, we have $x_i = w_i(x^+)$. We have therefore proved that in any core allocation, the set of players receiving a positive payoff must constitute a coalition belonging to $\phi(N)$.

To complete the induction step, take any $M \in \phi(N)$. Both if $M = N$ and if $M \neq N$, x^M (satisfying $x_i^M = w_i(M)$ for $i \in M$ and $x_i^M = 0$ for $i \neq M$) is feasible (if $M \neq N$ this follows from the induction hypothesis). We will now prove that this is a core allocation. Suppose not. Then there exists coalition X and allocation x' enforced by coalition X that makes all members of X better off. This implies that $x_i' > 0$ for all $i \in X$ (if $x_i' = 0$, then $x_i^M \geq 0 = x_i'$ and would contradict the fact that i is better off). So, $X \subset (x_i')^+$. If $(x_i')^+ = N$, then it $\gamma_X \geq (1-\alpha)\gamma_N$ (otherwise, X would not be able to enforce x'), then $M \cap X \neq \varnothing$. Take any player $i \in M \cap X$, and note that for such a player i, we have

$$\begin{aligned} x_i^M &= w_i(M) \\ &\geq w_i(N) \\ &= x_i'. \end{aligned}$$

This contradicts the fact that all members of coalition X are strictly better off by switching to x'. Now consider the case $(x_i')^+ \neq N$. Since $(x_i')^+ \neq \varnothing$, there exists $Y \subset N$, $Y \neq N$, such that $x' \subset Y$ and $x'|_Y$ is the core allocation for $\tilde{\Gamma}_Y = \tilde{\Gamma}(Y, \gamma|_Y, \alpha, v_Y(\cdot))$. By induction, $x_i' = w_i\left((x')^+\right)$ if $i \in (x')^+$. For x' to be enforceable, X must be a winning (within N) coalition. But since all members of X are better off, we have that $X \subset (x')^+$, so $(x')^+$ is also winning within N. By induction, $(x')^+ \in \phi(Y)$, so $(x')^+ \in \phi\left((x')^+\right)$. In other words, $(x')^+$ is winning and self-enforcing, therefore, $\gamma_{(x')^+} \geq \gamma_M$ (because $M \in \phi(Y)$). Both X and M are winning, therefore, there exists $i \in M \cap X$, and we have

$$\begin{aligned} x_i^M &= w_i(M) \\ &\geq w_i\left((x')^+\right) \\ &= x_i'. \end{aligned}$$

This implies that $i \in X$ does not prefer to switch to x', yielding a contradiction. Therefore, x^M is a core allocation.

We next prove that it is the only core allocation satisfying $x^+ = M$. Suppose, to obtain a contradiction, that there is another allocation x. This allocation, x, must satisfy $x_i = w_i(x^+)$ if $i \in x^+$, and $x_i = 0$ if $i \notin x^+$, which means $x = x^M$. Therefore, the specified core allocation is unique.

This completes the proof of the induction step, which proves the one-to-one correspondence between core allocations and coalitions from $\phi(N)$. If Assumption 1 holds, then $\phi(N)$ is a singleton by Theorem 1. If $\phi(N) = \{M\}$, then, as we have proved, x^M is a core allocation, moreover, any core allocation is x^M, and $\left(x^M\right)^+ = M$ is unique. This completes the proof of Theorem 3. ■

References

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

Ansolabehere, Stephen, James M. Snyder, Jr., Aaron B. Strauss, and Michael M. Ting (2005) "Voting Weights and Formateur Advantages in the Formation of Coalition Governments," *American Political Science Review* 99.

Arrow, Kenneth (1951) *Social Choice and Individual Values.* New York, Wiley&Sons.

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

Aumann, Robert and Roger Myerson (1988), "Endogenous Formation of Links between Players and of Coalitions," in A. Roth (ed.), The Shapley Value, Cambridge: Cambridge University Press, 175-191.

Austen-Smith, David and Jeffrey Banks (1999) *Positive Political Theory.* Ann Arbor, U.Michigan Press.

Axelrod, Robert (1970) *Conflict of Interest,* Markham: Chicago.

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.

Browne, Eric and Mark Franklin (1973) "Aspects of Coalition Payoffs in European Parliamentary Democracies." *American Political Science Review* 67: 453-469.

Browne, Eric and John Frendreis (1980) "Allocating Coalition Payoffs by Conventional Norm: An Assessment of the Evidence from Cabinet Coalition Situations," *American Journal of Political Science* 24: 753-768.

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.

Gamson, William A. (1961) "A Theory of Coalition Formation," *American Sociological Review* 26: 373-382.

Fudenberg, Drew and Jean Tirole (1991) *Game Theory,* MIT Press, Cambridge, MA,

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

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.

Isbell, J. R. (1956) "A Class of Majority Games," *Quarterly Journal of Mathematics Oxford Series*, 7, 183-187.

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

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

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

Myerson, Roger (1991) *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge, MA.

Nash, John. F. (1953) "Two Person Cooperative Games," *Econometrica* 21, 128-140.

Osborne, Martin and Ariel Rubinstein (1994) *A Course in Game Theory.* MIT Press, Cambridge, MA.

Peleg, Bezalel (1968) "On Weights of Constant-Sum Majority Games," *SIAM Journal of Applied Mathematics*, 16, 527-532.

Peleg, Bezalel (1968) "A Theory of Coalition Formation in Committees," *Journal of Mathematical Economics*, 7, 115-134.

Peleg, Bezalel and Joachim Rosenmuller (1992) "The Least Core, Nucleolus, and Kernel of Homogeneous Weighted Majority Games," *Games and Economic Behavior*, 4, 588-605.

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, Connecticut.

Schofield, Norman, and Michael Laver (1985) "Bargaining Theory and Portfolio Payoffs in European

Coalition Governments 1945-1983," *British Journal of Political Science* 15:143-164.

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

Selten, Reinhard (1975) "Reexamination of the Perfection Concept of Equilibrium in Extensive Games," *International Journal of Game Theory* 4: 25—55.

Sened, Itai (1996) "A Model of Coalition Formation: Theory and Evidence", *Journal of Politics,* Vol. 58, No. 2 (May, 1996), pp. 350-372.

Serrano, Roberto (2005) "Fifty Years of the Nash Program, 1953-2003," *Investigaciones Economicas* 29, 219-258.

Shapley, Lloyd S. and Martin Shubik (1954) "A Method for Evaluating the Distribution of Power in a Committee System," *American Political Science Review* 48: 787-792.

Warwick, Paul and James Druckman (2001) "Portfolio Salience and the Proportionality of Payoffs in Coalition Governments." *British Journal of Political Science* 31:627-649.